How can we optimise shortening of Hepatitis C treatment? Observational and interventional approaches

Leanne Gwyneth McCabe

Institute of Clinical Trials and Methodology, UCL

Submitted for the degree of Doctor of Philosophy

Declaration

I, Leanne Gwyneth McCabe confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:

Date: 30/06/2023

Abstract

Although direct-acting antivirals (DAAs) are shorter and better tolerated than older treatments for Hepatitis C (HCV), many HCV-infected patients who are not yet cured will find adhering to standard treatment lengths challenging. However, which patients might achieve cure with short-course treatment is unclear.

This thesis used data from two trials investigating short-course HCV treatment, STOP-HCV-1 and SEARCH-1, to explore factors associated with cure, time to treatment failure and with the kinetics of viral load (VL) rebound to help understand which patients might be most suitable for short-course treatment. Quality of life (QoL) with varying treatment durations and ribavirin was investigated to determine acceptability of treatment options. The statistical aspects of VIETNARMS, a HCV treatment trial with an early stopping mechanism, were assessed.

Baseline VL was strongly associated with all outcomes. The probability of cure was highest in those with HCV genotype 1b, CC IL28B genotype, no DAA resistance and no current substance abuse at baseline. Treatment failure happened quicker in older participants, those with genotype 1a and with shorter treatment lengths. Peak rebound VL was higher in those taking ribavirin, no DAA resistance and a higher Fibroscan score, and faster in those with IL28B genotype CC.

There were no clear differences in QoL between those taking different lengths of treatment and no lasting impact on QoL of taking ribavirin.

A statistically rigorous approach to monitoring the VIETNARMS design was developed: the probability of incorrectly or correctly stopping a group during follow-up was appropriate and overall power at the final analysis will be sufficiently high.

Overall, baseline VL was the main determinant of suitability for short-course treatment, with HCV genotype, IL28B genotype and pre-existing DAA resistance also key factors. Treatment lengths and ribavirin use can be tailored to these patient characteristics without compromising QoL and could be tested in well-designed trials.

Impact Statement

Chronic infection with the Hepatitis C virus (HCV) is now easily treatable with highly efficacious direct acting antivirals, which are shorter and better tolerated than older HCV treatments. It is estimated that only 62% of those with known HCV infection have received treatment; a high proportion of those still left to treat are in hard-to-reach populations who may struggle with adherence to standard durations of treatment (8-12 weeks). More research is needed to determine which patients are suitable for short-course treatment.

In this thesis I used data from the STOP-HCV-1 and SEARCH-1 trials to help understand better who might successfully achieve cure with short-course treatment and how to investigate several different strategies in one clinical trial where there is little prior evidence which are the ones most likely to be beneficial to patients.

I identified various factors that were associated with cure on short-course HCV treatment, primarily viral load at the start of treatment, which could be used to generate selection or exclusion criteria for receiving short-course treatment or for participation in a clinical trial. I also identified factors associated with time until treatment failure and the kinetics of VL rebound, which could also aid in the development of drug shortening strategies by suggesting which patients need longer treatment, that could then be tested in clinical trials.

Looking at quality of life (QoL) of trial participants, I did not find that this was related to length of short-course treatment. I found that although there was a trend for worse QoL for those taking ribavirin, this no longer existed by the timepoint when cure was determined. I also showed that some of the negative effect on QoL from ribavirin was mediated through anaemia, and therefore, timely treatment of anaemia might prevent these effects. With this information, longer treatment or adjunctive ribavirin could be considered when testing new strategies, as in VIETNARMS, or clinicians could feel more confident prescribing longer treatments or ribavirin if their main concern about using those regimens was worse QoL for the patient leading to potentially lower adherence.

It is my work on the VIETNARMS trial design that has had the most impact to date. When I assessed the VIETNARMS design, the trial had received funding, but was still in the design phase and had not yet received ethical approval. The development of the stopping guideline and the interim analysis monitoring plan were included both within the trial protocol and in presentations to approval committees in support of the trial. Although the monitoring plan could not be followed according to the anticipated timing of visits due to delays in recruitment caused by COVID-19 lockdowns, it was followed as closely as possible according to the number of outcomes expected. The VIETNARMS trial is currently in follow-up with all interim analyses

completed and a planned first-line last participant last visit in October 2023. This work was presented orally to clinical trial statisticians and methodologists at the International Clinical Trials Methodology Conference in 2019 and later published in Trials.

Acknowledgements

I would like to thank my supervisors Sarah Walker, Ian White and Graham Cooke for their advice and guidance. I would also like to thank the STOP-HCV-1 and SEARCH-1 trial teams and participants for the data used in the thesis.

UCL Research Paper Declaration Form

referencing the doctoral candidate's own published work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- *have been uploaded to a preprint server*
- *are in submission to a peer-reviewed publication*
- *have been published in a peer-reviewed publication, e.g. journal, textbook.*

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.

1. **For a research manuscript that has already been published** (if not yet published, please skip to section 2)

   a) **What is the title of the manuscript?**

   The design and statistical aspects of VIETNARMS: a strategic post-licensing trial of multiple oral direct-acting antiviral hepatitis C treatment strategies in Vietnam

   b) **Please include a link to or doi for the work**

   https://doi.org/10.1186/s13063-020-04350-x

   c) **Where was the work published?**

   Trials

   d) **Who published the work?** (e.g. OUP)

   BMC

   e) **When was the work published?**

   18/05/2020

   f) **List the manuscript's authors in the order they appear on the publication**

   Leanne McCabe, Ian R. White, Nguyen Van Vinh Chau, Eleanor Barnes, Sarah L. Pett, Graham S. Cooke & A. Sarah Walker on behalf of SEARCH investigators

   g) **Was the work peer reviewed?**

Yes

h) **Have you retained the copyright?**

Yes

i) **Was an earlier form of the manuscript uploaded to a preprint server?** (e.g. medRxiv). If 'Yes', please give a link or doi)

https://arxiv.org/abs/1911.02272

If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

☐

*I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.*

2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)

a) **What is the current title of the manuscript?**

Click or tap here to enter text.

b) **Has the manuscript been uploaded to a preprint server?** (e.g. medRxiv; if 'Yes', please give a link or doi)

Click or tap here to enter text.

c) **Where is the work intended to be published?** (e.g. journal names)

Click or tap here to enter text.

d) **List the manuscript's authors in the intended authorship order**

Click or tap here to enter text.

e) **Stage of publication** (e.g. in submission)

Click or tap here to enter text.

3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4)

LM performed the statistical analysis and wrote the first draft of the manuscript. GSC, NVVC, EB, SLP and ASW designed the trial. IRW, GSC and ASW provided critical comments on initial drafts of the manuscript. All authors read and approved the manuscript.

4. **In which chapter(s) of your thesis can this material be found?**

   5

5. **e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

   Candidate

   Click or tap here to enter text.

   Date:

   30/06/2023

   Supervisor/ Senior Author (where appropriate)

   Date

   20 June 2023

Table of contents

List of tables

List of figures

Abbreviations

AIC	Akaike's information criteria
ALT	alanine transaminase
ALP	alkaline phosphatase
AST	aspartate aminotransferase
AUROC	area under the receiving curve
BMI	body mass index
CI	confidence interval
DAA	direct acting antiviral
DMC	Data Monitoring Committee
D*X*	*X* days after enrolment into a study
eGFR	estimated glomerular filtration rate
EOT	end of treatment
EOT+*X*	*X* weeks after end of treatment
FIB-4	fibrosis-4 (index)
HCV	Hepatitis C virus
HIV	human immunodeficiency virus
HR	hazard ratio
IQR	interquartile range
LLOQ	lower limit of quantification
MCS	mental component summary (for SF-12)
OR	odds ratio
PCS	physical component summary (for SF-12)
PEG-IFN	pegylated-interferon
pOR	proportional odds ratio
QoL	quality of life
RAS	resistance associated substitution
RD	risk difference
RGT	response guided therapy
RNA	ribonucleic acid
RT D0	day 0 of retreatment
SD	standard deviation
SVR	sustained virologic response
SVR*X*	sustained virologic response *X* weeks after completing treatment
VAS	visual analogue scale (for EQ-5D-5L)

VL	viral load
WHO	World Health Organization

1 Introduction

1.1 Hepatitis C

Hepatitis C virus (HCV) is one of the five known hepatitis viruses and without treatment can lead to liver disease, cirrhosis, hepatocellular carcinoma and death. First identified in 1989 (1), there were an estimated 57 million people infected globally in 2020 (2) and it is one of the leading causes of death (3). Compared to 1990, the number of people estimated to be infected has decreased due to a lower infection rate, initially mainly due to a reduction in nosocomial infections (4) and an increase in mortality caused by those already infected advancing to late stage liver disease, but more recently and more substantially following the introduction of direct-acting antivirals (DAAs). In order to reduce the impact of HCV, in 2016 the WHO set targets for the elimination of HCV as a global health threat by 2030, by reducing incidence of chronic HCV infections by 90% and mortality from chronic HCV infections by 65% (5).

HCV initially starts as an acute infection, which is often asymptomatic, and approximately a third of those infected cure spontaneously within the first 6-12 months after infection without the need for treatment (6). Spontaneous clearance is significantly lower in males, those aged ≥45 years, those with IL28B genotypes CT or TT and those co-infected with HIV. It is also lower in those with a history of injecting drug use and current excessive alcohol use, who may also struggle to adhere to HCV treatment (7). For those who do not spontaneously clear the virus in this time period, their infection is considered chronic and cure can usually only be achieved with treatment.

HCV is estimated to affect between 0.5-1.6% of the global population, depending on region; prevalence is highest in the Eastern Mediterranean region (1.6%) followed by Europe (1.3%) with a prevalence of 0.8% in Africa and 0.5% elsewhere (8). However, there are large disparities in the proportion of those with a known diagnosis and who have accessed treatment. In the Eastern Mediterranean, it is estimated that 37% of people infected with HCV are aware of their status and 33% have accessed treatment, indicating a very high level of treatment in those who have been diagnosed. In Africa, it is estimated that 5% are aware of their status, but only 0.5% have accessed treatment.

There are 7 HCV genotypes, each consisting of numerous subtypes, with varied global distributions (4). In Europe and Oceania, the most prevalent genotypes are 1 and 3, with genotypes 1 and 2 being most common in North America. This has led to a large number of studies focused on genotypes 1 and 3, with limited research into other genotypes. Within the Middle East, particularly Egypt, Kuwait and Qatar, and parts of Sub-Saharan Africa, notably the Democratic Republic of Congo and Burundi, genotype 4 is the most common genotype.

Genotype 5 is mostly found in South Africa and Mozambique, with only a very small proportion of the distribution outside of these countries. Genotype 6 is most commonly found in South-East Asia, making up over half the genotype distribution in those infected in Cambodia, Laos and Vietnam.

HCV viral load (VL) is the measure of HCV RNA circulating in the body. After the initiation of treatment, VL declines in two phases: a fast first phase where virus is cleared from circulation and a slower second phase where infected cells die. There have been very few studies investigating the factors affecting viral load decay. Of the five studies I have identified, only ribavirin, low baseline HCV VL and low baseline ALT were associated with faster declines. Further detail into the mechanism of decline and associated factors is given in Section 3.2.

HCV cure is measured by sustained virologic response (SVR), which is defined as having undetectable levels of HCV RNA in the blood, usually at least 12 (SVR12) or 24 (SVR24) weeks after completing treatment. Those who have achieved SVR have significantly lower rates of morbidity and overall mortality, as well as mortality due to hepatocellular carcinoma, compared to both those who have not been treated and those who have been treated but did not achieve SVR (9). As well as preventing occurrence of future complications, SVR can also lead to a reduction in cirrhosis or stage of fibrosis (10).

1.2 Historical treatment with interferon

Initial treatment for HCV infection consisted of injectable interferon monotherapy for 24 or 48 weeks, with very low SVR rates of 6% and 20% respectively (11) (Figure 1.1). The addition of ribavirin in 1998 increased SVR rates significantly to 32%. SVR rates increased further after 2001 with the use of pegylated interferon (PEG-IFN), which increased both concentrations of interferon and its half-life, to approximately 50% after 48 weeks of treatment, although SVR rates could be as high as 80% in genotypes 2 and 3 (12).

SVR rates were affected by lack of adherence due to the long treatment regimens and poor tolerance due to adverse events, with 9-42% of patients requiring dose reductions and 7-13% stopping treatment altogether (12, 14, 15). Incomplete adherence due to poor tolerance was particularly an issue in those taking ribavirin over these long periods of time (15). With interferon, the majority of patients experienced flu-like symptoms at the beginning of treatment. Other common adverse events included fatigue, insomnia and alopecia. Severe adverse events were less common but included depression with suicidal ideation and autoimmune disease. Anaemia was very common with ribavirin (16), and was severe enough to require treatment with blood transfusions in some cases. Ribavirin also has teratogenic

effects so care needed to be taken to prevent pregnancy in the patient or patient's partner for up to 6 months after completing treatment.

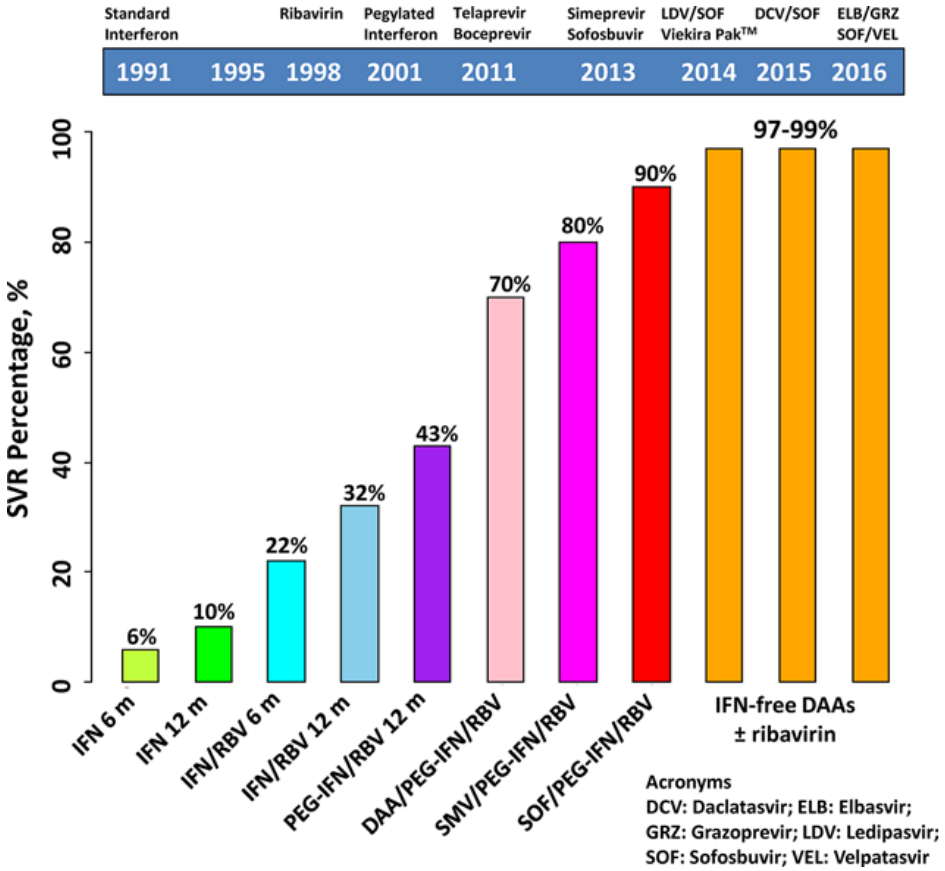


Figure 1.1: SVR rates over time for standard HCV treatments (reproduced from (13))

Image reproduced with the permission of the rights holder, John Wiley and Sons.

Though standard treatment with interferon was 24-48 weeks, patients who had a low baseline VL (<400,000 IU/ml) and undetectable RNA 4 weeks after starting treatment had comparable SVR rates with only 12 weeks of treatment (17, 18).

1.3 Direct-acting antivirals

DAAs are oral medications that have greatly improved SVR rates and have allowed more people infected with HCV to access treatment. First used in combination with interferon and ribavirin in 2011, DAA regimens are now 8-12 weeks in length, interferon-free and ribavirin-free and are the WHO recommended treatment for HCV (19). SVR rates with DAAs are up to 100%, depending on genotype and regimen (Figure 1.1). Adherence has increased and there are lower rates of adverse events, with <10% patients stopping treatment or experiencing a serious adverse event (20).

There are three classes of DAAs, each targeting a different protein involved in viral replication: NS3/4A protease inhibitors, NS5A inhibitors and NS5B inhibitors. Commonly used DAAs and their drug classes are listed in Table 1.1 (21).

The NS3/4A protease is an enzyme that cleaves the large polyprotein initially created from the translation of the viral RNA into the various structural and non-structural proteins that are necessary for viral replication. NS3/4A protease inhibitors prevent this cleavage from occurring and so the virus is unable to replicate. The NS5A protein is involved in many stages in viral replication including interacting with both other viral proteins, particularly NS5B, and the host cell to create a more favourable environment for viral replication, as well as having a role in viral assembly and release. While the exact mechanism of both NS5A itself and NS5A inhibitors is unclear, it is hypothesised that the inhibitors work to prevent multiple NS5A processes from occurring (22). NS5B is responsible for replicating viral RNA by catalysing the process of adding new nucleotides to complementary RNA strands, using the viral RNA as a template. The two NS5B inhibitors differ in their mechanism of action: sofosbuvir incorporates into the RNA chain created by NS5B and acts as a chain terminator by preventing further growth of the RNA, whereas dasabuvir binds to NS5B directly to inhibit further elongation of the RNA chain (23).

Table 1.1: Direct-acting antivirals and their classes

NS3/4A protease inhibitors	NS5A inhibitors	NS5B inhibitors
Glecaprevir*	Daclatasvir*	Dasabuvir
Grazoprevir	Elbasvir	Sofosbuvir*
Paritaprevir	Ledipasvir	
Simeprevir	Ombitasvir	
Voxilaprevir	Pibrentasvir*	
	Velpatasvir*	

*Used as part of WHO recommended pangenotypic regimens (19)

Recommended regimens combine classes of DAAs; these combination regimens are more effective and have a lower risk of drug resistance than a single class due to their different mechanism of action. WHO recommended pangenotypic (i.e. effective against every genotype of virus) regimens consist of the NS5B inhibitor sofosbuvir with a NS5A inhibitor, either daclatasvir or velpatasvir, or the NS3/4A protease inhibitor glecaprevir with the NS5A inhibitor pibrentasvir (19).

The initial cost of DAAs was high and ranged from US$900 to $95,000 per cure (24), but increased competition and development of more DAAs has led to a reduction in the cost of

treatment in some settings. Some national health systems have negotiated further reductions in the cost of DAAs reducing the cost burden of treatment (25) and voluntary licences allowing the production of DAAs by generic manufacturers have reduced the costs in many low and low-middle income countries (26). Although these lower costs have made DAA treatment more affordable for many people, countries with a high burden of HCV may still struggle to eradicate the disease and patients in countries with no access to generic drugs who are expected to pay for or contribute to the costs of their treatment may find it particularly hard.

1.4 Short-course DAA treatment for HCV

Despite DAAs being better tolerated with shorter treatment lengths, there are still barriers to treating hard-to-reach populations such as prisoners, people who inject drugs and the homeless. At the end of 2019, the WHO estimated that 9.4 million people with chronic HCV infection had received DAA treatment, which is 62% of those who knew their diagnosis but only 13% of the total of those infected (8). To achieve the WHO's target for 2030, it will be important to treat both those who struggle to adhere to treatment and those who are unable to afford treatment. Short-course treatments are more likely to be completed by the former group and are more likely to be affordable to the second group.

Literature on previous studies investigating short-course treatment is reviewed in Section 2.2. Of the twelve studies identified testing DAA treatment lengths of 6 weeks or shorter, the primary factor associated with SVR12 was a low baseline VL or a rapid response to treatment. In studies testing adjunctive ribavirin or PEG-IFN, higher SVR12 rates were seen in those groups compared to participants not receiving adjunctive drugs. Other factors associated with SVR12 were HCV genotype (1b compared to 1a, non-3 compared to 3) and lack of pre-existing resistance to prescribed DAAs.

The WHO has previously highlighted the need for research identifying predictive factors to determine which patients are suitable for short-course treatment (19). Not all HCV patients need the full duration of 8-12 weeks treatment to achieve cure, but it is not known which groups can take shorter treatment and what length of treatment they would need. The simplest strategy for allocating shortening treatment would be to prescribe a length of treatment based on a patient's individual characteristics, either based on large categories such as baseline VL and genotype or on a wider set of criteria to create a more individualised treatment regimen. Another strategy for shortening treatment is response guided therapy (RGT). RGT involves measuring the VL of a patient at a given time after starting treatment and if the VL is below a given cut-off, the treatment length can be shortened. Historically with PEG-IFN, the time point at which the VL was measured was after 28 days, but this time point would

not be informative with DAAs as most patients would be virologically suppressed and so a shorter time point is used, usually after 2 or 7 days.

1.5 Introduction to trials of short-course HCV treatment used within the thesis

Data from two trials investigating the use of short-course treatment were used within this thesis. The STOP-HCV-1 trial was a factorial trial conducted in the UK which investigated short-course HCV treatment, where the duration of therapy was based on baseline VL, as well as the use of adjunctive ribavirin. The SEARCH-1 trial was a single arm trial conducted in Vietnam, which opened to recruitment after the analysis of STOP-HCV-1 had taken place and determined length of treatment using VL at day 2 or 14, depending on severity of liver disease. It was also the pilot trial for VIETNARMS, described in Chapter 5, which, at the time of submission, had closed to recruitment and was in follow-up with an expected first-line last participant last visit date in October 2023.

1.5.1 The STOP-HCV-1 trial

STOP-HCV-1 was an open-label, non-inferiority randomised controlled trial testing biomarker-stratified short-course treatment with or without ribavirin within the UK (Figure 1.2) (27). Between 18th March 2016 and 28th August 2018, the trial recruited 204 adults infected with HCV genotype 1a, 1b or 4 for ≥6 months, with mild liver disease (Fibroscan score ≤7.1kPa) and a screening HCV viral load that was detectable but <10,000,000 IU/ml. Any participants co-infected with HIV also had to be HIV suppressed for >24 weeks at screening. Study sites were primarily in England (12 sites), with 1 site each in Scotland and Wales; the three largest recruiting areas were London (5 sites, 40% of total recruited), Brighton (1 site, 14% of total) and Nottingham (12% of total).

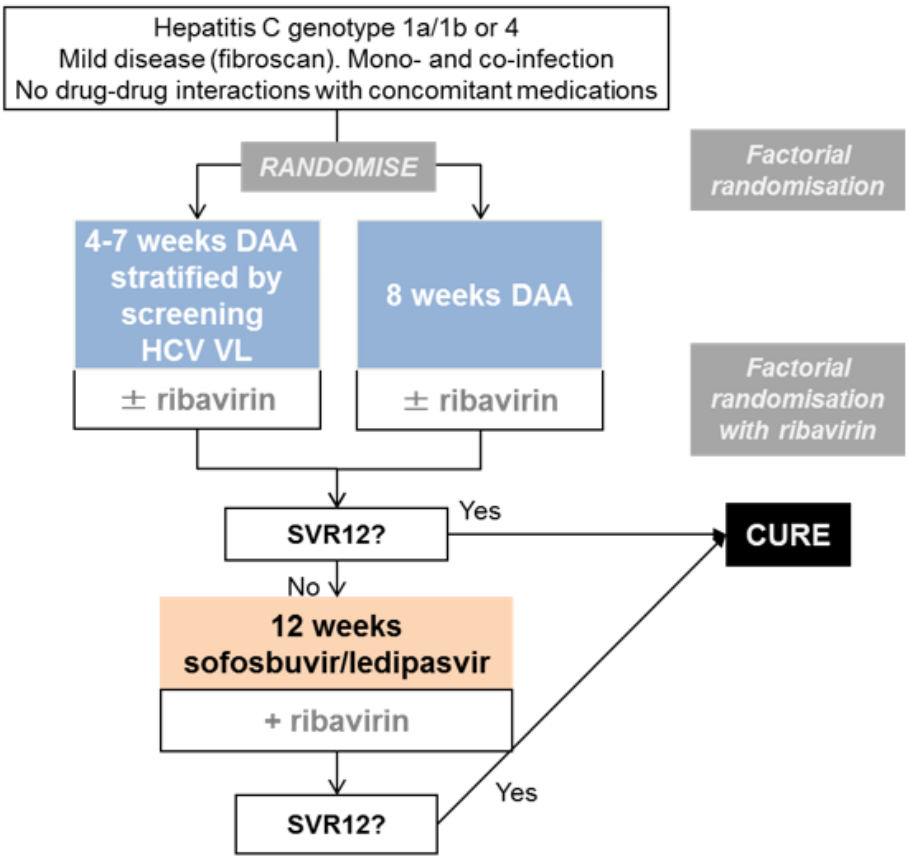


Figure 1.2: STOP-HCV-1 trial schema

Participants were factorially randomised to either 4-7 weeks of DAA with the precise duration determined using a sliding scale based on screening HCV viral load (Figure 1.3) vs 8 weeks of DAA (control), and also to receive ribavirin or not. First-line regimens were ombitasvir/paritaprevir/ritonavir for genotype 4 participants with additional dasabuvir for genotype 1 participants, and glecaprevir/pibrentasvir for all participants where locally available. Those who failed first-line treatment at any point received retreatment, which consisted of 12 weeks of sofosbuvir/ledipasvir and ribavirin. The primary outcomes of the trial were SVR12 after first-line or any necessary retreatment for the duration comparison and SVR12 after first-line treatment only for the ribavirin comparison.

In the STOP-HCV-1 trial protocol, treatment failure was defined as either:

"i. two consecutive measurements of HCV RNA >LLOQ (taken at least one week apart) after two consecutive visits with HCV RNA <LLOQ, at any time, with the latter confirmatory measurement also being >2000 IU/ml, or

ii. two consecutive measurements of HCV RNA (taken at least one week apart) that are >1 log10 increase above HCV RNA nadir on treatment and >2000 IU/ml, at any time."

The time of treatment failure was the first VL taken that constituted the first of the two VLs needed in the definition of treatment failure, with the second VL considered as only a confirmatory result.

At the onset of the trial, those randomised to the variable duration group received 4-6 weeks (VUS1) of treatment based on the screening HCV viral load. After the DMC meeting in April 2017, the Data Monitoring Committee (DMC) recommended that the treatment length should be altered based on a pre-specified adaptation detailed in the protocol. Participants randomised after 1st April 2017 received 4-7 weeks (VUS2) of treatment (Figure 1.3).

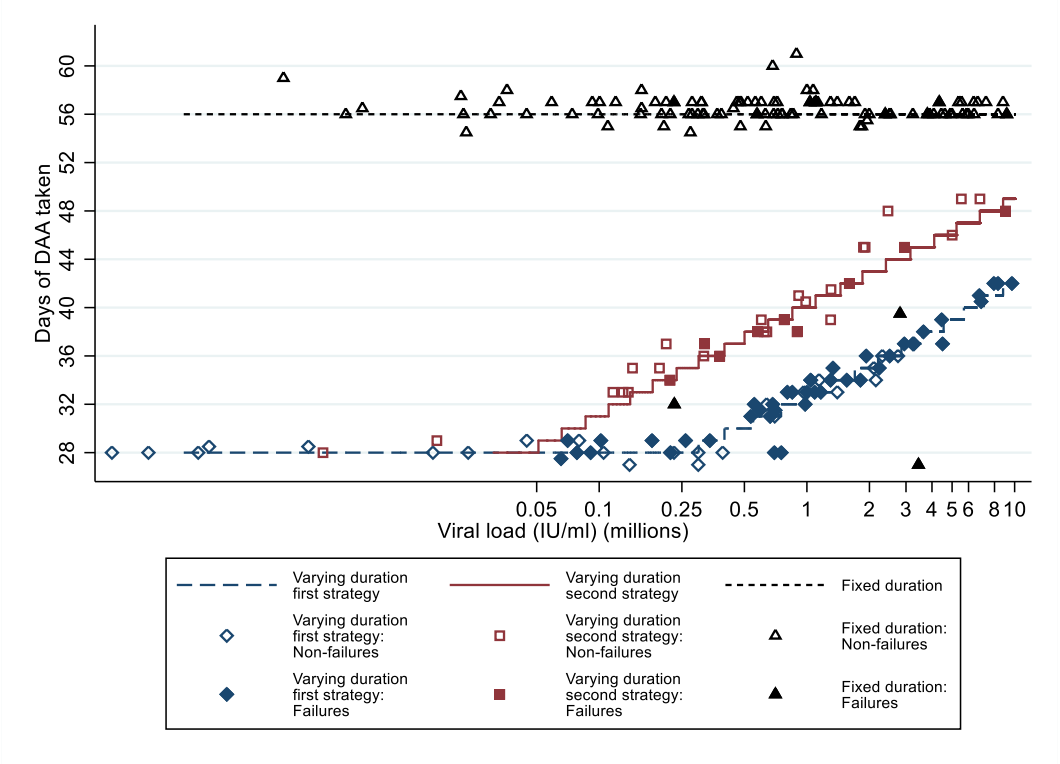


Figure 1.3: Allocated days of DAA based on baseline HCV VL with treatment length allocated to all participants within STOP-HCV-1

The trial was designed with a sample size of 408 participants, but the trial was stopped due to slow recruitment after only 204 were randomised, with 202 starting first-line treatment and included in analyses. Recruitment was slow due to not being able to access enough drug for treatment and, later, due to the unsuitability of the trial for those who were left to treat. Therefore, the trial participants were likely to resemble those with mild liver disease who were being treated at the time of the trial, they would be less likely to resemble those who were left to treat and would benefit most from the shortening strategy, however this would be the case for any trial testing short-course HCV treatment.

Within the participants included in these analyses, those randomised to receive fixed duration treatment took a mean (SD) 56.5 (1.0) days of first-line DAA, VUS1 participants took 32.5 (4.1) days of DAA and VUS2 participants took 39.8 (5.6) days of DAA.

Three participants were lost to follow-up and their first-line SVR12 status could not be otherwise ascertained: in total SVR12 status could be ascertained in 97/100 (97%) participants receiving variable duration treatment and 101/102 (99%) participants receiving fixed duration treatment. First-line SVR12 was 70% (95% CI 64, 75) overall. SVR12 in the fixed 8 week duration arm was 91% (95% CI 86, 97). In the variable duration arm, SVR12 was 48% (95% CI 39, 57) overall; however, despite the mean difference in DAAs received being only 7.3 days, there was a large difference in SVR12 between the VUS strategies with 36% (95% CI 25, 48) achieving SVR12 in the VUS1 group and 72% (95% CI 56, 87) in the VUS2 group (p=0.001 for the difference in effect of duration randomisation between the two VUS strategies).

There was no evidence of differences in first-line SVR12 rates between those who took ribavirin and those who did not (p=0.48), with SVR12 rates of 68% (95% CI 61, 76) and 72% (95% CI 65, 78), respectively. All participants who required retreatment achieved SVR12 after their retreatment. Baseline demographics and further results from all participants are detailed in the main trial paper (27).

1.5.2 The SEARCH-1 trial

SEARCH-1 was a single arm trial conducted in Vietnam testing the efficacy of shortened treatment based on on-treatment response, and recruited participants between 12th February 2019 and 4th March 2020 (Figure 1.4) (28). There were two strata within the trial: Stratum A was limited to participants with mild liver disease (Fibroscan score ≤7.1 kPa) who received either 4 weeks or 8 weeks of treatment dependent on a day 2 (D2) HCV VL of <500 or ≥500 IU/ml respectively; Stratum B was limited to cirrhotic participants (Fibroscan score ≥10.1 kPa) who received either 12 weeks or 24 weeks of treatment based on a D14 HCV VL of <500 or ≥500 IU/ml respectively. All participants received sofosbuvir/daclatasvir. In the SEARCH-1 trial protocol, treatment failure was defined identically to the STOP-HCV-1 trial. Only data from the 52 participants recruited in Stratum A were included in this thesis as both the participants (in terms of stage of disease) and treatment lengths in Stratum B were too different to those from the participants in STOP-HCV-1 (also fewer (only 41) participants were recruited in Stratum B by design limiting power for any comparisons).

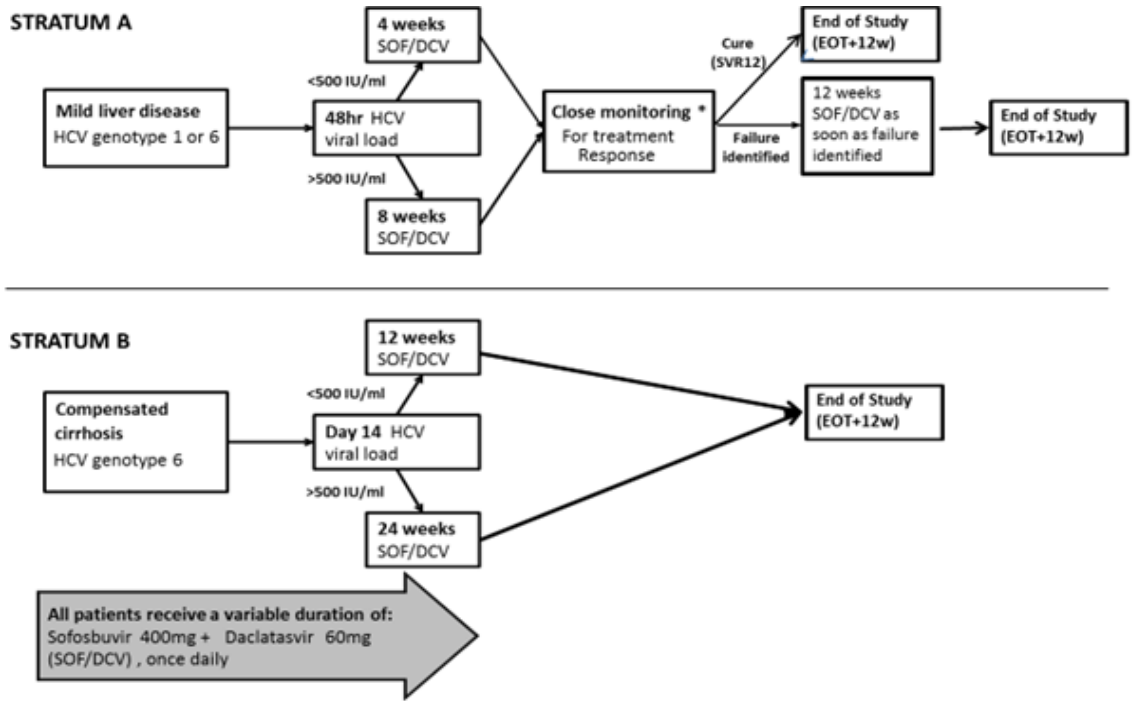


Figure 1.4: Schema of SEARCH-1

Of the 52 participants that were enrolled into Stratum A, 34 participants received 4 weeks treatment and 17 received 8 weeks. First-line SVR12 was 38/51 (75%; 95% CI 63, 86) overall, with all participants achieving cure after any necessary retreatment (28). One participant withdrew while on treatment and was excluded from analyses of the main trial and this thesis.

1.5.3 Previous work on the STOP-HCV-1 and SEARCH-1 trials

I worked as the trial statistician on both trials, liaising with the study team on a day-to-day basis and responsible for performing all statistical analyses. I worked with the data management teams to identify data queries and missing data. For SEARCH-1, I was also involved in the drafting and review of case report forms. I analysed the data for interim analyses for both trials (four for STOP-HCV-1 and one for SEARCH-1) and presented the full reports to the DMCs and, when the committees met, open sections of the report to the Trial Steering Committees. I also performed the final statistical analysis for both trials, presented the results to the wider study team, and provided the tables and figures for and reviewed the primary manuscripts.

1.6 Overview and objectives of the thesis

Although there have been attempts to determine which patients are suitable for short-course DAA treatment, many trials investigating short-course treatment are still unsuccessful (see Section 2.2 for further details on these trials). Different strategies for shortening therapy may be beneficial for different groups of patients, depending on their demographics and their

preferences for treatment, including tolerance for adjunctive drugs or need for monitoring during treatment. However, more research is needed to identify both these groups of patients and successful strategies.

The overall aim of this thesis was to investigate which patients might be suitable for short-course DAA treatment and how to trial suitable strategies for shortening or sparing HCV treatment. Specifically, my objectives were to:

- Explore potential factors that could help select patients for short-course HCV treatment based on the rates of treatment failure, the timing of treatment failure and, in those with treatment failure, how fast and how high viral rebound was.
- Examine the impact of short-course DAA treatment, ribavirin and SVR12 on quality of life to determine the acceptability of shortened treatment options for patients.
- Assess the statistical aspects of a complex clinical trial testing several strategies for shortening or sparing HCV treatment to ensure valid and reliable results.

In Chapter 2, I used the STOP-HCV-1 trial data to examine factors associated with SVR12 and used these to predict which groups of patients would have a high enough probability of cure such that shortened treatment could be a suitable strategy. Additionally, I also compared several methods of model selection to determine robustness of factor selection and probabilities of achieving SVR12.

In Chapter 3, I combined the SEARCH-1 data with the STOP-HCV-1 data to investigate factors associated with key aspects of treatment failure: specifically, the timing of when treatment failure occurred and, once it had occurred, the speed of VL rebound and peak VL rebound.

In Chapter 4, I used the STOP-HCV-1 trial data to examine the impact of short-course DAA treatment, adjunctive ribavirin and SVR12 on quality of life of HCV patients. As anaemia is a common adverse event from the use of ribavirin, I also investigated the impact of anaemia on quality of life specifically.

In Chapter 5, I examined the statistical aspects of the VIETNARMS trial, by assessing the operating characteristics of the stopping guideline, developing a plan for interim analyses, and investigating how the complex design affects the power of the trial.

In Chapter 6, I summarised and discussed my findings from the thesis.

2 Predictions of SVR12 with short-course DAA treatment: what factors are important and which patients have a high probability of cure

2.1 Introduction and aims

The overall aim of this chapter was to determine which factors are associated with SVR12 and how likely patients with those factors are to achieve SVR12, using the data collected from the STOP-HCV-1 trial which compared a variable treatment length of 4-7 weeks based on baseline VL to a fixed duration 8-week treatment.

Although short-course therapy will be important for curing those remaining with HCV who are currently difficult to treat, it will not be suitable for all patients. Identifying the patients who will benefit the most, and also the least, from short-course therapy allows it to be targeted and prevent treatment failure in patients unlikely to cure and waste resources in settings where these may be limited. Additionally, looking at the effect of different treatment options (here VUS1 (4-6 weeks) vs VUS2 (4-7 weeks)) and the addition of ribavirin can also be used to determine the relative benefits of these treatment options, primarily the increased probability of cure, which can then be weighed against the risks and costs of monitoring longer treatment or increased adverse events for individual patients.

My first aim was to determine which factors were associated with a higher probability of SVR12 in STOP-HCV-1. I first built a model including only baseline factors: this model is important as clinicians need to understand the probability of cure before treatment is initiated and is also useful where follow-up visits during or at the end of treatment are not possible. I also attempted to extend the baseline model by including on treatment and end of treatment factors. A model with these factors would be useful where a patient would be able to return during treatment and their probability of cure could be updated and treatment duration lengthened if necessary. Finally, I also examined models of baseline factors that excluded those tests that are less likely to be available in resource limited settings where there may be a greater need for short-course treatment due to costs.

My second aim was to examine the impacts of different model building processes on the selection of factors. This was performed due to the limited sample size and number of outcomes and the potentially large number of factors of interest.

My third aim was to generate probabilities of cure for specific groups of patients under different treatments based on the model selected. These could then be used to determine if specific groups of patients, based on the original model developed, would have a high enough probability of cure to attempt short-course treatment and, if so, what strategy would be

acceptable and would the addition of ribavirin be necessary to achieve sufficiently high cure rates.

2.2 Background

Predictors of cure with interferon and ribavirin have been well established. One of the main factors affecting SVR rates from treatment with these drugs was HCV genotype, with up to 80% of patients with genotypes 2 and 3 achieving SVR compared with only 50% of those with genotype 1 (12). Within genotype 1, patients who had a low baseline VL (≤600,000 IU/ml), were not black and had minimal fibrosis were more likely to achieve SVR (29). Females were also more likely to achieve SVR (30) as were younger patients who are ≤40 years old and those with a lower body weight (12, 31). Additionally, IL28B genotype SNP rs12979860 was associated with SVR, with those with the CC genotype having a higher SVR rate than those with the CT or TT genotype; due to different distributions within different ethnic groups, this may explain up to half the discrepancy in SVR between ethnicities (32). Of these factors, HCV genotype and baseline VL, as well as on treatment responses, were able to determine with reasonable accuracy who was able to cure on shorter interferon treatment as well as who needed longer treatment, and hence were used to stratify treatment duration in different populations.

For standard length DAA treatment, the presence of advanced liver fibrosis or cirrhosis is the most common indicator for reduced SVR12 rates, with almost all studies reporting it as an associated factor. Other common factors associated with reduced SVR12 include HIV coinfection and previous HCV treatment (33-35). Race has also been found to be a predictor of SVR using standard length DAAs, with black patients having a lower cure rate, but this may be due to differing rates of treatment completion (36, 37). For laboratory results, one study found that low baseline albumin and high bilirubin was associated with not achieving SVR12 (38), and two studies have found that baseline creatinine or eGFR are related to SVR rates but with associations in opposite directions (39, 40). Very few or no studies including participants receiving standard length DAA treatment have shown a difference in SVR rates based on sex, IL28B genotype or HCV genotype (other than genotype 3).

There have been a limited number of clinical trials that have tested short-course HCV treatment based on specific entry criteria: those trials that were more successful either had strict entry criteria or had minimum treatment lengths of 6 weeks. Lau *et al.* reported a 100% SVR12 after three weeks of treatment in participants with HCV genotype 1b who had no cirrhosis and with both a baseline HCV VL 10,000-10,000,000 IU/ml and HCV VL <500 IU/ml on day 2 of treatment (41). Another trial tested 4 weeks of DAA treatment in those aged 18-48

years with a BMI <30kg/m^2, mild liver disease and a baseline HCV VL ≤2,000,000 IU/ml, and reported a SVR12 rate of 75% (vs 90% with 4 weeks of DAA treatment plus PEG-IFN) (42). Two additional studies by Madsen *et al*. enrolled participants aged 18-49 years with mild liver disease and who were treatment naïve; all were given 4 weeks of treatment. In one study, overall SVR12 was 59%, but increased to 73% in those also receiving ribavirin (43); in the second overall SVR12 was 58%, but in participants with a baseline HCV VL <2,000,000 IU/ml this increased to 93% (44).

Other trials that have tested 4 week DAA treatment courses in participants with no cirrhosis and either genotype 1 or no prior HCV treatment have reported much lower SVR12 rates of 20-40% (45-48). When the treatment length was extended to 6 weeks, still shorter than the standard 8-12 weeks DAA treatment, either within the same study or in separate studies, the SVR12 rate increased to 57%-99% (46-52). Analysis examining factors associated with SVR12 within the 4 week treatment trials have suggested lower baseline HCV VL, younger age, HCV genotype (non-3 vs 3 or 1b vs 1a) and absence of baseline resistance to prescribed DAAs to be associated with higher SVR rates. Within the 6 week treatment trials, SVR12 was higher in those that did not have cirrhosis compared to those that did (47) and those who were treatment naïve (49).

2.3 Methods

As the duration of treatment related to screening viral load differed according to the randomisation in STOP-HCV-1, models were generated separately for the variable duration and the fixed duration group. All models were adjusted for ribavirin randomisation regardless of statistical significance and the variable duration model was additionally adjusted for VUS strategy (VUS1 or VUS2) as the interaction between strategy and duration randomisation was highly significant (p<0.0001) in the main analysis of STOP-HCV-1. Trial centre was not included as a factor due to the small number of participants recruited by some centres. Prescribed DAA regimen was also not included as a factor due to the very small number of participants receiving glecaprevir/pibrentasvir.

This analysis aimed to investigate associations between SVR12 and variable (VUS1/VUS2) or fixed 8 week duration treatment. Due to small numbers and perfect prediction, no participants with HCV genotype 4 could be included in the analysis (n=1 in each treatment group, both cured) and no participants with HCV genotype 1b could be included in the fixed duration analysis (n=17, all cured). Three participants in the fixed duration group were also excluded from the analysis as they took substantially less treatment than allocated (27, 32 and 40 days instead of 56). Therefore, 96 participants who received variable duration treatment and 80

who received fixed duration treatment have been included in the analysis within this chapter. All participants had mild liver disease (Fibroscan score ≤7.1kPa) due to the trial's inclusion criteria.

2.3.1 Factors included for selection in models

Initial models assessed potential baseline predictors only, including baseline demographics, resistance to prescribed DAAs, HCV VL (after $\log_{10}$ transformation) and baseline biochemistry test results, to allow for a model that could predict who would have the greatest chance of cure with short course treatment before the commencement of treatment. In total, all baseline data collected was considered in the baseline predictor model, apart from haematology test results, as these were not considered to be important factors based on previous research, and other medical conditions with low prevalence (<10%).

Biochemistry results at end of treatment (EOT) and adherence to DAAs (as reported by participants) (full list of variables in Table 2.1) were considered in a second model to assess if these improved the fit of the baseline model. Also considered was virologic response at days 3, 7, 14 and 28: relative change, absolute change and if suppressed were considered (if suppressed at day 3 was not considered due to perfect prediction).

The correlation between all potential factors was assessed using Spearman's rank correlation coefficients; factors with a high correlation (rho>0.8) were not tested together, apart from ALT and AST (rho=0.88) where potential collinearity was carefully considered.

Continuous factors were examined for outliers and those with large outliers, assessed visually, were truncated to the 1%, 5%, 95%, 99% percentile as appropriate. Non-linearity in effects of covariates was tested using fractional polynomials in the backwards elimination model; there was no evidence of non-linearity and so models estimated linear associations with continuous covariates in all models.

All previously identified factors in the literature were collected during STOP-HCV-1 and could be included in the analysis. However, adherence to DAAs was measured only by participant recall and may not be as accurate as other methods used to measure adherence within clinical trials, such as pill counts and directly observed therapy.

2.3.2 Missing data

As some *a priori* key factors had missing data (baseline resistance to DAAs, IL28B genotype, AST and the related FIB-4), these were multiply imputed separately in each duration comparison group using chained equations, with the uncertainty of the power transformation

in the fractional polynomial accounted for by the approximate Bayesian bootstrap (53). Predictive mean matching with 10 donors was used to create 25 imputations.

2.3.3 Model selection processes: overview

Due to the large number of potential factors and the small sample size and number of events, three methods were used to select models: backwards elimination, lasso and best subsets regression. The backwards elimination process starts with a 'full model' (one that contains all possible factors) and this model is refined by sequentially eliminating those factors that are the least significant. The process continues until one model remains, with all factors below the significance threshold. Lasso (least absolute shrinkage and selection operator) selects factors for a model by penalising the log-likelihood function of the model using a penalty which is a multiple (denoted λ) of the sum of the absolute values of all coefficients in the model (54). The model is selected as this penalisation shrinks the size of the coefficients of some factors to zero – these factors are then excluded from the model. The size of the model can be chosen by adjusting λ: a larger λ leads to smaller models. An adaptive lasso method uses multiple lassos, each using cross-validation; after each lasso factors with zero coefficients are discarded and factors with small coefficients are given penalty weights, which results in a smaller model than for a single cross-validation model. For best subsets regression, all potential models of a specific size are estimated and compared using a measure such as Akaike's information criteria (AIC). The best models of each size can then be compared, and an overall best model selected using the same measure.

All three methods used logistic regression and obtained estimates of the risk difference and 95% confidence intervals for the marginal effects of each factor holding other factors constant and averaging over all individuals. For the lasso model, risk differences were estimated from a standard logistic regression using the factors selected by the lasso process, although the standard errors from this method do not account for the extra variability generated by the lasso process. This extra variability is introduced as the aim of the process is to find factors that correlate with the outcome or with factors in the true model, not the true model itself. Other methods, such as cross-fit partialing-out logistic regression which uses multiple lassos (55), can be used to obtain odds ratios, however these are not easily compared to the estimates in the backwards elimination and best subsets regression models as additional control factors may be included in the lasso estimation models.

Risk differences for continuous factors were given for a 1/2 IQR unit increase. Interactions between factors were not tested for parsimony given the sample size, other than between VUS strategy and ribavirin randomisation.

2.3.4 Model selection processes: backwards elimination

Factor selection for the backwards elimination model was using Stata's mfp command. For continuous factors this was initially limited to 1 degree of freedom, which forces a linear association between the dependent and independent variables, but models with 2 degrees of freedom, non-linear but monotonic associations, were also tested where these would converge (binary factors are always chosen using 1 degree of freedom). Preliminary univariable analyses indicated that all associations with continuous factors were linear and limiting the degrees of freedom also ensured that the full multivariable model was able to achieve convergence, which did not always occur allowing higher degrees of freedom. Factors were retained in the model if $p<0.157$, which is equivalent to using AIC for factor selection. A higher threshold was used as the small sample size limited the amount of power to find appropriate factors.

The process was based initially on complete cases for all variables considered for inclusion; given the small sample size, an additional forwards selection step was used in which the factors not selected during the mfp process were then considered for inclusion in the model based on complete cases only for selected variables which had a larger sample size. A model of baseline factors was also built using multiply imputed data. The stability of the backwards elimination models using complete cases was assessed by bootstrapping the sample and model selection procedure 500 times and determining the proportion of models containing each factor.

2.3.5 Model selection processes: lasso

For the lasso models, both the cross-validation and adaptive lasso methods were used to select λ. For cross-validation, 10 folds were used. The cross-validation method was sufficient for the complete case data as it produced a model of comparable size to the backwards elimination and best subsets regression models. For the multiply imputed data, lasso was performed on the stacked multiply imputed data. Due to the size of the dataset, cross-validation did not perform as well as for the complete cases and a large model was selected. In this case, the adaptive lasso was preferred as it produced a smaller model comparable to the other two methods.

2.3.6 Model selection processes: best subsets regression

For the best subsets regression model, due to limits on computational time, models of up to size 10 were tested. For the complete cases (for all variables considered for inclusion), AIC was used to determine both the best model of each size and the best model overall. For the multiply imputed data, logistic regressions were run on the stacked dataset of all imputations. Although AIC is not a good indicator of model performance between the models when

observations are not independent, it was used to determine the best model of each size due to the lack of other suitable methods. To determine the best overall model, the Wald p-value using standard errors clustered on each participant was used. Starting from the smallest model and increasing the model size by one covariate each time, the Wald p-value of the risk difference of each new covariate that was added was estimated. The model such that adding the new covariate had a p-value>0.157 was the one selected as the best overall model.

2.3.7 Model comparison and prediction estimation

The performance and goodness-of-fit of each model selected by the different approaches above was analysed and compared using area under the ROC curves and deviance ratios, respectively. A deviance ratio is calculated using $(D_{null}-D)/D_{null}$ where D is the deviance of the model being assessed and D_{null} is the deviance when only a constant term is included in the model; it is comparable to the R^2 from a linear model. The optimism of the backwards elimination model, that is an estimate of how much the model selection procedure might lead to over-optimistic predictions of performance, was tested by bootstrapping the data 100 times and applying the selected model to each dataset to obtain an average area under the ROC curve. The optimism was then the difference between this average and the value derived from the original dataset. Patient-level predictions of the probability of achieving SVR12 from each selected model were compared using a scatter matrix and using Bland-Altman plots. Predictions were made from the final selected models of the probability of achieving cure based on different participant characteristics and durations of DAA treatment.

All analyses were performed in Stata v16.1.

2.4 Results

In this section I have first described baseline demographics and missing data in the STOP-HCV-1 trial as a whole. I then focused on assessing associations between cure and various factors for the variable DAA treatment duration group using different methods of model selection (backwards elimination logistic regression, lasso, best subsets regression), different populations (complete cases, multiply imputed data) and different groups of variables reflecting various clinical situations (baseline only, baseline plus on/end of treatment, baseline excluding high cost tests). For each model selection process, a model of baseline factors based on the complete cases and the multiply imputed data were developed. In the last section of the variable treatment duration section, I have derived probabilities of SVR12 for specific groups of patients and various treatment options from the best model. Finally, I assessed associations between cure and baseline factors for those who received fixed DAA treatment

duration, but due to the limited number of treatment failures no further work on this duration has been included.

2.4.1 Baseline demographics and missing data

Baseline demographics were broadly similar between the two treatment groups (Table 2.1). Overall, of those included in this analysis, 52 (29%) were female, median age was 46 (IQR 39, 53) years, 54 (32%) had IL28B genotype CC and 65 (37%) were coinfected with HIV. Median baseline HCV viral load was 740973 (IQR 247972, 1872136) IU/ml, 160 (90%) were infected with HCV genotype 1a, 23 (13%) had previously been unsuccessfully treated with interferon and 22 (13%) had baseline resistance to their prescribed DAA regimen.

Table 2.1: Baseline and on-treatment characteristics of the subset of participants included in analyses in this chapter

	Variable duration N=96	Variable duration: achieved SVR12 N=46 (48%)	Variable duration: failed first-line treatment N=50 (52%)	Fixed duration N=80; 74 (93%) achieved SVR
Randomised when variable strategy was VUS2	30 (31%)	22 (48%)	8 (16%)	28 (35%)
Randomised to receive ribavirin	47 (49%)	25 (54%)	22 (44%)	39 (49%)
Treated using dasabuvir/ombitasvir/paritaprevir/ritonavir	95 (99%)	45 (98%)	50 (100%)	79 (99%)
Treated using glecaprevir/pibrentasvir	1 (1%)	1 (2%)	0	1 (1%)
Baseline (pre-treatment) factors				
Age (years)	45 (38, 52)	42 (37, 48)	48 (37, 54)	47 (38, 54)
Female	28 (29%)	14 (30%)	14 (28%)	24 (30%)
BMI (kg/m^2)	25 (23, 27)	25 (23, 27)	25 (22, 26)	25 (22, 28)
Baseline HCV viral load (IU/ml)	801533 (270188, 1550000)	570000 (80000, 1148154)	936633 (575440, 1949845)	569249 (211613, 2264795)
Baseline HCV viral load (log_{10} IU/ml)	5.90 (5.43, 6.19)	5.74 (4.90, 6.06)	5.97 (5.76, 6.29)	5.76 (5.33, 6.36)
HCV genotype				
1a	80 (82%)	36 (78%)	44 (88%)	80 (100%)
1b	16 (17%)	10 (22%)	6 (12%)	0
IL28B genotype				
CC	31 (33%)	18 (41%)	13 (27%)	23 (30%)
CT or TT	62 (67%)	26 (59%)	36 (73%)	53 (70%)
Fibroscan score (kPa)	5.0 (4.2, 5.9)	5.3 (4.2, 5.9)	4.7 (4.2, 5.9)	4.8 (4.1, 5.3)
HIV coinfected	31 (32%)	14 (30%)	17 (34%)	34 (43%)
Baseline resistance to prescribed DAA regimen	10 (11%)	2 (5%)	8 (17%)	12 (16%)

	Variable duration N=96	Variable duration: achieved SVR12 N=46 (48%)	Variable duration: failed first-line treatment N=50 (52%)	Fixed duration N=80; 74 (93%) achieved SVR
Previously unsuccessfully treated with interferon	12 (13%)	2 (4%)	10 (20%)	11 (14%)
ALT (U/l)	50 (34, 90)	55 (31, 102)	47 (34, 75)	54 (34, 90)
AST (U/l)	39 (29, 57)	40 (30, 58)	38 (28, 53)	41 (31, 58)
ALP (U/l)	71 (59, 87)	72 (58, 96)	69 (61, 81)	74 (60, 91)
Albumin (g/l)	44 (42, 47)	44 (42, 47)	44 (42, 47)	45 (41, 47)
Bilirubin (µmol/l)	8 (6, 11)	9 (6, 13)	8 (6, 10)	8 (6, 12)
Creatinine (µmol/l)	74 (67, 84)	74 (67, 82)	74 (64, 86)	75 (65, 87)
FIB-4*	0.97 (0.77, 1.36)	0.87 (0.69, 1.23)	1.04 (0.85, 1.36)	1.02 (0.81, 1.51)
Current depression	22 (23%)	12 (26%)	10 (20%)	14 (18%)
Current illicit substance abuse	11 (11%)	2 (4%)	9 (18%)	14 (18%)
On-treatment factors				
HCV VL <LLOQ at D7	17 (18%)	12 (27%)	5 (10%)	20 (25%)
HCV VL <LLOQ at D14	40 (43%)	25 (57%)	15 (31%)	40 (51%)
HCV VL <LLOQ at D28	78 (84%)	41 (89%)	37 (79%)	59 (75%)
Finished treatment ≥2 days early**	4 (4%)	1 (2%)	3 (6%)	0
Missed any dose	23 (24%)	10 (22%)	13 (26%)	25 (31%)
Absolute decline in HCV VL to D3 (log_{10} IU/ml)	3.42 (3.04, 3.81)	3.23 (2.72, 3.72)	3.51 (3.24, 3.84)	3.44 (3.09, 3.84)
Relative decline in HCV VL to D3 (%)	59 (54, 65)	60 (53, 65)	58 (54, 62)	59 (51, 66)
Absolute decline in HCV VL to D7 (log_{10} IU/ml)	3.93 (3.55, 4.36)	3.74 (3.03, 4.24)	4.13 (3.66, 4.38)	3.91 (3.43, 4.48)
Relative decline in HCV VL to D7 (%)	68 (62, 74)	70 (62, 73)	67 (62, 74)	68 (60, 77)

	Variable duration N=96	Variable duration: achieved SVR12 N=46 (48%)	Variable duration: failed first-line treatment N=50 (52%)	Fixed duration N=80; 74 (93%) achieved SVR
Absolute decline in HCV VL to D14 (log_{10} IU/ml)	4.36 (3.97, 4.86)	4.22 (3.41, 4.85)	4.45 (4.13, 4.86)	4.32 (3.69, 4.77)
Relative decline in HCV VL to D14 (%)	76 (71, 80)	76 (71, 79)	75 (71, 80)	76 (66, 80)
ALT at EOT (U/l)	19 (14, 23)	20 (16, 26)	17 (14, 22)	17 (12, 20)
AST at EOT (U/l)	21 (18, 26)	21 (19, 26)	20 (17, 26)	20 (17, 24)
ALP at EOT (U/l	73 (61, 88)	78 (61, 90)	71 (61, 87)	79 (67, 99)
Albumin at EOT (g/l)	45 (42, 46)	45 (41, 46)	45 (42, 47)	45 (41, 47)
Bilirubin at EOT (µmol/l)	10 (7, 15)	11 (7, 15)	10 (6, 14)	9 (6, 15)
Creatinine at EOT (µmol/l)	75 (65, 84)	74 (65, 83)	75 (65, 85)	77 (66, 92)
FIB-4* at EOT	0.90 (0.65, 1.16)	0.83 (0.63, 1.11)	0.97 (0.74, 1.23)	0.91 (0.64, 1.31)

*Fibrosis-4 index (calculated as age*AST/(platelets*$ALT^{1/2}$) (56)

**Two participants stopped treatment 2 days early (38/40 and 39/41 days), one stopped three days early (28/31) and one stopped four days early (28/32 days).

Note: n (%) or median (interquartile range). In total, 7 participants had a missing IL28B genotype, 12 missing baseline resistance, 20 missing baseline AST and FIB-4, 9 missing D3 HCV VL, 3 missing D7 HCV VL, 4 missing D14 HCV VL, 4 missing D28 HCV VL, 1 missing EOT ALT and ALP, 19 missing EOT AST, 2 missing EOT bilirubin and albumin, 1 missing EOT creatinine, 21 missing EOT FIB-4. For the absolute and relative declines in VL, values <LLOQ are treated as equal to LLOQ (LLOQ varies by site and dilution, but is 12 or 15 IU/ml in most cases).

Of the baseline factors considered in this analysis, four had missing data: IL28B (4% missing), baseline resistance associated substitutions (RAS) (6%), baseline AST and therefore baseline FIB-4 (both 10%) (Figure 2.1). In the variable duration group, no participant had all of these factors missing, only two had both AST and either IL28B or RAS missing, and no participant had both IL28B and RAS missing (Figure 2.2). This was similar in the fixed duration group, with one participant missing both IL28B and RAS data (Figure 2.3).

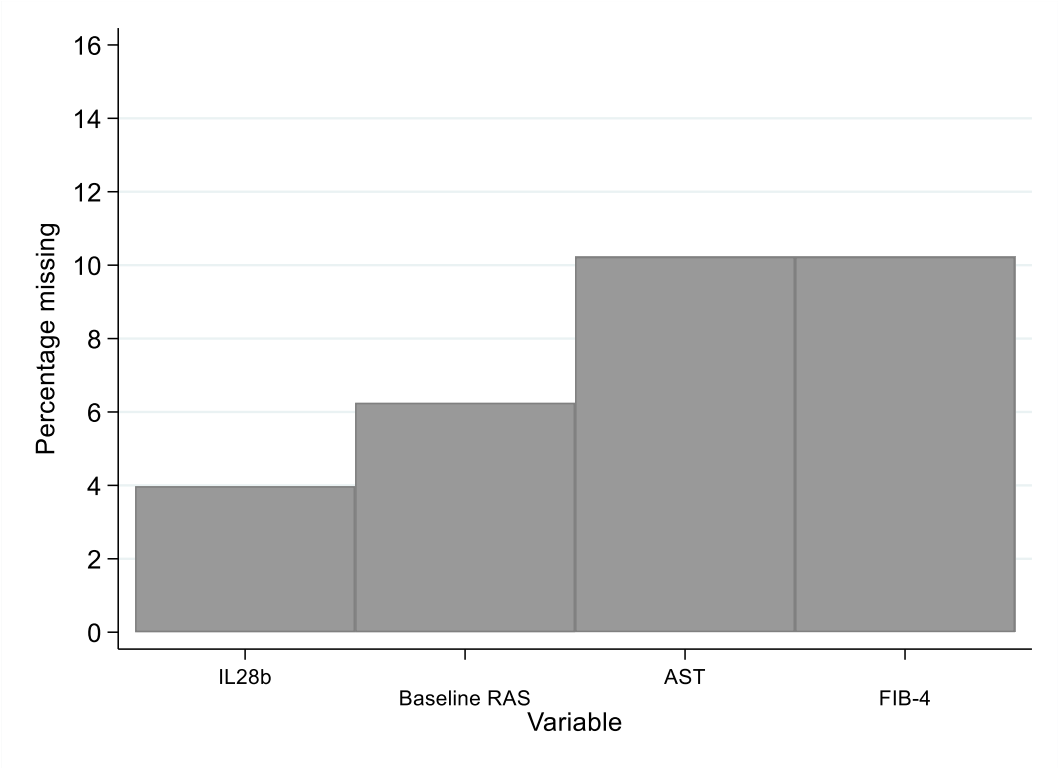


Figure 2.1: Percentage of missing data in key factors

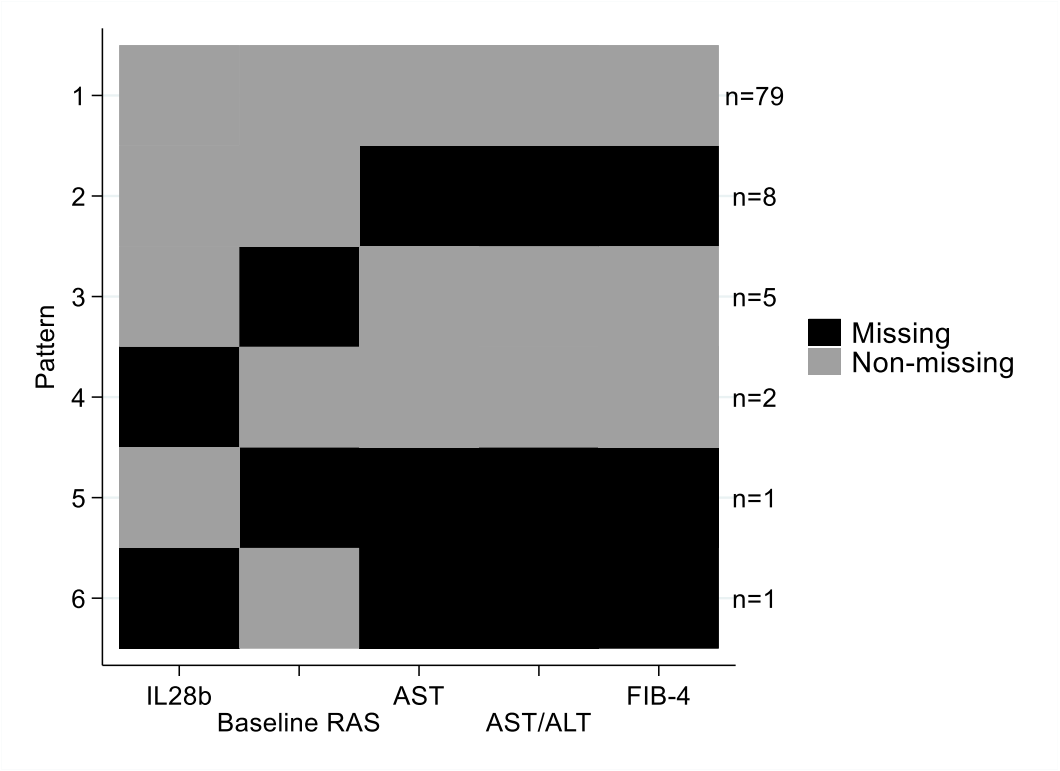


Figure 2.2: Patterns of missing data in variable duration group

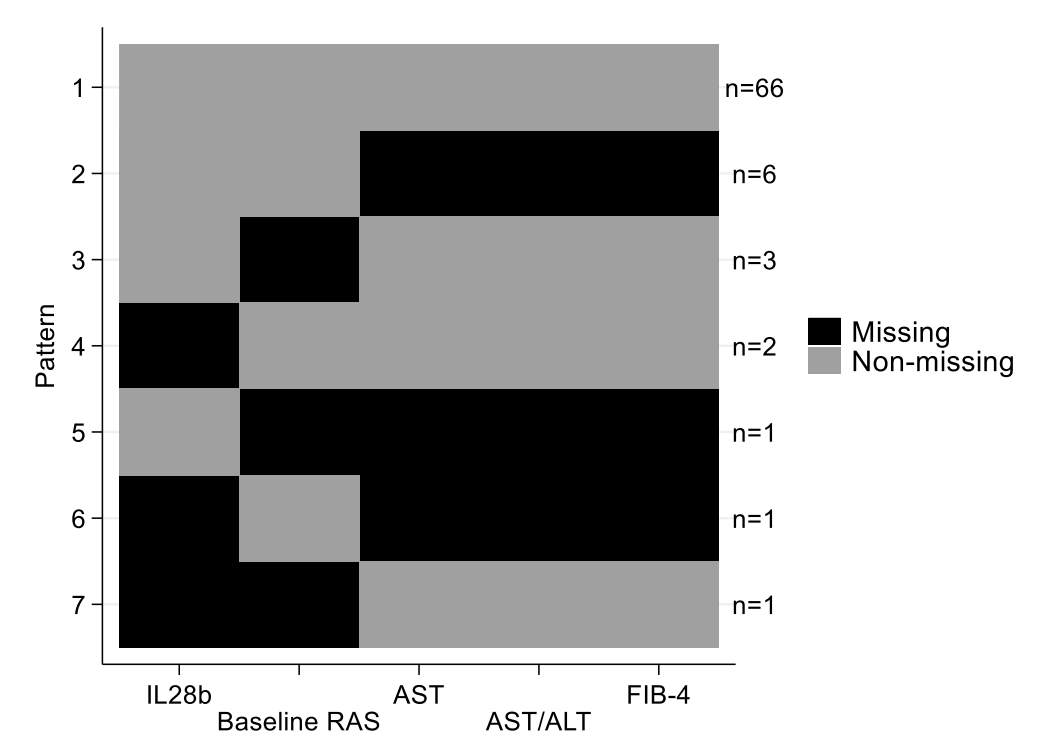


Figure 2.3: Patterns of missing data in fixed duration group

2.4.2 Associations with SVR12 using a variable DAA treatment duration: baseline factors identified using a backwards elimination process

2.4.2.1 Variable duration DAAs: baseline model restricted to complete cases

The complete case analysis included 87/96 (91%) participants. HCV genotype 1b, IL28B genotype CC, no baseline resistance to prescribed DAA, no current illicit substance abuse, low baseline HCV VL and low baseline bilirubin were independently associated with SVR12 after adjustment for ribavirin randomisation and VUS treatment strategy as well as these factors (Table 2.2). Although backwards elimination was based on AIC, all factors were significant based on the conventional level of 5%, other than IL28B genotype (p=0.09). Previous unsuccessful treatment with PEG-IFN ± ribavirin was marginally associated with lower SVR12 in the model partially adjusted for ribavirin randomisation and VUS strategy only (p=0.09), but not in the fully adjusted model (p=0.89). There was a highly significant effect of baseline HCV VL, with SVR12 11% lower (95% CI -15%, -7%; p<0.0001) per 0.4 $\log_{10}$ higher VL (half the width of the interquartile range). However, other categorical factors had much larger effects. Those infected with HCV genotype 1b had a 35% higher SVR12 (95% CI 18%, 51%; p<0.001) compared with HCV genotype 1a and those with IL28B genotype CC a 13% higher SVR12 (95% CI -2%, 18%; p=0.09) than those with CT or TT. The presence of any resistance mutation to prescribed DAAs lowered the cure rate by 37% (95% CI -57%, -17%; p<0.001). Participants who reported current illicit substance abuse at randomisation into the trial were 28% less likely (95% CI -53%, -4%; p=0.02) to achieve SVR12 than those who did not. While most factors had a large effect on cure rate, baseline bilirubin had a smaller effect and for each 2.5 µmol/l higher bilirubin (half the width of the interquartile range), SVR12 was 6% higher (95% CI 0.5%, 11%; p=0.03). Of note, this effect was about half the size of the effect of baseline HCV VL. Within these participants a higher bilirubin was associated with increased SVR12, which is unexpected, but all bilirubin results were within the normal range (<21 µmol/l) so this result may not hold with truly higher bilirubin results.

Those who received VUS2 (mean (SD) 39.8 (5.6) days DAAs) had a significantly higher cure rate than those who received VUS1 (mean (SD) 32.5 (4.1) days DAAs) with a 48% higher SVR12 (95% CI 35%, 61%; p<0.0001) after adjusting for the ribavirin randomisation and all other factors.

Although ribavirin randomisation was included in the model, there was no evidence of a difference in cure rates between those randomised to receive ribavirin and those who were not. This was the case for both the partially adjusted (adjusting only for VUS strategy) and fully adjusted models (p=0.23 and p=0.11, respectively) although there was slightly more evidence for an effect in the fully adjusted model. The fully adjusted effect was relatively modest and

equal to the partially adjusted effect (11% (95% CI -3%, 25%)). There was no evidence of interaction between ribavirin and VUS strategy (p=0.80).

Table 2.2: Variable duration model estimates using complete cases and backwards elimination to determine associated factors for SVR12 (N=87)

	Partially adjusted RD (95% CI)	p-value	Fully adjusted RD (95% CI)	p-value
Randomised to receive ribavirin	11% (-7%, 30%)	0.23	11% (-3%, 25%)	0.11
VUS strategy: VUS2 vs VUS1	37% (18%, 57%)	<0.001	48% (35%, 61%)	<0.001
HCV Genotype: 1b vs 1a	26% (4%, 49%)	0.02	35% (18%, 51%)	<0.001
IL28B Genotype: CC vs CT/TT	19% (-1%, 38%)	0.06	13% (-2%, 18%)	0.09
Baseline DAA resistance	-28% (-55%, -1%)	0.04	-37% (-57%, -17%)	<0.001
Current substance abuse	-28% (-56%, -0.1%)	0.05	-28% (-53%, -4%)	0.02
Baseline HCV viral load (per 0.4 $\log_{10}$ IU/ml higher)	-10% (-15%, -6%)	<0.001	-11% (-15%, -7%)	<0.001
Bilirubin (per 2.5 µmol/l higher)	6% (-0.3%, 12%)	0.06	6% (0.5%, 11%)	0.03
Previous unsuccessful treatment	-26% (-55%, 4%)	0.09	-	-
Age (per 6.5 years older)	-4% (-10%, 1%)	0.13	-	-
Female	9% (-12%, 29%)	0.37	-	-
BMI (per 2 kg/m^2 higher)	2% (-3%, 7%)	0.36	-	-
Fibroscan score (per 0.9 kPa higher)	3% (-5%, 11%)	0.51	-	-
HIV positive	-12% (-32%, 7%)	0.22	-	-
ALT (per 28 IU/l higher)	1% (-3%, 5%)	0.49	-	-
AST (per 14 IU/l higher)	2% (-2%, 7%)	0.31	-	-
ALP (per 14 IU/ higher)	5% (-1%, 11%)	0.13	-	-
Albumin (per 2.5 g/l higher)	3% (-4%, 9%)	0.43	-	-
Creatinine (per 9 µmol/l higher)	-3% (-9%, 4%)	0.44	-	-
FIB-4 (per 0.3 higher)	-1% (-6%, 7%)	0.83	-	-
Current depression	3% (-19%, 26%)	0.78	-	-

Note: partially adjusted models adjusted for ribavirin and VUS strategy only. All estimates are averaging over all individuals in the model. RD: risk difference.

There was no evidence that the fully adjusted model was poorly calibrated to the data (Hosmer-Lemeshow p=0.84). The model had high discrimination with an area under the ROC curve of 0.93 (95% CI 0.88, 0.98) (Figure 2.4). There was no optimism in the fully adjusted model: the median bootstrapped area under the ROC curve was 0.95 (IQR 0.87, 0.99) and the mean difference in area under the ROC curves between the fully adjusted model and the bootstrapped models was 0.01 (SD 0.03). However, this difference was highly significant (one sample t-test that the difference is zero: p<0.001).

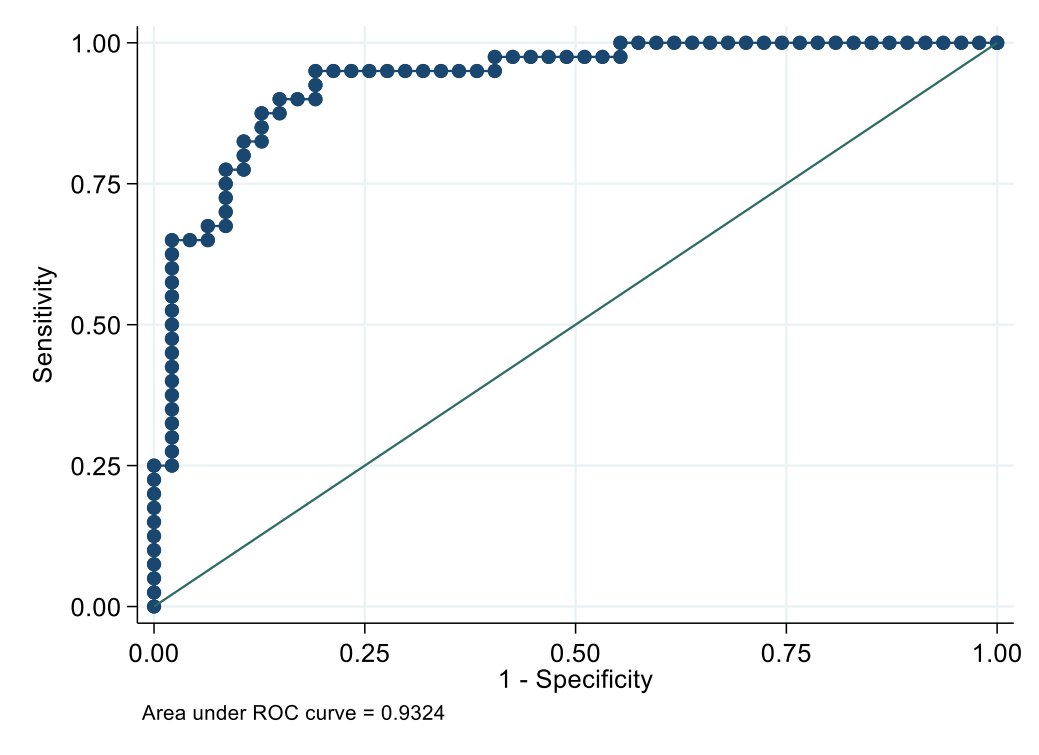


Figure 2.4: ROC curve for predictions from the variable duration model restricted to complete cases

The most common factor selected in the bootstrap models was HCV VL, which was selected in 50% of models (Figure 2.5). The two next most commonly selected factors in the bootstrap models, bilirubin (47% of bootstrap models) and current substance abuse (47%) were also selected for the baseline model on complete cases. However, the remaining factors were selected in a small number of bootstrap models: HCV genotype was selected in 9% of bootstrap models, IL28B in 16% of bootstrap models, and baseline DAA resistance in 17% of bootstrap models. Fibroscan score, ALT, albumin, creatinine and FIB-4 were all selected in ≥40% of bootstrap models, and several other factors in 30-40% of bootstrap models, but not in the baseline model suggesting that there was no clear discrimination in associated factors.

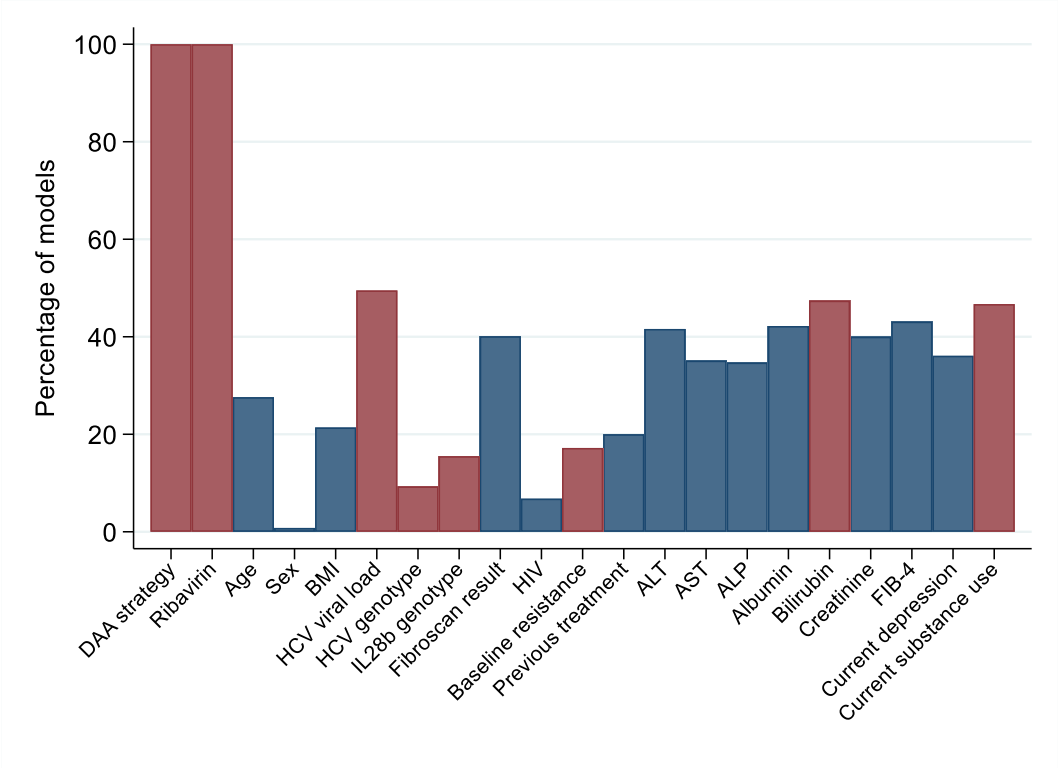


Figure 2.5: Factor selection in variable duration bootstrap models

Note: all models were required to include VUS strategy and ribavirin. Red bars denote factors selected in the baseline model, and blue bars factors considered but not selected.

2.4.2.2 Variable duration DAAs: models considering on-treatment factors in addition to the baseline model restricted to complete cases

None of the on-treatment factors tested met the criteria for inclusion into the model, with all p>0.16. The factors tested were suppression at D7, D14, and D28, relative and absolute declines in HCV VL from baseline to D3, D7 and D14, missing any dose, finishing treatment ≥2 days than allocated, and results of ALT, AST, ALP, bilirubin, albumin, creatinine and FIB-4 at EOT. The on-treatment factors with the strongest evidence for association with SVR12 were missing any DAA dose (p=0.17), and both the relative and absolute decline in HCV VL from randomisation to day 3 (both p=0.18).

2.4.2.3 Variable duration DAAs: models only considering subsets of baseline factors restricted to complete cases

Restricting the available baseline factors for selection to those that would be most readily available at the point a patient started treatment in resource limited settings did not lead to a change in factor selection in the remaining factors offered for inclusion in the model (Table 2.3). In both models, the effect of VUS2 was slightly smaller than for the original model, decreasing by 3-4%, whereas the effect of current substance abuse was stronger, increasing by

2-5%. The effect of baseline HCV VL and bilirubin was similar. In the model that included HCV genotype, the effect of being infected with genotype 1b compared to 1a was a lot weaker, decreasing by 13%. As expected, since the excluded factors were strongly associated with SVR12, the discrimination of both of the restricted models was lower than the original model, although still fairly high at 0.88 and 0.86.

Table 2.3: Variable duration model estimates based on the restricted set of baseline factors (N=96)

	Adjusted RD without considering IL28B or DAA resistance (95% CI)	**p-value**	**Adjusted RD without considering IL28B, DAA resistance or HCV genotype (95% CI)**	**p-value**
Randomised to receive ribavirin	11% (-4%, 26%)	0.15	8% (-7%, 24%)	0.30
VUS strategy: VUS2 vs VUS1	45% (32%, 59%)	<0.001	44% (29%, 58%)	<0.001
Genotype: 1b vs 1a	22% (3%, 40%)	0.03	-	-
Current substance abuse	-30% (-53%, -8%)	0.008	-33% (-54%, -12%)	0.002
Baseline HCV viral load (per 0.4 $\log_{10}$ IU/ml higher)	-11% (-15%, -7%)	<0.001	-11% (-15%, -8%)	<0.001
Baseline bilirubin (per 2.5 µmol/l higher)	7% (1%, 12%)	0.02	7% (1%, 12%)	0.01
AUROC (95% CI)	0.88 (0.81, 0.95)		0.86 (0.78, 0.93)	

RD: risk difference. AUROC: area under the ROC curve.

2.4.2.4 Variable duration DAAs: baseline model using multiply imputed data

After multiply imputing the missing data, baseline Fibroscan score was selected for the model instead of bilirubin, even though the Spearman correlation between these two factors was low (rho=0.11). Although Fibroscan score was associated with SVR12 in the fully adjusted multiply imputed model (p=0.07), it was not in the partially adjusted model (p=0.51) – however, it was selected in 40% of bootstrap models based on complete cases (Figure 2.5). The presence of HCV VL in the model strengthened the evidence of the association between Fibroscan score

and SVR12 (Fibroscan score p=0.37 when HCV VL excluded from the model; Spearman's rho between HCV VL and Fibroscan 0.07). Other selected factors were the same as for the model based on complete cases, but as two of the imputed factors were included in the model, some estimates changed slightly (Table 2.4).

Table 2.4: Variable duration model estimates using multiply imputed data and backwards elimination (N=96)

	Partially adjusted RD (95% CI)	**p-value**	**Fully adjusted RD (95% CI)**	**p-value**
Randomised to receive ribavirin	11% (-7%, 30%)	0.23	11% (-4%,25%)	0.16
VUS strategy: VUS2 vs VUS1	37% (18%, 57%)	<0.001	42% (28%, 56%)	<0.001
Genotype: 1b vs 1a	26% (4%, 49%)	0.02	28% (12%, 45%)	0.001
IL28B: CC vs CT/TT	18% (-2%, 38%)	0.07	15% (-0.1%, 30%)	0.05
Baseline resistance to DAAs	-27% (-56%, 2%)	0.06	-40% (-58%, -22%)	<0.001
Current substance abuse	-28% (-56%, -0.1%)	0.05	-26% (-53%, -0.4%)	0.05
Baseline HCV viral load (per 0.4 $\log_{10}$ IU/ml higher)	-26% (-37%, -15%)	<0.001	-11% (-15%, -1%)	<0.001
Fibroscan score (per 0.9 kPa higher)	3% (-5%, 11%)	0.51	6% (-1%, 12%)	0.07
AST (per 14 IU/l higher)	2% (-3%, 7%)	0.37	-	-
FIB-4 (per 0.3 higher)	0% (-6%, 6%)	0.99	-	-

Note: partially adjusted models adjusted for ribavirin and VUS strategy. All estimates are averaging over all over factors in the model. Only factors that have been imputed or included in the adjusted model included here, see Table 2.2 for estimates from partially adjusted models for other factors. RD: risk difference.

The greatest change in the estimates of association of the factors between the complete case and multiply imputed models was in the effect of HCV genotype; this decreased by 7% so now those with HCV genotype 1b had a 28% higher SVR12 (95% CI 12%, 45%) compared to 35% higher in the complete cases. The effect of VUS2 differed by 6% in the point estimate, decreasing from 48% in the complete case model to 42% (95% CI 28%, 56%) in the multiply imputed model. Current substance abuse, IL28B genotype and baseline resistance all had slightly different estimates (-26% vs -28%, 15% vs 13%, and -40% vs -37% respectively), while the effect of baseline HCV VL remained the same.

All factors were still significant at the p<0.05 level, but the evidence supporting effects of IL28B genotype strengthened (p=0.05 vs 0.09 in complete cases) and of current substance abuse weakened (p=0.05 vs p=0.03).

2.4.3 Variable duration DAA models: identifying factors using other methods for model selection

2.4.3.1 Variable duration models: lasso using complete cases

The model selected when using cross-validation was similar to that from backwards elimination, except ALP replaced baseline resistance to DAAs (ALP was selected in 35% of bootstrap models for complete cases vs baseline resistance in 17%, Figure 2.5). The model selected using adaptive lasso was smaller, as expected due to the process of selecting factors, and contained only HCV genotype, HCV VL and bilirubin in addition to ribavirin randomisation and VUS strategy (forced into the model by design). Table 2.5 shows the lasso coefficients from both models. Estimates from the two models were similar, apart from those for VUS strategy and bilirubin so any predictions of SVR12 from the adaptive lasso model will be affected by these factors much more than predictions from the cross-validation model.

Neither lasso model fitted the data as well as the backwards elimination model as their deviance ratios were smaller than that from the backwards elimination model (0.37 and 0.32 for the cross-validation and adaptive cross-validation models respectively vs 0.57 for the backwards elimination model on complete cases). Both models also had lower discrimination than the backwards elimination model with areas under the ROC curve of 0.89 (95% CI 0.82, 0.96) and 0.85 (95% CI 0.77, 0.93) for the cross-validation and adaptive cross-validation methods respectively.

2.4.3.2 Variable duration models: lasso using multiply imputed data

Using multiple imputation, both lasso models contained more factors than those restricted to the complete cases due to the larger sample size provided by the stacked data (Table 2.6). In addition to the factors selected with the complete cases (Table 2.5), the cross-validation method selected DAA resistance, albumin, creatinine, Fibroscan score and FIB-4; the adaptive cross-validation model additionally selected IL28B, current substance abuse and DAA resistance. The model selected by adaptive lasso is the same as that selected with backwards elimination on the multiply imputed data. Of the factors selected in both the complete cases and the multiply imputed data by lasso, the effects of VUS strategy and bilirubin vary the most – both within and between the datasets.

Table 2.5: Lasso variable duration model estimates using complete cases (N=93)

	Cross-validation			Adaptive lasso		
	Lasso coefficients	**Adjusted RD (95% CI)**	**p-value**	**Lasso coefficients**	**Adjusted RD (95% CI)**	**p-value**
Randomised to receive ribavirin	1.32	14% (-2%, 29%)	0.09	1.43	12% (-3%, 28%)	0.12
VUS strategy: VUS2 vs VUS1	4.32	46% (33%, 59%)	<0.001	6.05	47% (33%, 62%)	<0.001
Genotype: 1b vs 1a	0.47	30% (12%, 48%)	0.001	0.34	25% (6%, 44%)	0.01
IL28B: CC vs CT/TT	0.89	13% (-3%, 29%)	0.11	-	-	-
Current substance abuse	1.12	-25% (-50%, 0%)	0.052	-	-	-
Baseline HCV VL (per 0.4 $\log_{10}$ IU/ml higher)	0.68	-11% (-15%, -7%)	<0.001	0.56	-10% (-14%, 7%)	0.02
Bilirubin (per 2.5 µmol/l higher)	2.55	6% (1%, 12%)	0.03	4.67	7% (1%, 12%)	0.02
ALP (per 14 U/l)	2.64	0.2% (-6%, 6%)	0.95	-	-	-

Note: The lasso coefficients are exponentials of the original coefficients. All estimates are averaging over all over factors in the model. RD: risk difference.

Table 2.6: Lasso variable duration model estimates using multiply imputed data (N=97)

	Cross-validation			Adaptive lasso		
	Lasso coefficients	**Adjusted RD (95% CI)**	**p-value**	**Lasso coefficients**	**Adjusted RD (95% CI)**	**p-value**
Randomised to receive ribavirin	1.39	10% (-6%, 25%)	0.23	1.49	12% (-2%, 26%)	0.10
VUS strategy: VUS2 vs VUS1	2.98	45% (31%, 59%)	<0.001	3.69	44% (30%, 57%)	<0.001
Genotype: 1b vs 1a	0.66	24% (4%, 44%)	0.02	0.53	27% (10%, 44%)	0.002
IL28B: CC vs CT/TT	0.85	11% (7%, 29%)	0.22	0.76	11% (-5%, 28%)	0.17
Current substance abuse	1.24	-26% (-52%, 0%)	0.053	1.38	-27% (-53%, 2%)	0.04
Baseline DAA resistance	1.53	-41% (-59%, -22%)	<0.001	2.07	-37% (-56%, 18%)	<0.001
Baseline HCV VL (per 0.4 $\log_{10}$ IU/ml higher)	0.66	-11% (-14%, -6%)	<0.001	0.55	-11% (-14%, 7%)	<0.001
Bilirubin (per 2.5 µmol/l higher)	1.88	4% (-1%, 10%)	0.12	2.50	5% (-1%, 10%)	0.08
ALP (per 14 U/l higher)	1.50	2% (-4%, 8%)	0.51	-	-	-
Albumin (per 2.5 g/l higher)	2.57	4% (-2%, 9%)	0.20	-	-	-
Creatinine (per 9 µmol/l higher)	0.66	-4% (-9%, 2%)	0.18	-	-	-
Fibroscan score (per 0.9 kPa higher)	1.05	4% (-2%, 10%)	0.17	-	-	-
FIB-4 (per 0.3 higher)	1.01	1% (-4%, 7%)	0.60	-	-	-

Note: The lasso coefficients are exponentials of the original coefficients. All estimates are averaging over all over factors in the model. RD: risk difference. Shading shows differences with the lasso models on complete cases.

2.4.3.3 Variable duration models: best subsets regression

In the complete cases, the model with the lowest AIC overall had 7 extra factors in addition to ribavirin and VUS strategy (53.88): no model larger than this had a smaller AIC (Table 2.7). The factors in this model were HCV VL, HCV genotype, bilirubin, IL28B genotype, current substance abuse, age and current depression.

The largest model such that no model containing one extra covariate led to a statistically significant decrease in AIC (-3.84) was the model containing 5 extra covariates; the difference in AIC between the best model containing 5 covariates (55.75) and the best model containing 6 covariates (54.78) was +1.63. This model, the one on which the rest of the analysis will focus, contained HCV VL, HCV genotype, IL28B genotype, current substance abuse, and bilirubin. This model differed from the backwards elimination model only in that it did not include resistance to DAAs. Resistance to DAAs was not selected in any of the best models of any size using complete cases.

Although the model selected using best subsets regression did not contain baseline resistance to prescribed DAAs, which had a strong effect in the model selected using backwards elimination (risk difference -37%), the estimates of the other factors in the model were similar (Table 2.8). The estimates for the effect of IL28B, baseline HCV VL and bilirubin were exactly the same to the nearest percentage, though there was slightly less evidence for the association with IL28B (p=0.11 vs 0.09). The largest difference in estimates was for the effect of HCV genotype which decreased from 35% to 30% in the best subsets regression model.

Both the deviance ratio (0.45 vs 0.57) and the area under the ROC curve (0.90 (95% CI 0.84, 0.96) vs 0.93) were lower than for the model selected by backwards elimination. However, both measures were higher than for the models chosen by lasso.

The models selected by best subsets regression using the multiply imputed data are also shown in Table 2.7. For models including 1-5 factors in addition to ribavirin and VUS strategy, the Wald p-value of the new factor added to increase the size of the best model was $p<0.157$. The factor that increased the size of the model from 6 to 7 factors was current depression; its Wald p-value when added to the best-fitting model with 6 factors (which happened to be nested) was p=0.16. This suggested that the optimal model from best subsets regression for the multiply imputed data was the model containing HCV VL, HCV genotype, baseline resistance to prescribed DAAs, IL28B genotype, Fibroscan score and current substance abuse. This was the same model as created with backwards elimination on multiply imputed data (Table 2.4).

Table 2.7: Best performing models as defined by lowest AIC of each subset of models containing up to 10 extra factors

Number of factors in model	Complete cases for all factors only (N=79)		Multiply imputed data (N=97)		
	Factors in best model in addition to ribavirin and VUS strategy	AIC	Factors in best model in addition to ribavirin and VUS strategy	AIC	p-value of next variable*
1	HCV VL	80.85	HCV VL	2382.41	0.008
2	HCV VL, HCV genotype	65.95	HCV VL, HCV genotype	2220.44	<0.001
3	HCV VL, HCV genotype, bilirubin	60.91	HCV VL, HCV genotype, RAS	1994.53	0.03
4	HCV VL, HCV genotype, bilirubin, IL28B	57.38	HCV VL, HCV genotype, RAS, IL28B	1841.64	0.08
5	HCV VL, HCV genotype, bilirubin, IL28B, substance abuse	55.75	HCV VL, HCV genotype, RAS, IL28B, Fibroscan score	1715.87	0.06
6	HCV VL, HCV genotype, bilirubin, IL28B, substance abuse, age	54.78	HCV VL, HCV genotype, RAS, IL28B, Fibroscan score, substance abuse	1647.42	0.16
7	HCV VL, HCV genotype, bilirubin, IL28B, substance abuse, age, depression	53.88	HCV VL, HCV genotype, RAS, IL28B, Fibroscan score, substance abuse, depression	1580.62	0.30
8	-	-	HCV VL, HCV genotype, RAS, IL28B, Fibroscan score, substance abuse, depression, albumin	1530.30	0.15
9	-	-	HCV VL, HCV genotype, RAS, IL28B, Fibroscan score, substance abuse, depression, albumin, FIB-4	1490.57	0.16
10	-	-	HCV VL, HCV genotype, RAS, IL28B, Fibroscan score, substance abuse, depression, albumin, FIB-4, previous unsuccessful treatment	1460.74	-

*Wald p-value of the risk difference of the additional factor in the next largest model: used to discriminate between models instead of AIC as the observations not independent.
Note: In the complete case model, no model containing 8 extra factors or more improved on the model with 7 factors. RAS: resistance associated substitution (i.e. baseline resistance to DAAs). Shading indicates variables which are in one model or the other but not both models using complete cases and multiply imputed data when comparing models containing the same number of variables.

Table 2.8: Variable duration model estimates using complete cases and best subsets regression (N=93)

	Adjusted RD (95% CI)	**p-value**
Randomised to receive ribavirin	14% (-1%, 28%)	0.07
VUS strategy: VUS2 vs VUS1	46% (33%, 59%)	<0.001
Genotype: 1b vs 1a	30% (13%, 48%)	<0.001
IL28B: CC vs CT/TT	13% (-3%, 29%)	0.11
Current substance abuse	-25% (-49%, 0.1%)	0.05
Baseline HCV viral load (per 0.4 $\log_{10}$ IU/ml higher)	-11% (-15%, -7%)	<0.001
Bilirubin (per 2.5 µmol/l higher)	6% (1%, 11%)	0.03

Note: All estimates are averaging over all factors in the model. RD: risk difference.

2.4.3.4 Further comparison of models

A summary of the factors selected in each model building process for both complete cases and multiply imputed data is provided in Table 2.9. All complete case models contained HCV genotype, IL28B genotype, baseline HCV VL, baseline bilirubin and current substance abuse. Additionally, the backwards elimination model contained baseline resistance to prescribed DAAs and the lasso model contained ALP. For the multiply imputed data, all models contained baseline HCV VL, HCV genotype, IL28B genotype, baseline resistance to prescribed DAAs and current substance abuse. Fibroscan score was additionally selected in the backwards elimination and best subsets regression models and bilirubin was selected in the lasso model (Spearman's rho between Fibroscan score and bilirubin 0.11).

Table 2.9: Summary of factor selection for each model selection process

	Factors selected using complete cases			Factors selected using multiply imputed data		
	Backwards elimination	**Lasso**	**Best subsets regression**	**Backwards elimination**	**Lasso**	**Best subsets regression**
Baseline HCV VL	✓	✓	✓	✓	✓	✓
Genotype: 1b vs 1a	✓	✓	✓	✓	✓	✓
IL28B: CC vs CT/TT	✓	✓	✓	✓	✓	✓
Baseline resistance to DAAs	✓			✓	✓	✓
Current substance abuse	✓	✓	✓	✓	✓	✓
Bilirubin	✓	✓	✓		✓	
ALP		✓				
Fibroscan score				✓		✓
AUROC (95% CI)	0.93 (0.88, 0.98)	0.89 (0.82, 0.96)	0.90 (0.84, 0.96)	0.92 (0.86, 0.98)	0.91 (0.86, 0.97)	0.92 (0.86, 0.98)
Location of results	Table 2.2	Table 2.5	Table 2.8	Table 2.4	Table 2.6	Table 2.4

Note: For the complete cases, the method used for the lasso model was cross-validation. For the multiply imputed data, the method used was adaptive cross-validation. Adaptive cross-validation lasso was used for the multiply imputed data to account for the large, stacked dataset.

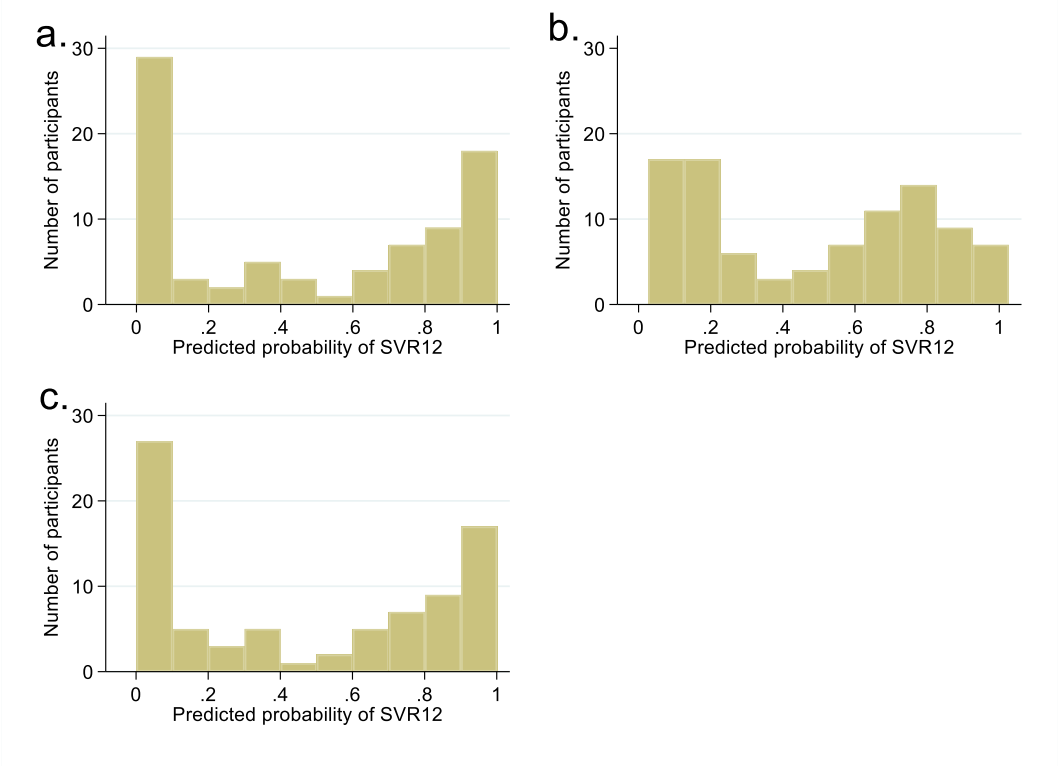


Figure 2.6: Distribution of predicted probabilities of SVR12 from each model using complete cases only (a) backwards elimination model, (b) lasso model, (c) best subsets regression model

Note: backwards elimination model is best subsets regression model plus baseline resistance to DAAs; lasso is best subsets regression model plus ALP, see Table 2.9.

In each model the predicted probabilities of cure tended to be either very high or very low (Figure 2.6). For the backwards elimination and best subset regression models, 58% and 56%, respectively, of participants had a predicted probability of SVR12 of either <10% or >90%; for the lasso model this was only 19%, however 55% had a predicted probability of either <20% or >80%.

Between the models, there was high positive correlation between their participant-level predicted probabilities of SVR12 (Figure 2.7). The Spearman rank correlation coefficient between the backwards elimination and lasso models was 0.94; between the backwards elimination and best subsets regression models 0.97; between the lasso and best subsets regression models 0.95. There were two clear discrepancies between the backwards elimination model (in which probability of SVR12<0.9) and the best subsets regression model (probability of SVR12>0.9): one of these was a participant who achieved SVR12 with HCV genotype 1a, IL28B genotype CT and reported current substance abuse; the other participant

did not achieve SVR12 despite receiving VUS2 and no baseline resistance to DAAs (other factors were as expected for their SVR12 status for both participants).

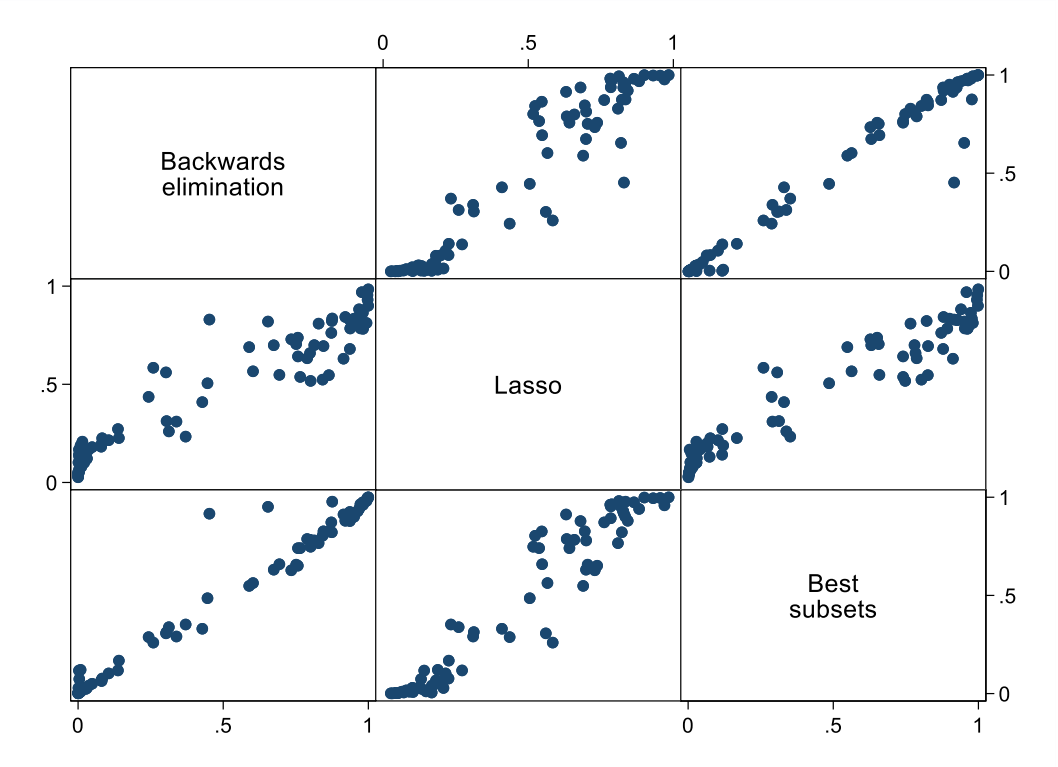


Figure 2.7: Comparison of predicted probability of SVR12 from each model using complete cases only

The predictions of SVR12 from each complete case model were also compared in Figure 2.8-Figure 2.10 using Bland-Altman plots. The observed mean difference between the predicted probabilities of SVR12 between each pair of models was 0, but there was variability between their 95% limits of agreement. The closest agreement was between the backwards elimination and best subsets regression models, with 95% limits of agreement of (-0.143, 0.144), whereas these were (-0.280, 0.284) and (-0.249, 0.246) for the comparison between the lasso model and the backwards elimination and best subset regression models respectively. All comparisons of the predictions of SVR12 show that the agreement between the models is strongest at the lowest and highest probabilities of SVR12.

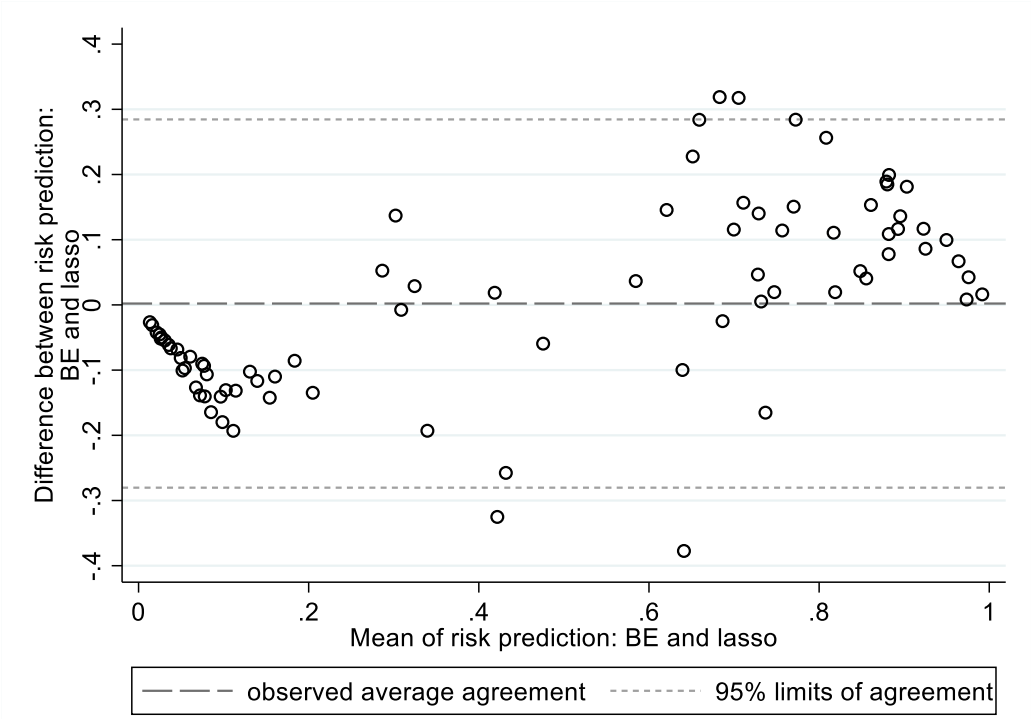


Figure 2.8: Bland-Altman plot comparing SVR12 prediction between backwards elimination and lasso models

BE: backwards elimination model.

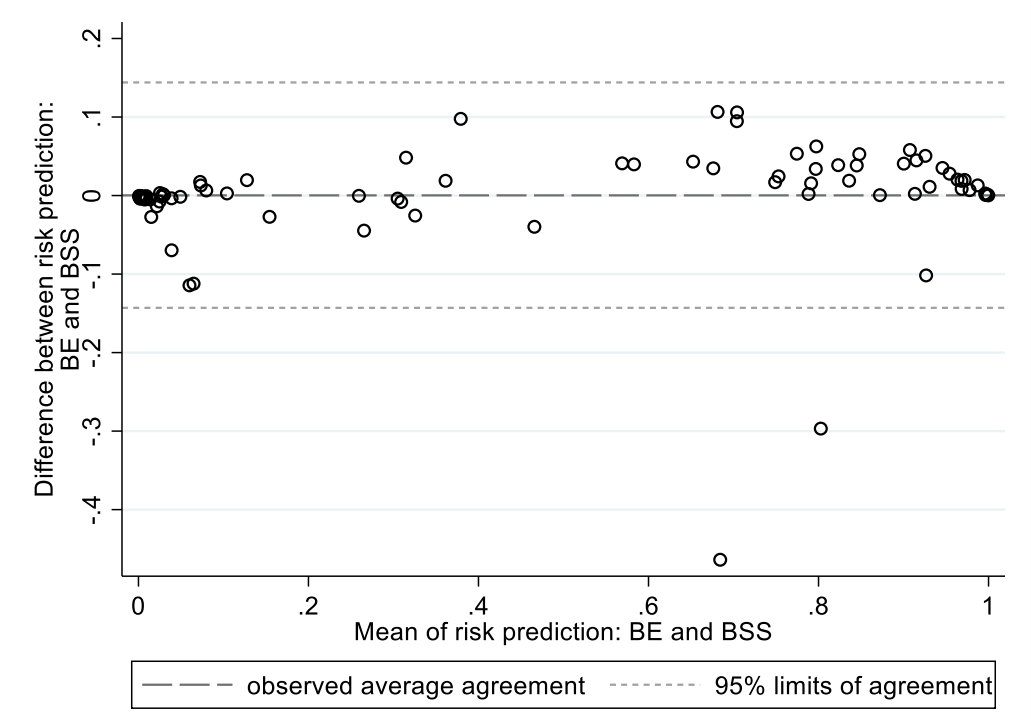


Figure 2.9: Bland-Altman plot comparing SVR12 prediction between backwards elimination and best subsets models

BE: backwards elimination model; BSS: best subsets regression model.

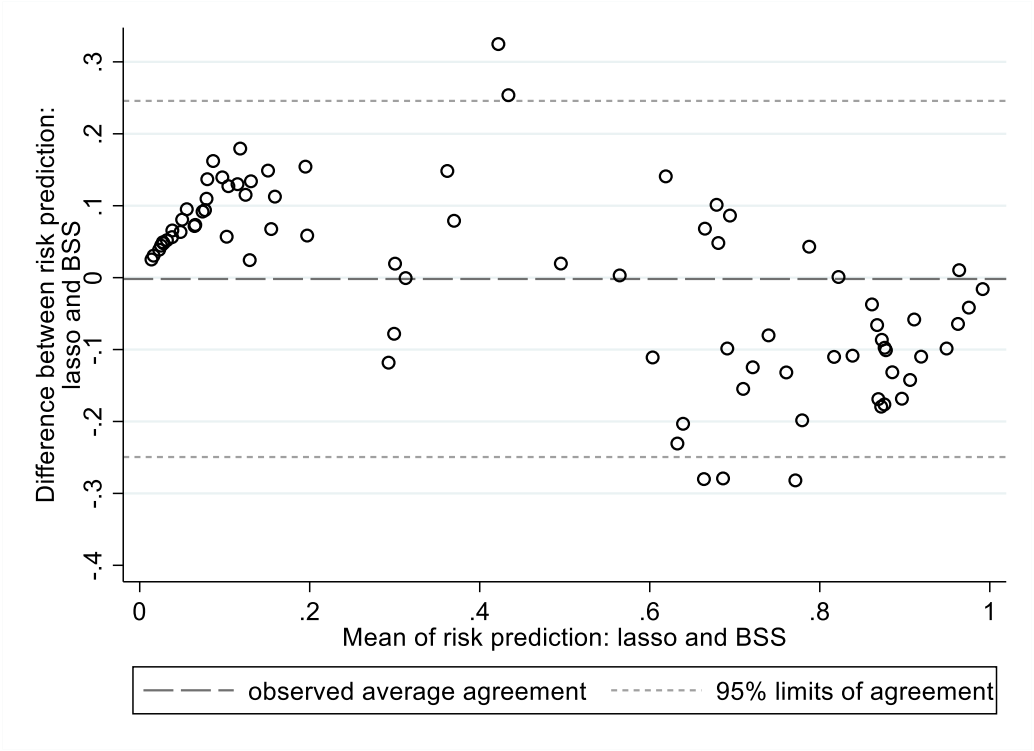


Figure 2.10: Bland-Altman plot comparing SVR12 prediction between lasso and best subset models

BSS: best subsets regression model.

2.4.4 Predictions of SVR12 using a variable DAA treatment duration from the model selected by backwards elimination

Predictions of cure were taken from the backwards elimination model, for both the complete cases and the multiply imputed data, as this model performed better than the lasso and best subsets regression models by having the highest area under the ROC curve and therefore the greatest ability to discriminate between those who did and did not achieve SVR12 under the variable duration treatment. Predicted probability of SVR12 for new patients with a range of characteristics from the complete case model are displayed in Figure 2.11-Figure 2.14 and Table 2.10-Table 2.11; predictions from the multiply imputed data model are displayed in Figure 2.15-Figure 2.18 and Table 2.12-Table 2.13. Given the relatively small magnitude of the effect of bilirubin, all predictions using the complete case data were made at the mean bilirubin for the participants in the model (9 µmol/l). For the multiply imputed data, predictions were made at the mean Fibroscan score for participants in the model (5.0 kPa).

2.4.4.1 Predictions from the complete case model

Patients infected with HCV genotype 1a, who do not have resistance to prescribed DAAs and have no current substance abuse have the biggest difference between the probabilities of cure between VUS1 and VUS2 (Figure 2.11, Figure 2.12). For those with IL28B genotype CT/TT, the

probability of cure without (and with) ribavirin decreases from 97% (99%) with screening VL of 100,000 IU/ml to 42% (66%) at 4,000,000 IU/ml for those taking VUS2 and from 37% (60%) to 1% (3%) for those taking VUS1 over the same VL range (although the decline in cure rates occurs mostly above 500,000 IU/ml in VUS1).

Patients infected with HCV genotype 1b and who have no resistance to prescribed DAAs and no current substance abuse have an extremely high prediction of cure (>95%) for all baseline HCV VL if treated with VUS2, with or without ribavirin (Figure 2.13, Figure 2.14). When treated with VUS1, the prediction of cure is similarly high for those with the lowest baseline HCV VL, but decreases quickly as HCV VL increases in those with IL28B genotype CT/TT to <77% for treatment with or without ribavirin for VLs >1,000,000 IU/ml. The probability of cure for those with genotype CC decreases more slowly as HCV VL increases, but is still <80% at 1,000,000 IU/ml if treated without ribavirin or 2,500,000 IU/ml if treated with ribavirin.

For patients infected with HCV genotype 1b who have baseline resistance to prescribed DAAs, but no current substance abuse, there is also a very large increase in predicted SVR12 rate between VUS1 and VUS2, with the largest difference in those with IL28B genotype CC. For those with IL28B genotype CC, the probability of cure without (and with) ribavirin decreases from 99% (>99%) at 100,000 IU/ml to 60% (80%) at 4,000,000 IU/ml for those taking VUS2 and from 54% (76%) to 2% (7%) for those taking VUS1 over the same VL range.

Participants with resistance to DAAs and current substance abuse have a very low probability of cure for most baseline HCV VLs, regardless of treatment strategy used, HCV genotype or IL28B genotype. For participants with HCV genotype 1b and HCV VL in the range 100,000-500,000 IU/ml, the probability of cure when using VUS2 is 64%-26% without ribavirin (83%-48% with ribavirin) in patients with IL28B genotype CT/TT and 85%-51% (94%-74%) in patients with CC.

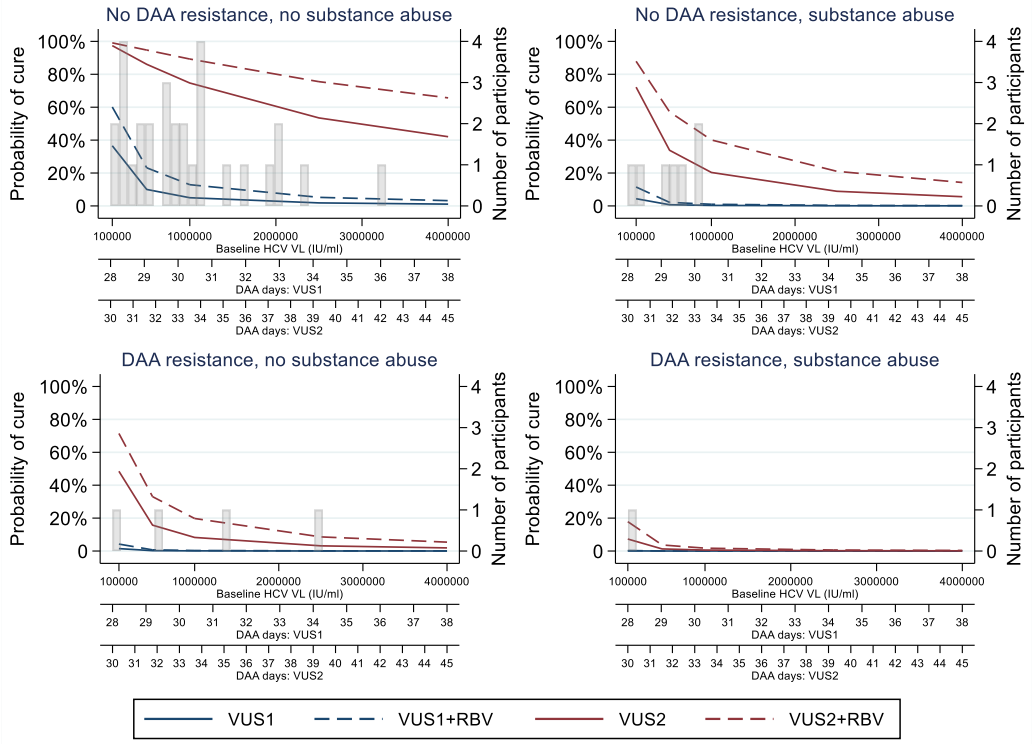


Figure 2.11: Predicted probability of SVR12 for patients with HCV genotype 1a and IL28B genotype CT/TT (complete cases)

Note: For Figure 2.11-Figure 2.18, the left y-axis is the probability of cure for a given baseline HCV VL, the right x-axis is the number of participants with that baseline HCV VL and the relevant HCV genotype, IL28B genotype, DAA resistance and current substance abuse.

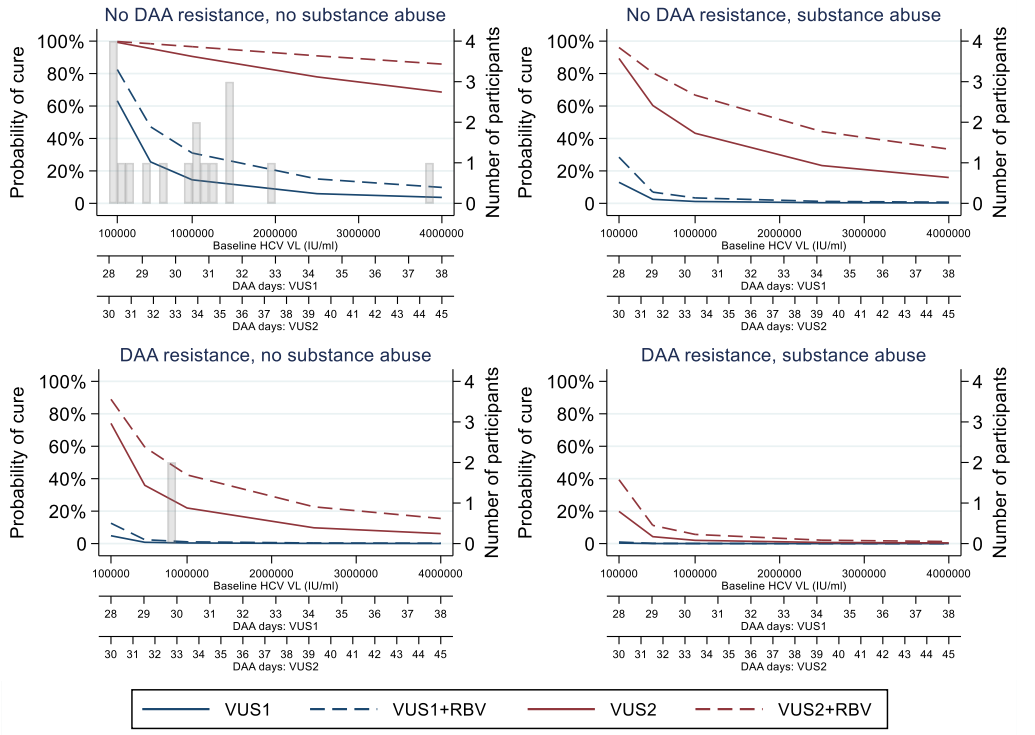


Figure 2.12: Predicted probability of SVR12 for patients with HCV genotype 1a and IL28B genotype CC (complete cases)

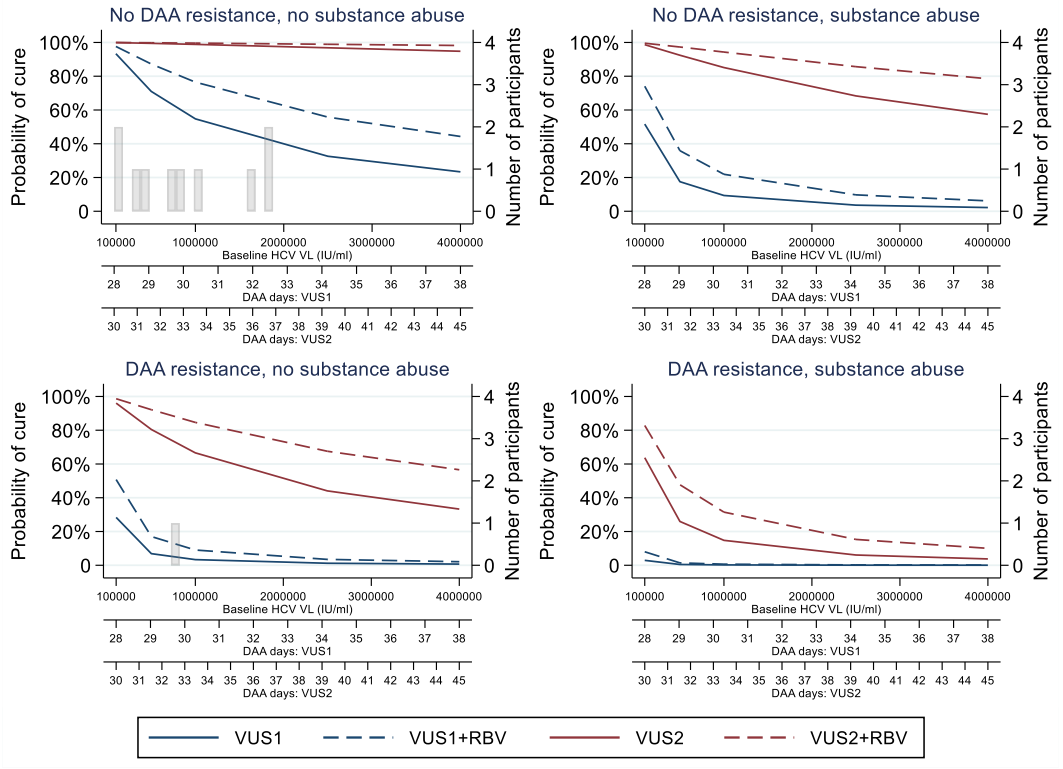


Figure 2.13: Predicted probability of SVR12 for patients with HCV genotype 1b and IL28B genotype CT/TT (complete cases)

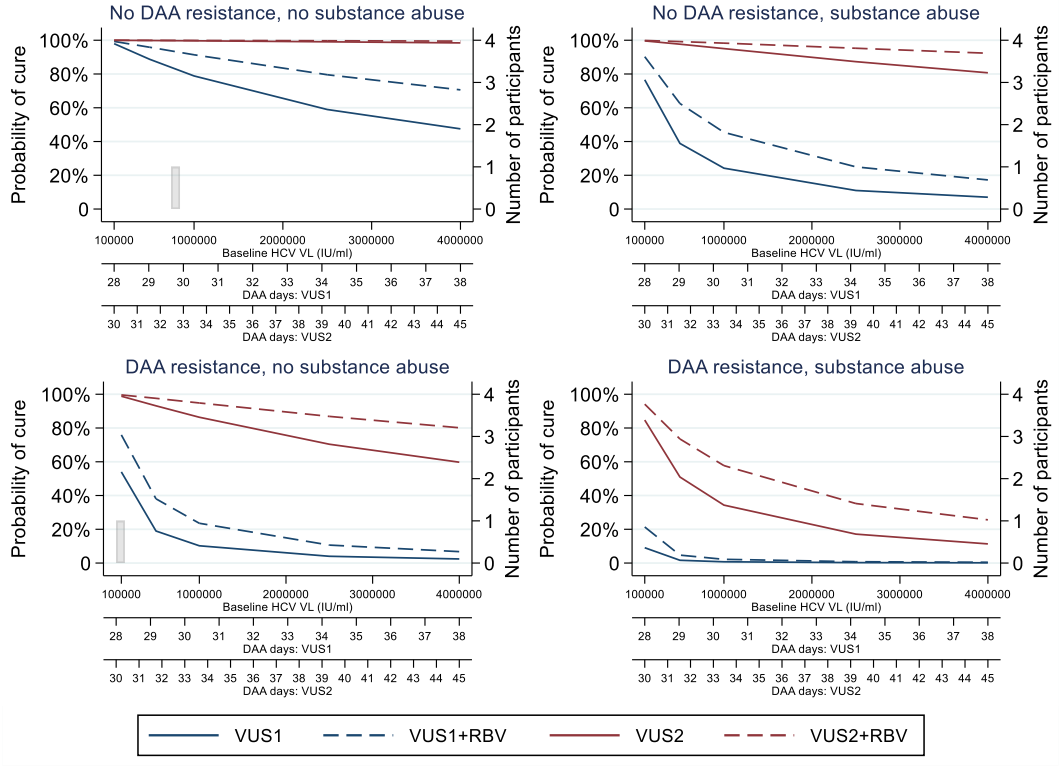


Figure 2.14: Predicted probability of SVR12 for patients with HCV genotype 1b and IL28B genotype CC (complete cases)

Table 2.10: Predicted probabilities of cure for those with HCV genotype 1a (complete cases)

Baseline HCV VL (IU/ml)	IL28B genotype	DAA resistance	Current substance abuse	VUS1 DAA days	VUS2 DAA days	Probability of cure with VUS strategy:			
						VUS1	VUS1+RBV	VUS2	VUS2+RBV
100,000	CT/TT	No	No	28	31	0.366	0.601	0.974	0.991
100,000	CT/TT	No	Yes	28	31	0.044	0.115	0.722	0.879
100,000	CT/TT	Yes	No	28	31	0.015	0.042	0.485	0.715
100,000	CT/TT	Yes	Yes	28	31	0.001	0.003	0.073	0.178
100,000	CC	No	No	28	31	0.632	0.825	0.992	0.997
100,000	CC	No	Yes	28	31	0.130	0.285	0.893	0.961
100,000	CC	Yes	No	28	31	0.048	0.126	0.741	0.890
100,000	CC	Yes	Yes	28	31	0.003	0.010	0.199	0.394
500,000	CT/TT	No	No	30	37	0.100	0.232	0.861	0.947
500,000	CT/TT	No	Yes	30	37	0.007	0.022	0.338	0.572
500,000	CT/TT	Yes	No	30	37	0.002	0.007	0.157	0.330
500,000	CT/TT	Yes	Yes	30	37	0.000	0.000	0.013	0.037
500,000	CC	No	No	30	37	0.256	0.472	0.954	0.984
500,000	CC	No	Yes	30	37	0.025	0.069	0.603	0.805
500,000	CC	Yes	No	30	37	0.008	0.024	0.359	0.595
500,000	CC	Yes	Yes	30	37	0.001	0.002	0.043	0.112
1,000,000	CT/TT	No	No	32	40	0.050	0.129	0.747	0.893
1,000,000	CT/TT	No	Yes	32	40	0.003	0.010	0.204	0.401
1,000,000	CT/TT	Yes	No	32	40	0.001	0.003	0.083	0.198
1,000,000	CT/TT	Yes	Yes	32	40	0.000	0.000	0.006	0.017
1,000,000	CC	No	No	32	40	0.145	0.310	0.906	0.966
1,000,000	CC	No	Yes	32	40	0.012	0.034	0.432	0.667

Baseline HCV VL (IU/ml)	IL28B genotype	DAA resistance	Current substance abuse	VUS1 DAA days	VUS2 DAA days	Probability of cure with VUS strategy:			
						VUS1	VUS1+RBV	VUS2	VUS2+RBV
1,000,000	CC	Yes	No	32	40	0.004	0.011	0.219	0.424
1,000,000	CC	Yes	Yes	32	40	0.000	0.001	0.020	0.057
2,500,000	CT/TT	No	No	36	44	0.018	0.052	0.535	0.755
2,500,000	CT/TT	No	Yes	36	44	0.001	0.004	0.089	0.210
2,500,000	CT/TT	Yes	No	36	44	0.000	0.001	0.032	0.086
2,500,000	CT/TT	Yes	Yes	36	44	0.000	0.000	0.002	0.006
2,500,000	CC	No	No	36	44	0.060	0.150	0.780	0.909
2,500,000	CC	No	Yes	36	44	0.004	0.012	0.233	0.442
2,500,000	CC	Yes	No	36	44	0.001	0.004	0.098	0.227
2,500,000	CC	Yes	Yes	36	44	0.000	0.000	0.007	0.021
4,000,000	CT/TT	No	No	38	45	0.011	0.032	0.421	0.657
4,000,000	CT/TT	No	Yes	38	45	0.001	0.002	0.056	0.142
4,000,000	CT/TT	Yes	No	38	45	0.000	0.001	0.019	0.054
4,000,000	CT/TT	Yes	Yes	38	45	0.000	0.000	0.001	0.004
4,000,000	CC	No	No	38	45	0.037	0.098	0.686	0.858
4,000,000	CC	No	Yes	38	45	0.002	0.007	0.160	0.334
4,000,000	CC	Yes	No	38	45	0.001	0.002	0.062	0.155
4,000,000	CC	Yes	Yes	38	45	0.000	0.000	0.004	0.013

Note: categories of factor that decrease the probability of cure have been shaded

Table 2.11: Predicted probabilities of cure for those with HCV genotype 1b (complete cases)

Baseline HCV VL (IU/ml)	IL28B genotype	DAA resistance	Current substance abuse	VUS1 DAA days	VUS2 DAA days	Probability of cure with VUS strategy:			
						VUS1	VUS1+RBV	VUS2	VUS2+RBV
100,000	CT/TT	No	No	28	31	0.933	0.977	0.999	1.000
100,000	CT/TT	No	Yes	28	31	0.517	0.741	0.987	0.996
100,000	CT/TT	Yes	No	28	31	0.284	0.508	0.961	0.987
100,000	CT/TT	Yes	Yes	28	31	0.029	0.080	0.637	0.828
100,000	CC	No	No	28	31	0.980	0.993	1.000	1.000
100,000	CC	No	Yes	28	31	0.766	0.903	0.996	0.999
100,000	CC	Yes	No	28	31	0.540	0.759	0.988	0.996
100,000	CC	Yes	Yes	28	31	0.091	0.214	0.847	0.941
500,000	CT/TT	No	No	30	37	0.710	0.872	0.995	0.998
500,000	CT/TT	No	Yes	30	37	0.175	0.359	0.924	0.973
500,000	CT/TT	Yes	No	30	37	0.069	0.170	0.804	0.922
500,000	CT/TT	Yes	Yes	30	37	0.005	0.014	0.260	0.477
500,000	CC	No	No	30	37	0.887	0.958	0.999	1.000
500,000	CC	No	Yes	30	37	0.389	0.625	0.977	0.992
500,000	CC	Yes	No	30	37	0.190	0.381	0.931	0.976
500,000	CC	Yes	Yes	30	37	0.016	0.047	0.510	0.735
1,000,000	CT/TT	No	No	32	40	0.547	0.765	0.989	0.996
1,000,000	CT/TT	No	Yes	32	40	0.094	0.219	0.851	0.943
1,000,000	CT/TT	Yes	No	32	40	0.033	0.090	0.666	0.846
1,000,000	CT/TT	Yes	Yes	32	40	0.002	0.007	0.148	0.315

Baseline HCV VL (IU/ml)	IL28B genotype	DAA resistance	Current substance abuse	VUS1 DAA days	VUS2 DAA days	Probability of cure with VUS strategy:			
						VUS1	VUS1+RBV	VUS2	VUS2+RBV
1,000,000	CC	No	No	32	40	0.789	0.914	0.997	0.999
1,000,000	CC	No	Yes	32	40	0.242	0.455	0.950	0.983
1,000,000	CC	Yes	No	32	40	0.102	0.236	0.864	0.949
1,000,000	CC	Yes	Yes	32	40	0.008	0.022	0.343	0.577
2,500,000	CT/TT	No	No	36	44	0.326	0.558	0.968	0.989
2,500,000	CT/TT	No	Yes	36	44	0.036	0.097	0.683	0.857
2,500,000	CT/TT	Yes	No	36	44	0.012	0.035	0.441	0.676
2,500,000	CT/TT	Yes	Yes	36	44	0.001	0.002	0.061	0.153
2,500,000	CC	No	No	36	44	0.590	0.796	0.991	0.997
2,500,000	CC	No	Yes	36	44	0.110	0.250	0.873	0.953
2,500,000	CC	Yes	No	36	44	0.040	0.106	0.704	0.869
2,500,000	CC	Yes	Yes	36	44	0.003	0.008	0.171	0.353
4,000,000	CT/TT	No	No	38	45	0.234	0.443	0.948	0.982
4,000,000	CT/TT	No	Yes	38	45	0.022	0.062	0.574	0.785
4,000,000	CT/TT	Yes	No	38	45	0.007	0.021	0.333	0.566
4,000,000	CT/TT	Yes	Yes	38	45	0.000	0.001	0.038	0.100
4,000,000	CC	No	No	38	45	0.475	0.706	0.984	0.995
4,000,000	CC	No	Yes	38	45	0.070	0.173	0.807	0.923
4,000,000	CC	Yes	No	38	45	0.024	0.068	0.597	0.801
4,000,000	CC	Yes	Yes	38	45	0.002	0.005	0.113	0.256

Note: categories of factor that decrease the probability of cure have been shaded

2.4.4.2 Predictions from the multiply imputed data model

Predictions from the multiply imputed model, which includes Fibroscan (5.0 kPa) rather than bilirubin, were in most cases either similar or only very slightly different than those from the complete case model. However, for those infected with HCV genotype 1b and with DAA resistance, the predicted probability of cure with VUS2 is considerably smaller in the multiply imputed model. The largest difference is in those with IL28B genotype CT/TT and no substance abuse: in the complete case model the probability of cure decreases from 96% without ribavirin (98% with ribavirin) at 100,000 IU/ml to 33% (57%) over the same VL range; in the multiply imputed data model this changes from 81% (91%) to 13% (25%).

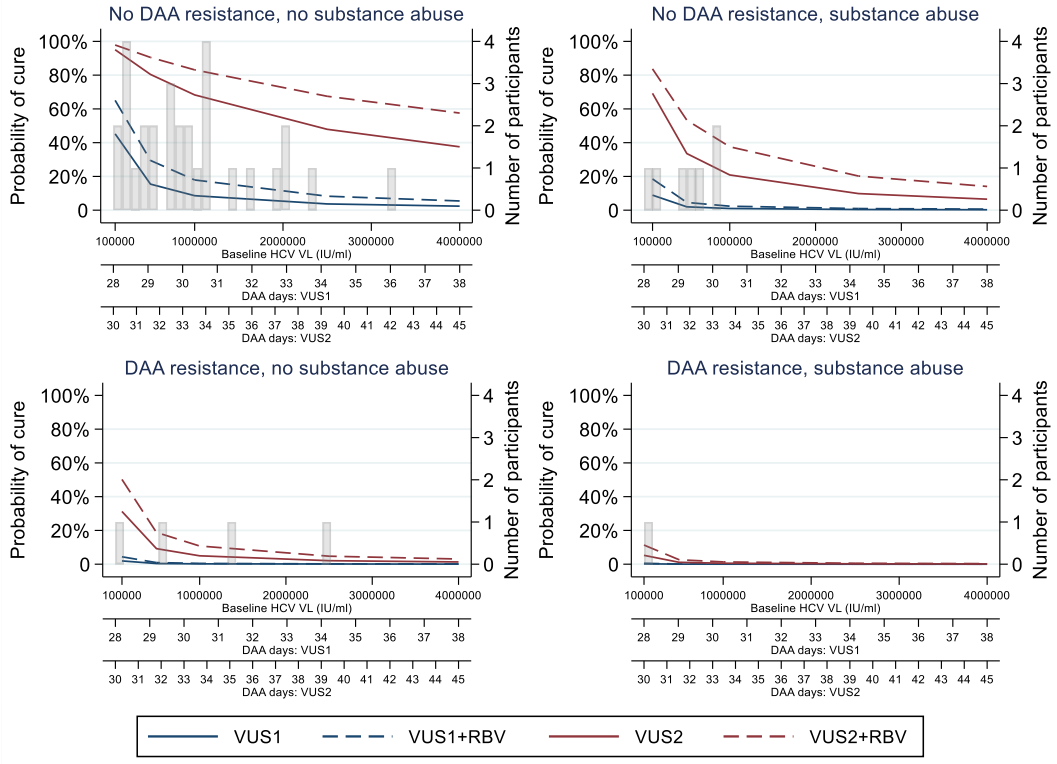


Figure 2.15: Predicted probability of SVR12 for patients with HCV genotype 1a and IL28B genotype CT/TT (multiply imputed data)

Note: For Figure 2.11-Figure 2.18, the left y-axis is the probability of cure for a given baseline HCV VL, the right x-axis is the number of participants with that baseline HCV VL and the relevant HCV genotype, IL28B genotype, DAA resistance and current substance abuse.

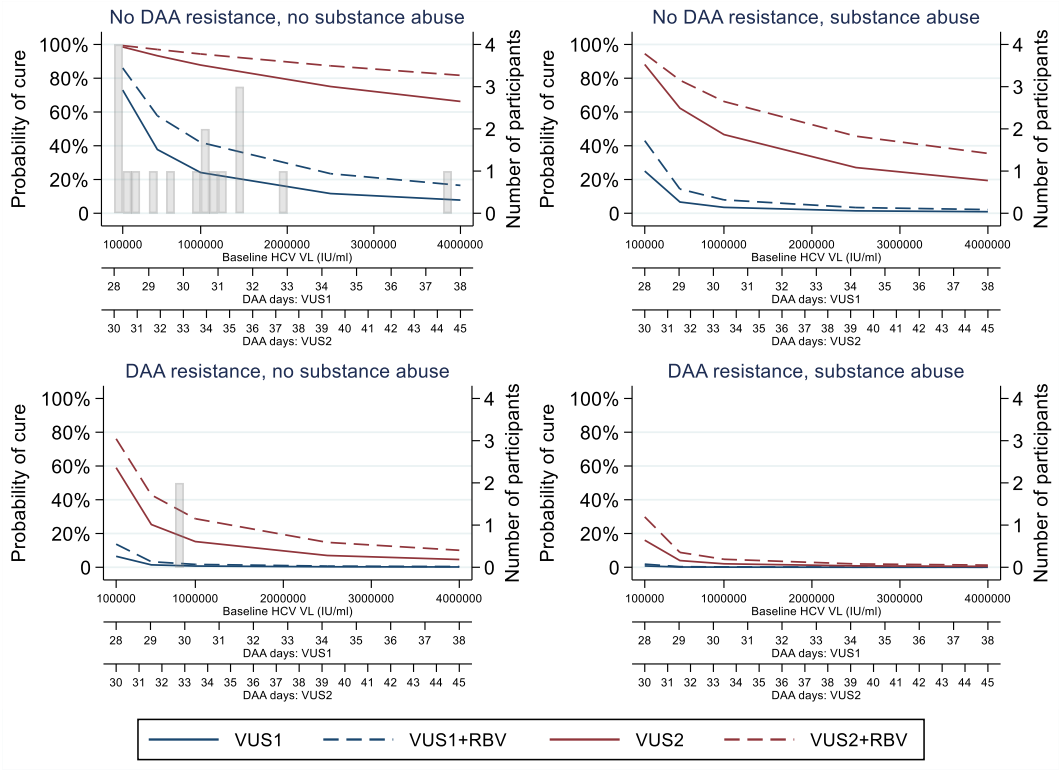


Figure 2.16: Predicted probability of SVR12 for patients with HCV genotype 1a and IL28B genotype CC (multiply imputed data)

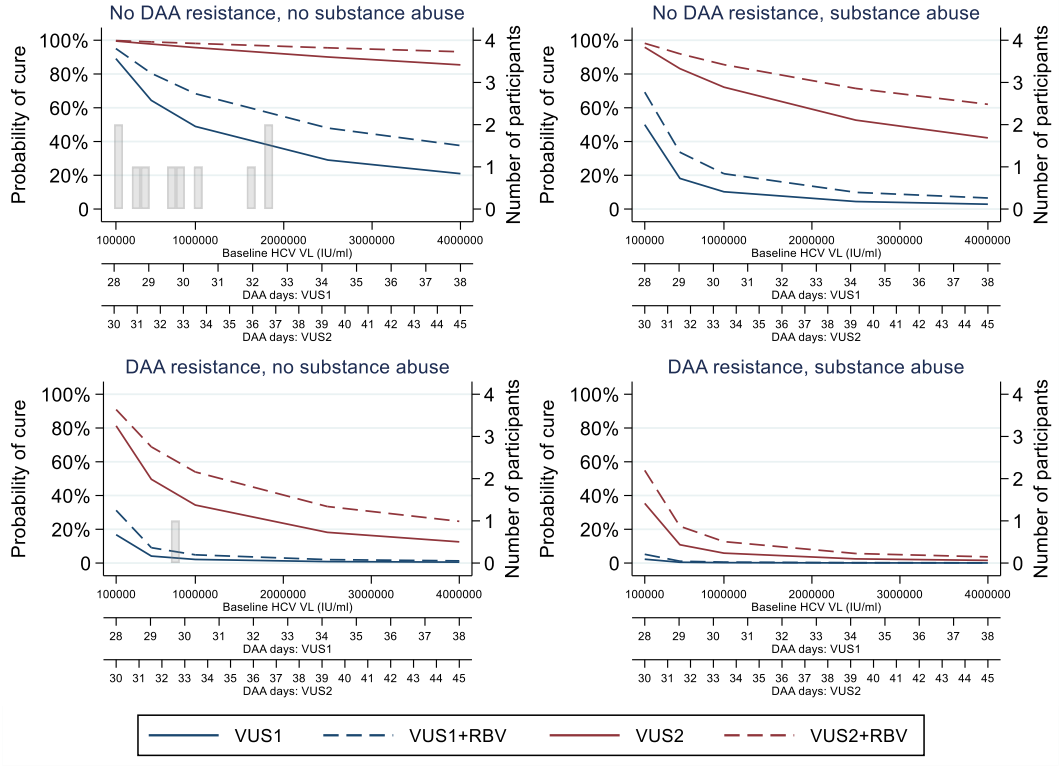


Figure 2.17: Predicted probability of SVR12 for patients with HCV genotype 1b and IL28B genotype CT/TT (multiply imputed data)

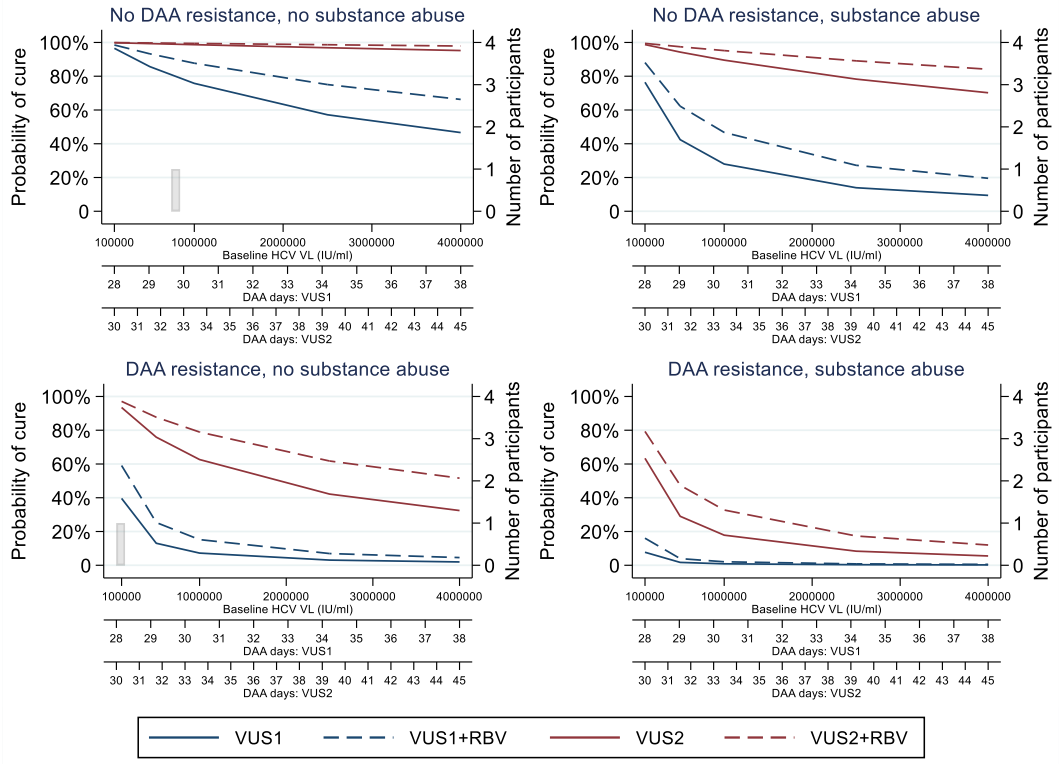


Figure 2.18: Predicted probability of SVR12 for patients with HCV genotype 1b and IL28B genotype CC (multiply imputed data)

Table 2.12: Predicted probabilities of cure for those with HCV genotype 1a (multiply imputed data)

Baseline HCV VL (IU/ml)	IL28B genotype	DAA resistance	Current substance abuse	VUS1 DAA days	VUS2 DAA days	Probability of cure with VUS strategy:			
						VUS1	VUS1+RBV	VUS2	VUS2+RBV
100,000	CT/TT	No	No	28	31	0.452	0.651	0.951	0.979
100,000	CT/TT	No	Yes	28	31	0.089	0.185	0.692	0.837
100,000	CT/TT	Yes	No	28	31	0.019	0.044	0.313	0.502
100,000	CT/TT	Yes	Yes	28	31	0.002	0.005	0.053	0.114
100,000	CC	No	No	28	31	0.730	0.861	0.985	0.994
100,000	CC	No	Yes	28	31	0.250	0.430	0.882	0.945
100,000	CC	Yes	No	28	31	0.065	0.137	0.590	0.761
100,000	CC	Yes	Yes	28	31	0.008	0.018	0.160	0.299
500,000	CT/TT	No	No	30	37	0.154	0.295	0.804	0.905
500,000	CT/TT	No	Yes	30	37	0.020	0.046	0.336	0.533
500,000	CT/TT	Yes	No	30	37	0.004	0.009	0.092	0.188
500,000	CT/TT	Yes	Yes	30	37	0.000	0.001	0.011	0.026
500,000	CC	No	No	30	37	0.378	0.578	0.933	0.970
500,000	CC	No	Yes	30	37	0.067	0.144	0.623	0.788
500,000	CC	Yes	No	30	37	0.014	0.033	0.253	0.429
500,000	CC	Yes	Yes	30	37	0.002	0.004	0.040	0.088
1,000,000	CT/TT	No	No	32	40	0.086	0.179	0.683	0.831
1,000,000	CT/TT	No	Yes	32	40	0.010	0.024	0.209	0.376
1,000,000	CT/TT	Yes	No	32	40	0.002	0.005	0.050	0.108
1,000,000	CT/TT	Yes	Yes	32	40	0.000	0.001	0.006	0.014
1,000,000	CC	No	No	32	40	0.241	0.420	0.878	0.944
1,000,000	CC	No	Yes	32	40	0.035	0.079	0.466	0.662

Baseline HCV VL (IU/ml)	IL28B genotype	DAA resistance	Current substance abuse	VUS1 DAA days	VUS2 DAA days	Probability of cure with VUS strategy:			
						VUS1	VUS1+RBV	VUS2	VUS2+RBV
1,000,000	CC	Yes	No	32	40	0.007	0.017	0.152	0.288
1,000,000	CC	Yes	Yes	32	40	0.001	0.002	0.021	0.047
2,500,000	CT/TT	No	No	36	44	0.037	0.083	0.479	0.675
2,500,000	CT/TT	No	Yes	36	44	0.004	0.010	0.098	0.202
2,500,000	CT/TT	Yes	No	36	44	0.001	0.002	0.021	0.048
2,500,000	CT/TT	Yes	Yes	36	44	0.000	0.000	0.002	0.006
2,500,000	CC	No	No	36	44	0.117	0.235	0.751	0.874
2,500,000	CC	No	Yes	36	44	0.014	0.034	0.271	0.457
2,500,000	CC	Yes	No	36	44	0.003	0.007	0.070	0.148
2,500,000	CC	Yes	Yes	36	44	0.000	0.001	0.008	0.020
4,000,000	CT/TT	No	No	38	45	0.024	0.055	0.376	0.575
4,000,000	CT/TT	No	Yes	38	45	0.003	0.006	0.065	0.141
4,000,000	CT/TT	Yes	No	38	45	0.001	0.001	0.013	0.031
4,000,000	CT/TT	Yes	Yes	38	45	0.000	0.000	0.001	0.004
4,000,000	CC	No	No	38	45	0.078	0.165	0.663	0.817
4,000,000	CC	No	Yes	38	45	0.009	0.022	0.194	0.355
4,000,000	CC	Yes	No	38	45	0.002	0.004	0.046	0.101
4,000,000	CC	Yes	Yes	38	45	0.000	0.000	0.005	0.013

Note: categories of factor that decrease the probability of cure have been shaded

Table 2.13: Predicted probabilities of cure for those with HCV genotype 1b (multiply imputed data)

Baseline HCV VL (IU/ml)	IL28B genotype	DAA resistance	Current substance abuse	VUS1 DAA days	VUS2 DAA days	Probability of cure with VUS strategy:			
						VUS1	VUS1+RBV	VUS2	VUS2+RBV
100,000	CT/TT	No	No	28	31	0.891	0.951	0.995	0.998
100,000	CT/TT	No	Yes	28	31	0.500	0.693	0.959	0.982
100,000	CT/TT	Yes	No	28	31	0.168	0.312	0.813	0.909
100,000	CT/TT	Yes	Yes	28	31	0.023	0.052	0.354	0.549
100,000	CC	No	No	28	31	0.966	0.985	0.999	0.999
100,000	CC	No	Yes	28	31	0.765	0.881	0.987	0.995
100,000	CC	Yes	No	28	31	0.396	0.591	0.935	0.971
100,000	CC	Yes	Yes	28	31	0.077	0.161	0.634	0.793
500,000	CT/TT	No	No	30	37	0.644	0.804	0.977	0.990
500,000	CT/TT	No	Yes	30	37	0.181	0.336	0.832	0.920
500,000	CT/TT	Yes	No	30	37	0.041	0.091	0.496	0.688
500,000	CT/TT	Yes	Yes	30	37	0.005	0.011	0.108	0.218
500,000	CC	No	No	30	37	0.857	0.933	0.993	0.997
500,000	CC	No	Yes	30	37	0.424	0.623	0.943	0.975
500,000	CC	Yes	No	30	37	0.130	0.253	0.759	0.877
500,000	CC	Yes	Yes	30	37	0.017	0.040	0.290	0.474
1,000,000	CT/TT	No	No	32	40	0.489	0.683	0.957	0.981
1,000,000	CT/TT	No	Yes	32	40	0.102	0.209	0.722	0.856
1,000,000	CT/TT	Yes	No	32	40	0.021	0.049	0.344	0.540
1,000,000	CT/TT	Yes	Yes	32	40	0.002	0.006	0.059	0.127
1,000,000	CC	No	No	32	40	0.758	0.877	0.987	0.994

Baseline HCV VL (IU/ml)	IL28B genotype	DAA resistance	Current substance abuse	VUS1 DAA days	VUS2 DAA days	Probability of cure with VUS strategy:			
						VUS1	VUS1+RBV	VUS2	VUS2+RBV
1,000,000	CC	No	Yes	32	40	0.280	0.467	0.895	0.952
1,000,000	CC	Yes	No	32	40	0.072	0.152	0.626	0.789
1,000,000	CC	Yes	Yes	32	40	0.009	0.021	0.179	0.327
2,500,000	CT/TT	No	No	36	44	0.290	0.480	0.901	0.955
2,500,000	CT/TT	No	Yes	36	44	0.044	0.099	0.527	0.714
2,500,000	CT/TT	Yes	No	36	44	0.009	0.020	0.182	0.335
2,500,000	CT/TT	Yes	Yes	36	44	0.001	0.002	0.025	0.056
2,500,000	CC	No	No	36	44	0.572	0.751	0.969	0.987
2,500,000	CC	No	Yes	36	44	0.140	0.272	0.783	0.891
2,500,000	CC	Yes	No	36	44	0.031	0.070	0.422	0.618
2,500,000	CC	Yes	Yes	36	44	0.004	0.008	0.084	0.174
4,000,000	CT/TT	No	No	38	45	0.210	0.376	0.854	0.932
4,000,000	CT/TT	No	Yes	38	45	0.029	0.066	0.422	0.621
4,000,000	CT/TT	Yes	No	38	45	0.006	0.013	0.126	0.247
4,000,000	CT/TT	Yes	Yes	38	45	0.001	0.001	0.016	0.037
4,000,000	CC	No	No	38	45	0.466	0.663	0.952	0.979
4,000,000	CC	No	Yes	38	45	0.094	0.196	0.702	0.842
4,000,000	CC	Yes	No	38	45	0.020	0.046	0.324	0.516
4,000,000	CC	Yes	Yes	38	45	0.002	0.005	0.055	0.120

Note: categories of factor that decrease the probability of cure have been shaded

2.4.5 Associations with SVR12 using fixed 8 week DAA treatment duration: baseline factors identified using backwards elimination

Using data from the complete cases and after adjusting for the ribavirin randomisation, no baseline factor was strongly associated with achieving SVR12 (Table 2.14). The factors which were included in the fully adjusted model (using backwards elimination based on AIC), baseline albumin and creatinine levels, had only moderate associations with achieving SVR12 after adjusting for the ribavirin randomisation only (p=0.17 for each). Univariably, there was no evidence of association between ribavirin and SVR12 even at the higher significance threshold used for backwards elimination (p=0.38).

In the adjusted model, there was very little change in the estimates compared to the unadjusted models, with the greatest difference seen in the effect of ribavirin, which reduced cure rate in those receiving 8 week's treatment, by a further 4%. However, the adjusted model did provide stronger evidence for the effect of all three factors, with albumin and creatinine either achieving or almost achieving conventional significance at the 5% level (p=0.047 and 0.057, respectively; Spearman correlation between albumin and creatinine rho=0.26). However, the effects of albumin and creatinine were small. For each 2.5g/l higher albumin, the cure rate was only 4% higher (95% CI 0.1%, 9%) and for each 9µmol/l higher creatinine, the cure rate was only 4% lower (95% CI -8%, 0.1%). There was still limited evidence for the effect of ribavirin (p=0.13).

After imputing the missing data and performing backwards elimination, the same model was selected as for the complete cases. As neither albumin nor creatinine contained missing data, the estimates shown in Table 2.14 are based on the full sample of participants and so were the same for the multiply imputed data.

The area under the ROC curve was 0.83 (Figure 2.19); however the 95% confidence interval was large (0.67-0.99), due to the low number of treatment failures in the group (6/80, 8%) and the lack of strong predictive factors in the selected model.

Only AST and FIB-4, which were not in the selected model above, were chosen in >40% of bootstrap models (Figure 2.20). Other commonly chosen factors that were not included in the selected model were BMI (39% of models), Fibroscan score (38%), sex (36%) and HIV coinfection (32%). The factors that were in the selected model, albumin and creatinine, featured in only 5% and 28% of bootstrap models respectively. Age and ALT featured in the fewest bootstrap models (1% each). HCV VL, which was the most commonly selected factor in the variable duration bootstrap models (50%) was selected in only 6% of fixed duration bootstrap models.

Table 2.14: Fixed duration model estimates using complete cases (N=80)

	Partially adjusted RD (95% CI)	p-value	Fully adjusted RD (95% CI)	p-value
Randomised to receive ribavirin	-5% (-17%, 6%)	0.38	-9% (-21%, 3%)	0.13
Albumin (per 2.5 g/l higher)	3% (-1%, 6%)	0.17	4% (0.1%, 9%)	0.047
Creatinine (per 9 µmol/l higher)	-2% (6%, 1%)	0.17	-4% (-8%, 0.1%)	0.057
Female	5% (-6%, 16%)	0.40		
IL28B Genotype: CC vs CT/TT	-1% (-15%, 12%)	0.86		
Baseline DAA resistance	-0.4% (-17%, 17%)	0.97		
HIV positive	-2% (-13%, 10%)	0.80		
Previous unsuccessful treatment	-2% (-20%, 16%)	0.83		
Baseline HCV viral load (per 0.4 $\log_{10}$ IU/ml higher)	-2% (-5%, 1%)	0.22		
Age (per 6.5 years older)	-3% (-7%, 2%)	0.25		
BMI (per 2 kg/m^2 higher)	1% (-2%, 4%)	0.55		
Fibroscan score (per 0.9 kPa higher)	2% (-3%, 8%)	0.43		
ALT (per 28 IU/l higher)	-0.4% (-3%, 2%)	0.70		
AST (per 14 IU/l higher)	-1% (-3%, 2%)	0.71		
ALP (per 14 IU/ higher l)	1% (-3%, 5%)	0.56		
Baseline bilirubin (per 2.5 µmol/l higher)	0.4% (-3%, 3%)	0.80		
FIB-4 (per 0.3 higher)	-1% (-4%, 1%)	0.27		

Note: partially adjusted models adjusted for ribavirin. All estimates are averaging over all over factors in the model. Estimates for current substance use and current depression could not be calculated due to perfect prediction.
RD: risk difference.

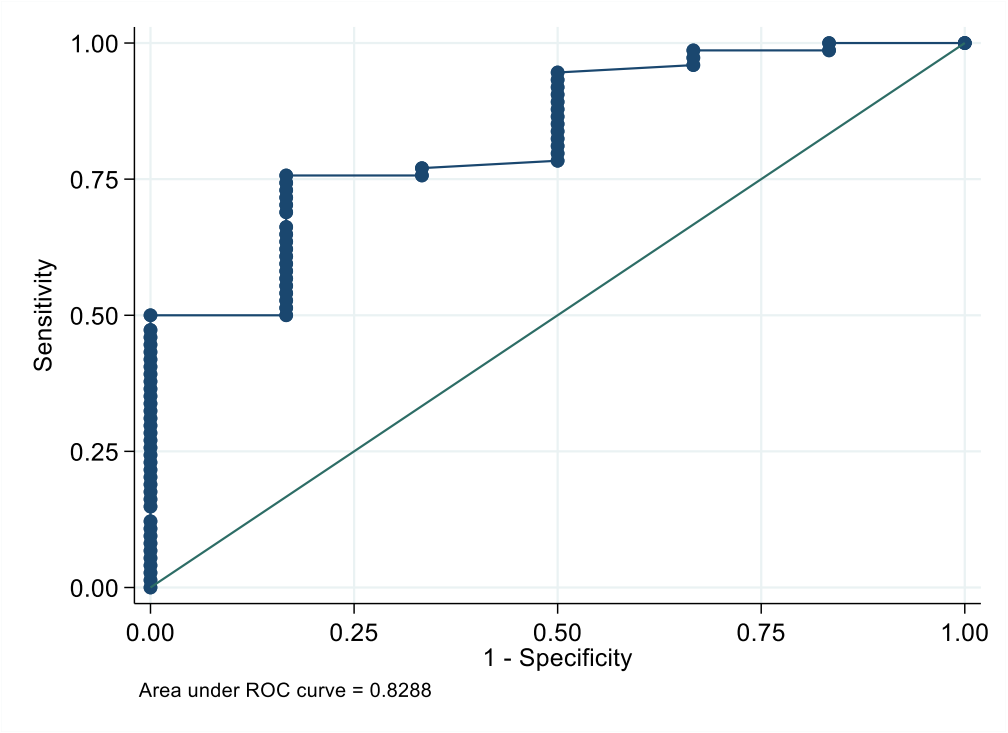


Figure 2.19: ROC curve for the fixed duration model

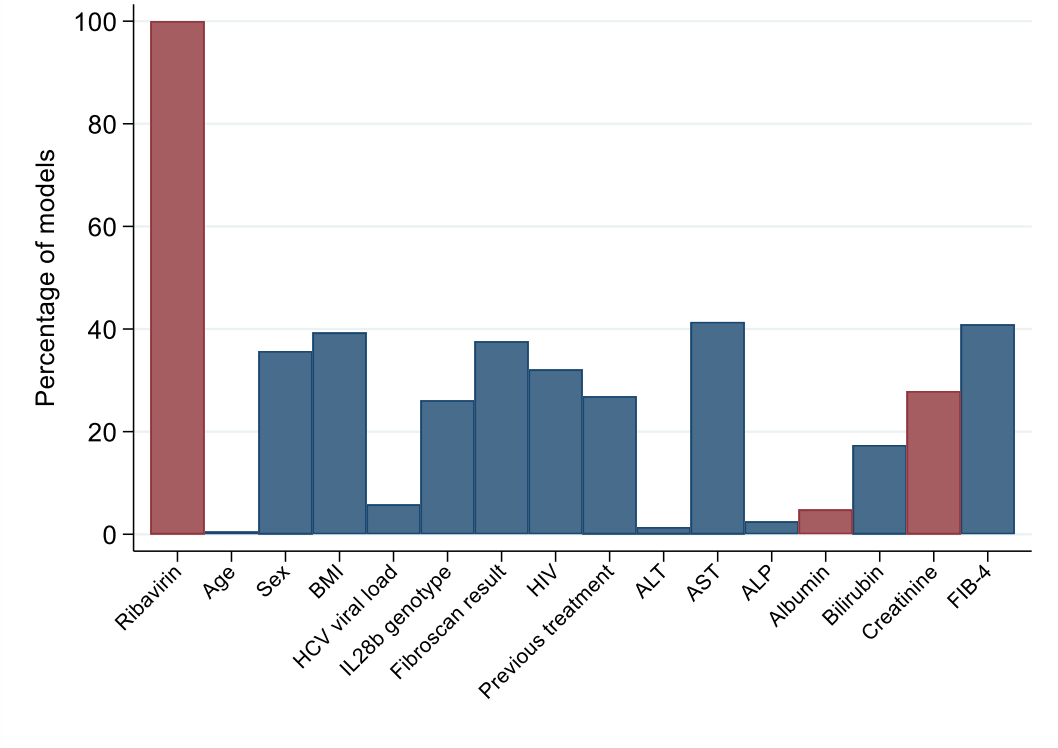


Figure 2.20: Factor selection for the fixed duration bootstrap models

Note: ribavirin was forced into all models. Red bars denote factors selected in the baseline model, and blue bars factors considered but not selected.

2.5 Discussion

In all the models of SVR12 in the variable duration group, HCV genotype 1b, IL28B genotype CC, no current substance abuse and lower baseline HCV VL were associated with higher SVR12. Additionally, no baseline resistance to prescribed DAAs, higher bilirubin, higher Fibroscan score and higher ALP were selected in some models as having a positive association with SVR12. Although the length of treatment was determined by HCV VL, it still had a strong association with and moderately large effect on SVR12 (-11% for every 0.4 $\log_{10}$ IU/ml higher) that did not vary across the models. The evidence was also very strong for the large effect of HCV genotype (27-35% higher SVR12 with 1b vs 1a) and baseline resistance to DAAs (37-40% lower SVR12, when included in the model) on SVR12. Although IL28B was included in all models, the strength of evidence of its effect varied (11-15% higher SVR12; p=0.05-0.11). In the multiply imputed data, Fibroscan result was selected instead of bilirubin; these factors were not correlated but had similar effect sizes on SVR12 and higher values of both can indicate more severe liver disease.

Predictions of cure showed that patients with no resistance to prescribed DAAs and no current substance abuse, regardless of IL28B genotype, have a very high chance of cure when treated with VUS2 at all baseline HCV VL for patients with HCV genotype 1b and for lower baseline HCV VL for patients with HCV genotype 1a. For patients with resistance to prescribed DAAs and current substance abuse, only those with HCV genotype 1b and receiving VUS2 with ribavirin have an acceptable prediction of cure (>80%). For many patients, extending the treatment length from VUS1 to VUS2 increases the prediction of cure substantially, even though the extra number of days of treatment is modest: the benefit of an extra 3-7 days of DAA treatment is most clear in those infected with HCV genotype 1a with no baseline resistance to prescribed DAAs or current substance abuse and those infected with HCV genotype 1b who may have resistance to prescribed DAAs, but who have no current substance abuse. For those patients who also have a higher baseline HCV VL, adding ribavirin to the shorter treatment regimen also increases the prediction of cure to a more acceptable rate.

For participants receiving 8-week fixed duration of therapy, higher albumin and lower creatinine levels were associated with SVR12. However, the effects were small (4% for each 2.5 g/l and 9 µmol/l increase, respectively) and the evidence for the association was not large. This model had limited ability to predict who would cure on 8-week treatment. Given the high cure rate (93%), 8 weeks of DAA treatment is likely too effective a treatment to determine strong predictors of cure as the effects of these factors have been weakened. Although the overall cure rate was high and no strong predictors were determined, it is possible that not every participant had a similarly high chance of cure. However, of the factors that were tested all

univariable effects were small (≤5% risk difference) so it is likely that for most participants, there was not a large difference in the chance of cure.

Within the studies of short-course treatment, a low baseline HCV VL (most commonly defined as <2,000,000 IU/ml) was widely identified as a factor associated with SVR12 (44-47, 50). Patients with a higher viral load have higher levels of virus circulating through their body that needs to be cleared before they can become cured and this will likely require longer treatment.

Resistance to prescribed DAAs has also been identified as a predictive factor for SVR12 with short-course treatment in some studies (45, 47), although not universally (48). Longer treatments are advised for patients with identified DAA resistance to overcome the lower effectiveness of the drugs (57) - this is supported by my analysis as higher rates of SVR12 were predicted for those with resistance taking VUS2 compared to VUS1, particularly for those with no current substance abuse.

One study has previously reported that participants with HCV genotype 1b have higher rates of SVR12 in short-course treatment (45), with other studies recruiting too few genotype 1b participants to determine a difference or not providing the breakdown of SVR12 results by subtype. One possible explanation for the difference in SVR12 is that the 1a genotype may be more likely to develop drug resistance on treatment than the 1b genotype (58). Although baseline resistance to DAAs was also selected for inclusion in the model, any drug resistance developed during the course of treatment and treatment failure could not be included, as it could only be ascertained in those who went onto fail, and therefore this effect may be represented by HCV genotype. In the trial, there was a numeric difference in the emergence of new resistance at treatment failure between the genotype subtypes: 13/50 (26%) of participants with genotype 1a had a new RAS compared to 1/6 (17%) participants with genotype 1b although this was not significant due to the small numbers (p=0.53).

Previous studies have identified the CC variant of the IL28B genotype as being associated with higher rates of SVR12 (47, 50). As well as previous research showing that the CC genotype was associated with better cure rates with PEG-IFN, it is also strongly associated with spontaneous cure of HCV (59). The mechanism of association between IL28B genotype with DAA outcomes is unclear, but the CC variant may aid in preventing rebound of HCV at lower levels of viral load, created by the DAAs, than the CT or TT variants through the same higher innate immune response to HCV that also increases the chance of spontaneous cure in some patients.

In previous studies of short-course treatment, there was no comparison of SVR12 in those reporting and not reporting current drug use, however the one study recruiting (exclusively)

participants who inject drugs did see relatively high SVR rates of 75-94% (42). Participants in the STOP-HCV-1 trial who reported current substance abuse at enrolment may have had lower adherence that was not measured accurately during the trial. While there is evidence that reduced adherence does not greatly impact SVR in full length treatment (60, 61), for short-course treatment there is evidence of a greater impact (44). It is possible that some of the participants reporting current substance abuse were participating in high-risk behaviour for reinfection, but there was no evidence that this was the cause of any treatment failure.

Lower bilirubin has previously been identified as a factor associated with SVR12 (38), which appears to contrast with the model developed in this chapter. However, this previous study dichotomised bilirubin into a high and low group using a cut off of 1.2mg/dl, whereas the highest baseline bilirubin observed in the STOP-HCV-1 trial was below this (17µmol/l or 0.99mg/dl) so all participants would have been considered as having low bilirubin levels using that study definition (and also within usual normal ranges in general). Although higher bilirubin was associated with SVR12 in my model, this may only hold for values of bilirubin within a normal range. Additionally, while higher levels of bilirubin can indicate severe liver disease and are incorporated into common prognostic scores for cirrhosis, such as the Child-Pugh score, low indirect bilirubin has also been found to be associated with advanced liver fibrosis in chronic HCV patients (62). There is also evidence that enzymes and metabolites involved in the catabolism of heme, of which bilirubin is an end product, may help with reducing viral replication and preventing liver damage and therefore a higher level of bilirubin at baseline may indicate a higher level of these protective enzymes and metabolites (63-65).

In the backwards elimination and best subsets regression models using multiply imputed data, Fibroscan score was selected in the model instead of bilirubin. Liver fibrosis, and particularly cirrhosis, is known to reduce efficacy of HCV treatment and all trials examining short-course treatment either exclude those with higher levels of fibrosis or cirrhosis completely from the trial or they are given longer treatment lengths than those with no or lower fibrosis or cirrhosis. In my analysis, a higher Fibroscan score was associated with higher rates of SVR12 and, in conjunction with the definition of mild liver disease/no or low levels of fibrosis (<7.1kPa) for inclusion within the STOP-HCV-1 trial, it is therefore more likely that in these participants the higher Fibroscan score is measuring higher levels of inflammation within the liver rather than fibrosis. Although Fibroscan score replaced bilirubin in the models using multiply imputed data and higher levels of bilirubin are often seen in patients with liver inflammation, the correlation between bilirubin and Fibroscan score was very low (spearman rho=0.11) so it is unlikely that bilirubin and Fibroscan score are representing the same mechanism of improving cure rates in each model.

A strength of the variable duration models, particularly those selected by backwards elimination and best subsets regression, is that they had high discrimination and predicted most participants to either have a very low or very high chance of cure. This indicates that the factors chosen are important and have strong effects on the probability of cure. Additionally, all model selection processes selected similar factors, with four factors being chosen in all six models selected, and predictions of SVR12 from each model selection process were also similar. This shows that factor selection was fairly robust and that the factors were not likely chosen solely by chance. The validity and generalisability of my analysis is also strengthened by finding factors that have been identified in previous research and there was no contradiction to previously identified strong predictors of SVR12.

The main limitation of this analysis is the small sample size, particularly for the fixed duration model which also had a high proportion of participants achieving SVR12. Due to the unique nature of the sliding scale mechanism for determining the variable length of treatment based on HCV VL in STOP-HCV-1 (Figure 1.3), combining STOP-HCV-1 data with other datasets that determined length of DAA treatment via another method would not have been a simple process. The small sample size means that in the predicted probabilities of SVR12 for some groups of patients and levels of HCV VL, there are very few or no STOP-HCV-1 participants that are contributing data, which may affect the validity of estimates which are essentially predicting out of sample. However, my analysis is larger than any currently published study examining short-course treatment, where the current maximum enrolled to a particular regimen is 31 participants (41, 42, 44-48). Additionally, the selected model has not been externally validated due to a lack of suitably sized datasets; the SEARCH-1 trial conducted in Vietnam, pilot to the VIETNARMS trial in Chapter 5 and included in the analysis in Chapter 3, enrolled and reported only 16 outcomes (10 cures) for participants infected with HCV genotype 1.

A limitation of the model is that, due to the assumption of no interactions, it assumes a constant increase in cure rate between VUS1 and VUS2 regardless of baseline HCV VL. This is particularly evident in the predictions of cure for patients with lower baseline HCV VL where the difference in DAA treatment days between the two strategies is either minimal or zero, but the difference in predictions is, for some groups of patients, large. As the model has not been externally validated, the accuracy of the predictions at the extremes of the VLs has not been assessed. A potential option to handle this problem in future work would be to include VUS strategy not as a binary factor (either VUS1 or VUS2), but as a factor that could take three options (VUS1, VUS2, or VUS1/VUS2 where treatment lengths are the same). I chose not to test interactions within this analysis, other than between VUS strategy and ribavirin, due to the

small sample size so key potential interactions between factors, which would have aided in prediction, may have been missed – although it is possible that even if they were considered in the model selection, they would have not been selected due to a lack of power in the analysis.

An additional limitation with the analysis is that, unlike for the effect of ribavirin, participants were not randomised to receive VUS1 or VUS2 and instead this was determined by when they were randomised into the trial. As this was not a randomised comparison, there may be unmeasured confounders that have not been accounted for within the analysis: the main confounder of concern is changes in other aspects of management over calendar time. It is therefore possible that the effects of VUS1/VUS2 may not be causal and instead be biased by this confounding.

Within this chapter I have developed a model that could be used to predict the probability of cure of a patient starting DAA treatment based on several patient characteristics with options for treatment (including ribavirin, slightly lengthening treatment from VUS1 to VUS2). To make these predictions more clinically useful, future work could involve developing an online system for clinicians to easily access and input the relevant characteristics of their patient. The clinician could then determine the probability of cure for various treatment lengths and the potential addition of ribavirin and decide the best treatment, if any, based on the patient's tolerance for treatment length or their availability for monitored treatment, such as hospital admission or incarceration, and their tolerance for potential ribavirin related adverse events.

2.6 Key findings

- SVR12 is associated with lower baseline HCV VL, HCV genotype 1b, IL28B genotype CC, absence of baseline resistance to prescribed DAAs, no current substance abuse and either higher baseline bilirubin or higher baseline Fibroscan score (higher within the normal range of results for both bilirubin and Fibroscan score) in participants taking short course variable duration therapy.
- Factor selection is largely robust to methods of model selection.
- Extending treatment from VUS1 to VUS2 is most beneficial for either:
    - Patients infected with HCV genotype 1a with no baseline resistance to prescribed DAAs and no current substance abuse, or;
    - Patients infected with HCV genotype 1b with baseline resistance to prescribed DAAs, but no current substance abuse.
- Patients with both baseline resistance to prescribed DAAs and current substance abuse have a very low probability of SVR12, regardless of baseline HCV VL, VUS

strategy or other factor associated with SVR12 and are not suitable for short-course treatment.

3 Viral load rebound kinetics and timing of treatment failure with short-course DAA treatment

3.1 Introduction and aims

The overall goal of this chapter was to explore other aspects around treatment failure, specifically rebound viral loads and the timing of treatment failure, to further develop understanding of which patients may be suitable for short-course therapy and why some patients require longer treatment. This chapter follows on from the work in Chapter 2, which looked at factors that affect SVR12 as a binary outcome in STOP-HCV-1. Here I incorporated data from the Vietnamese SEARCH-1 trial alongside the UK STOP-HCV-1 trial previously used and investigated factors that affect the time to rebound after EOT and speed of VL rebound. Factors identified using data from both trials are likely to be more generalisable to different treatment lengths and shortening strategies, as the duration of treatment as a baseline factor could be included, whereas this was not possible for STOP-HCV-1 where there was a direct link between baseline HCV VL and length of treatment.

The first specific aim of this chapter was to assess potential factors which influenced how soon after stopping treatment participants started to rebound. The second aim was to model the mean rebound VL overall, allowing for variation between participants. The third aim was to determine which factors affect rebound VL kinetics and which aspect of the rebound they impacted: either the rate of the initial rebound after VL started to increase, the peak VL during rebound or any change in VL after this peak.

3.2 Background

From the review of the literature, I found no previous studies investigating VL rebound after treatment failure with DAAs. This is likely due to the very high cure rates observed with standard duration DAA therapy which means very few patients experience a VL rebound and so data is very limited. The search was conducted both using Pubmed and Google Scholar with broad criteria: the search used the terms “hepatitis c” or “HCV” and “viral load rebound” or “viral rebound”, and was not limited to a specific treatment length, setting or time period.

Although I did not find mathematical models of viral load rebound after treatment failure, mathematical models of viral load decay have been well described and may provide insight into rebound kinetics. The standard model of viral load decline during treatment is a biphasic model (66). HCV RNA declines rapidly during the first phase as the virus is cleared from circulation and not replaced as virion production and release from infected cells is blocked. The second phase is slower and is caused by the death of infected cells; infection of new cells is limited due to the lower levels of circulating virus. It is theorised that with DAAs, the

intracellular levels of virus decrease such that infected cells become cured, further reducing the levels of RNA being released, and leading to a faster second phase decline than with PEG-IFN (67). In patients where the majority of hepatocytes are infected, viral load decay follows a triphasic model in which the first phase remains the same as the biphasic model and the second phase becomes the third phase with an introduction of an intermediate "shoulder phase" (68). In the "shoulder phase", viral load declines very slowly or remains constant due to the rate of proliferation of infected cells remaining high due to their higher prevalence. The differences in the two declines are displayed in Figure 3.1.

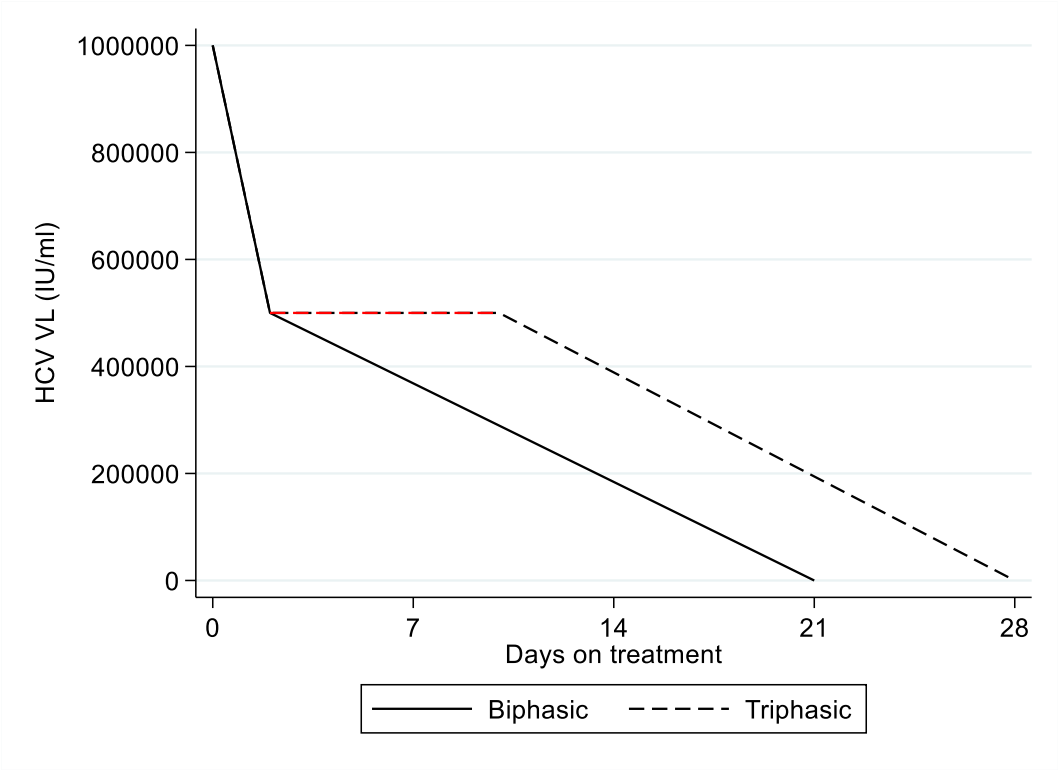


Figure 3.1: Biphasic and triphasic declines of HCV viral load

Note: red dashed line is the shoulder phase of the triphasic decline. Data is for illustrative purposes only and not taken from real measurements.

A small number of studies have attempted to determine what factors affect viral kinetics with DAA treatment. In one study, participants who took ribavirin with DAAs had faster declines in VL from baseline to day 1, and two weeks after starting treatment had lower RNA levels and were more likely to have undetectable virus (69). Of note, in STOP-HCV-1, I did not find evidence of similar effects (p=0.89, 0.84 and 0.25 comparing $\log_{10}$ HCV VL between those randomised to ribavirin vs no ribavirin adjusted for baseline at days 3, 7, and 14, respectively; unpublished findings). In another study that altered treatment length based on a predicted time to cure determined from the standard biphasic model, those with lower baseline HCV VL

and ALT had faster VL declines from baseline to day 28, a shorter predicted time to cure and were therefore more likely to have shorter DAA treatment (70). However, other studies have found no evidence of difference in kinetics across a large number of factors including age, sex, ethnicity, HCV genotype, coinfection with HIV/taking ART, fibrosis score/cirrhosis, previous HCV treatment and in one case resistance to DAAs (71-73). Previous research with PEG-IFN treatment suggested that IL28B genotype CC was associated with a faster first phase of VL decay, and that the second phase was faster in those with lower baseline HCV VL and higher baseline ALT (66, 74, 75). There is also evidence that ribavirin in conjunction with PEG-IFN increased viral load decay in the second phase of VL decline using the standard model and the third phase in a triphasic model (76, 77).

3.3 Methods

As for the analysis in Chapter 2, three participants with very poor adherence and two participants with HCV genotype 4 from the STOP-HCV-1 trial were excluded from the analyses in this chapter. Two participants in STOP-HCV-1 with SVR12 but who failed later (at the final trial follow-up 24 weeks after EOT) have been excluded from the analysis of kinetics of VL rebound, but included in the time to failure analysis, considered cured and censored at their EOT+12 visit. All SEARCH-1 participants with a SVR12 result have been included in the analyses. All analyses were performed in Stata v17.0.

3.3.1 Statistical analysis for time to treatment failure

As the exact time of treatment failure was unknown, interval censored Cox models were used to estimate the hazard of treatment failure. Time was measured from EOT due to the varying treatment lengths and because failure while receiving treatment occurred in only one participant in STOP-HCV-1, who was therefore excluded from the analysis. Participants who achieved SVR12 were also included in this analysis, right censored at their EOT+12 visit (the final post EOT assessment for those achieving cure in SEARCH-1). Those who did not achieve SVR12 contributed an interval censored failure, with the failure time occurring sometime between their last prior VL measurement up to their first VL measurement meeting the definition of failure (given in Section 1.5.1). A multivariable model was built using backwards elimination, with exit p=0.05. A threshold of p=0.05 was used instead of p=0.156 as used in Chapter 2 due to the larger sample and the increased power in the model. Only baseline (pre-treatment) factors were considered in this analysis, but baseline laboratory results were included as in Chapter 2, given the larger number of participants. EOT factors were not tested in this model to align with the models in Chapter 2 as well as the kinetics model described below.

For most continuous factors in both models, I chose to present estimates and 95% confidence intervals for a change of half the interquartile range, corresponding approximately to moving from one quartile to the next. For length of DAA treatment and time between visits, which have a natural scale, I used weeks.

3.3.2 Statistical analysis for VL rebound

Analysis considered kinetics of VL rebound up to 12 weeks post EOT. All VL measurements from the time of the first VL to meet the failure criteria up to and including 10 weeks after the initial failure VL or the time of the retreatment D0 VL, whichever came first, were included in the analysis. VLs after these time points were excluded, because either the participant's VL would have declined due to starting retreatment or data was sparse, suggesting that those that did not start retreatment closer to rebound might not have been generalisable and results may have been heavily influenced by any outliers. In total, only 7 (3% of total 221 measurements) VL results from 6 participants were excluded due to being past 10 weeks. Initial descriptive analysis suggested two phases in the VL rebound: an initial phase up to the peak rebound VL with a slight decline after this peak. To model this behaviour, a linear spline with one knot at the median time for the confirmatory VL result (2.14 weeks after the initial failure VL) was used and this time point was set to zero in the model so that the model intercept corresponded to the approximate peak rebound VL.

Linear mixed models were used to analyse the trajectories of HCV VLs. Random effects were used to allow for individual participant differences in the peak rebound VL and the slope from treatment failure VL to peak rebound VL; there was no evidence of variation in the slope after the peak VL (standard error for the estimate of variance in base model 7.00×10^{-10}) so this was not included.

Multivariable models were built using backwards elimination including, for each factor, a main effect corresponding to the impact of the factor on the peak VL, and interactions between the factor and each of the two slopes representing the different phases of the rebound. Factors were included in the model if the p-value for the term was <0.05, always including the main effect (representing the effect on the peak) if either interaction with time met this threshold. As the number of terms in the full model was already large compared with the size of the dataset, laboratory results were not included in the primary multivariable model but were added to the selected model and terms kept if $p<0.05$ as for the other factors. In addition to the baseline factors tested in Chapter 2, I also considered whether the participant's VL was <LLOQ (or <15 IU/ml for assays with greater sensitivity) at EOT for inclusion in the model, and the time between the initial failure VL and the prior visit, in order to account for the differences in visit schedule between the two trials, whereby STOP-HCV-1 had visits every 4

weeks after EOT, whereas SEARCH-1 had visits every 3-4 days up to EOT+4 where it switched to a 4 weekly visit schedule. Other EOT factors were not considered due to the large number of factors already being tested. The predicted peak rebound HCV VL from the final model was compared to the recorded baseline HCV VL using Spearman correlation.

3.3.3 Missing data

For both models, missing data in covariates was assumed to be missing at random and analysis was performed on the complete case data only (based on completeness of covariates, not VL measurements). Multiple imputation was not performed for the kinetics model due to the large numbers of factors and interactions that would need to be imputed; multiple imputation was not performed for the time until treatment failure to align with the kinetics model and also due to the lack of an established method so any method would have to be developed and tested, which is outside the scope of this thesis.

3.4 Results

3.4.1 Baseline demographics

Overall, 51 participants (13/51 (25%) with a treatment failure) from the SEARCH-1 trial were included in the analysis and 193 participants (56/193 (41%) with a treatment failure) from the STOP-HCV-1 trial (Table 3.1). Within those with a treatment failure, similar percentages took 28 days, 29-35 and 36-49 days of DAA treatment (29%, 32% and 30% respectively) with only 6 (9%) taking 56 days, which contrasts with the group of participants who cured where the majority (108, 62%) took 56 days of treatment and the second largest group took 28 days (36, 21%). Overall, ribavirin was taken by 96 (39%) participants, all of which were enrolled into STOP-HCV-1. The median (IQR) age of participants was 46 (38, 54) years, 89 (36%) participants were female, the predominant HCV genotype was 1a (170 (70%) participants), median (IQR) baseline HCV VL was 817892 (254325, 2251158) IU/ml, 92 (39%) had the IL28B genotype CC, 67 (27%) were coinfected with HIV (all in STOP-HCV-1) and 71 (31%) had baseline resistance to their prescribed DAAs.

Table 3.1: Demographics of participants included in the analysis

	Experienced treatment failure N=69 (column %)	**Achieved SVR12 N=175 (column %)**
Trial design		
Trial enrolled into		
STOP-HCV-1	56 (81%)	137 (78%)

	Experienced treatment failure N=69 (column %)	Achieved SVR12 N=175 (column %)
SEARCH-1	13 (19%)	38 (22%)
DAA days received	33 (28, 38)	56 (33, 56)
28	20 (29%)	36 (21%)
29-35	22 (32%)	14 (8%)
36-49	21 (30%)	17 (10%)
56	6 (9%)	108 (62%)
Received ribavirin	27 (39%)	69 (39%)
Treated using dasabuvir/ombitasvir/paritaprevir/ritonavir	56 (81%)	135 (77%)
Treated using glecaprevir/pibrentasvir	0	2 (1%)
Treated using sofosbuvir/daclatasvir	13 (19%)	38 (22%)
Baseline (pre-treatment) factors		
Age (years)	48 (40, 54)	45 (37, 54)
Female	23 (33%)	66 (38%)
BMI (kg/m^2)	24.6 (22.1, 25.9)	24.7 (21.8, 27.2)
Baseline HCV viral load (IU/ml)	1009081 (525856, 2327300)	776616 (213225, 2238721)
Baseline HCV viral load (log$_{10}$ IU/ml)	6.00 (5.72, 6.37)	5.89 (5.32, 6.35)
HCV genotype		
1a	54 (78%)	116 (66%)
1b	8 (12%)	37 (21%)
6	7 (10%)	22 (13%)
IL28B genotype		
CC	22 (33%)	70 (42%)
CT or TT	45 (67%)	96 (58%)
Fibroscan score (kPa)	4.8 (4.4, 5.9)	5.2 (4.4, 6.1)
HIV coinfected	20 (29%)	47 (27%)
Baseline resistance to prescribed DAA regimen	22 (33%)	49 (30%)
ALT (U/l)	45 (33, 74)	53 (30, 87)
AST (U/l)	38 (27, 52)	36 (28, 56)
ALP (U/l)	70 (60, 85)	72 (59, 88)
Albumin (g/l)	44 (42, 46)	45 (42, 46)

	Experienced treatment failure N=69 (column %)	Achieved SVR12 N=175 (column %)
Bilirubin (µmol/l)	8 (7, 10)	9 (7, 12)
Creatinine (µmol/l)	73 (63, 86)	73 (64, 81)
FIB-4*	1.05 (0.82, 1.54)	0.98 (0.74, 1.43)
Current depression	10 (14%)	28 (16%)
Current illicit substance abuse	9 (13%)	18 (10%)
Other factors		
HCV VL<LLOQ at end of treatment	59 (86%)	148 (86%)
DAA resistance at treatment failure	32 (51%)	-
Days between failure and previous visit	28 (23, 28) [3, 36]	-

*Fibrosis-4 index (calculated as age*AST/(platelets*$ALT^{1/2}$) (56)
Note: n (%) or median (interquartile range) [range]. Data split between failures and cures as kinetics analysis limited only to data from failure participants, therefore differences between the groups not tested here (see Chapter 2 for comparisons in STOP-HCV-1 participants). In total, 11 participants were missing IL28B genotype, 15 missing baseline DAA resistance, 6 missing DAA resistance at treatment failure, 4 missing substance use data, 20 missing AST and FIB-4, and 1 missing ALT, ALP, albumin, bilirubin and creatinine. Two cured participants were missing EOT VL data.

3.4.2 Factors associated with time to treatment failure

In the STOP-HCV-1 trial, 39 (70%) participants first met the failure criteria at their EOT+4 visit, 15 (27%) at their EOT+8 visit and 2 (3%) at their EOT+12 visit (Figure 3.2, panel (a)). In SEARCH-1, 7 (54%) had a treatment failure either at or before their EOT+4 visit (visits were scheduled every 3-4 days before this timepoint, cumulated for comparison with STOP-HCV-1 in panel (b)), 5 (38%) at their EOT+8 visit and 1 (8%) at their EOT+12 visit. However, not all visits took place exactly as scheduled so there was some variation as to when the treatment failure was first recorded, as seen in Figure 3.3 in the estimated hazard function from the interval-censored regression.

Being in SEARCH-1, taking longer DAA treatment, having HCV genotype 1b, having a lower baseline HCV VL and being younger were all independently associated with a lower hazard of treatment failure and therefore a longer time until treatment failure (Table 3.2).

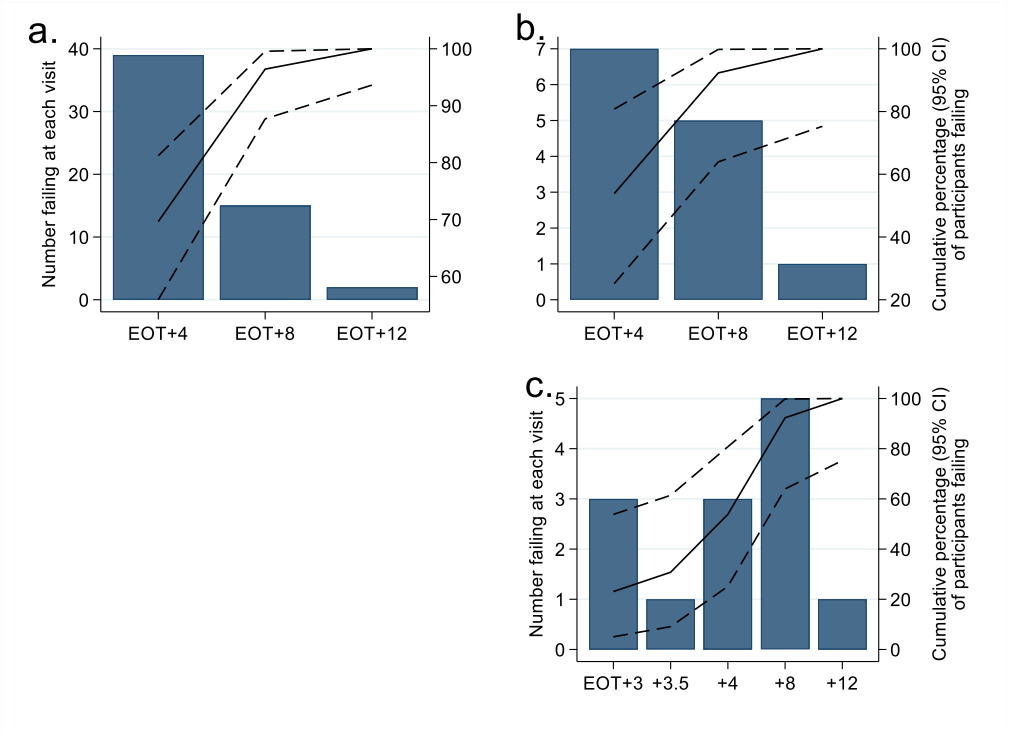


Figure 3.2: Timing and numbers of treatment failure (N=69: (a) 56 in STOP-HCV1 by week post EOT, (b) 13 in SEARCH-1 by week matching STOP-HCV-1 and (c) by sub-categorised week post EOT)

Note: Left x-axis shows the number of participants failing at each time point, right x-axis shows cumulative percentage of participants failing at each time point out of those who ever failed. In SEARCH-1, EOT+3.5 is 24 days after EOT.

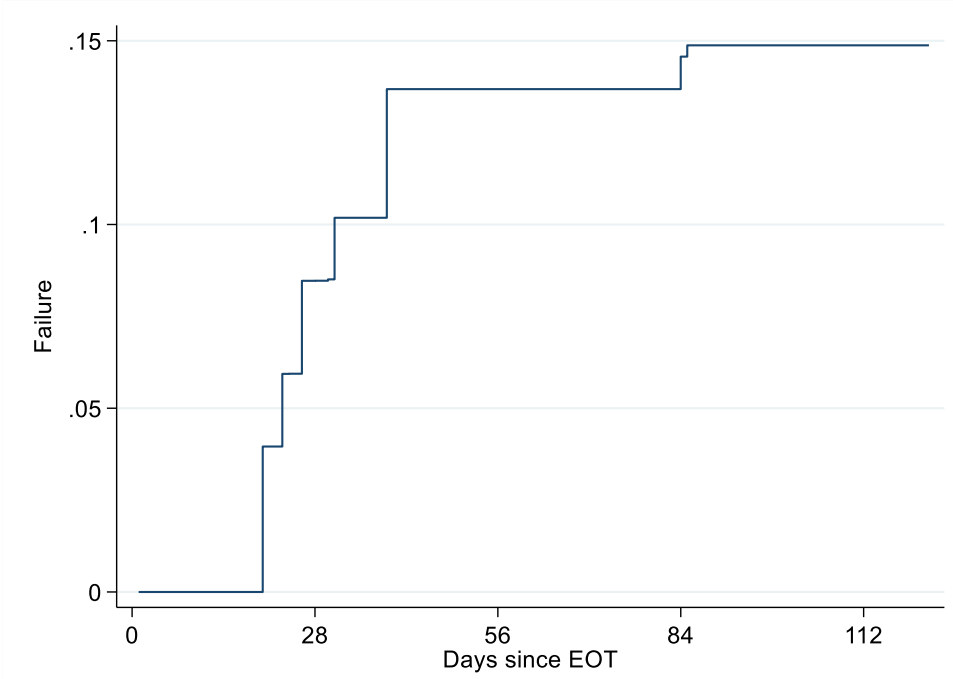


Figure 3.3: Hazard function for time to treatment failure from interval censored Cox regression

Note: Function estimated using adjusted model from Table 3.2 at the mean of each factor.

Table 3.2: Association between factors and time to treatment failure (N=242)

	Hazard Ratio (95% CI)	p-value	Fully adjusted Hazard Ratio (95% CI)	p-value
Trial: SEARCH-1 vs STOP-HCV-1	0.84 (0.45, 1.56)	0.58	0.23 (0.08, 0.68)	0.008
DAA days (per 1 day longer)	0.93 (0.91, 0.96)	<0.001	0.87 (0.84, 0.91)	<0.001
Genotype: 1b vs 1a	0.49 (0.23, 1.07)	0.07	0.36 (0.15, 0.86)	0.02
6 vs 1a	0.70 (0.32, 1.54)	0.38	0.58 (0.17, 1.98)	0.39
Baseline HCV viral load (per 0.33 $\log_{10}$ IU/ml higher)	1.14 (1.00, 1.29)	0.05	1.49 (1.26, 1.77)	<0.001
Age (per 8 years older)	1.12 (0.94, 1.33)	0.19	1.21 (1.04, 1.42)	0.02
Received ribavirin	0.96 (0.59, 1.56)	0.87	-	-
Female	0.82 (0.49, 1.36)	0.45	-	-
BMI (per 2 kg/m^2 higher)	0.95 (0.83, 1.09)	0.43	-	-
Baseline DAA resistance	1.06 (0.63, 1.78)	0.83	-	-
HIV positive	1.10 (0.66, 1.84)	0.70	-	-
IL28B: CC vs CT/TT	0.68 (0.41, 1.14)	0.14	-	-
Current substance abuse	1.30 (0.66, 2.56)	0.45	-	-
Current depression	0.94 (0.49, 1.83)	0.87	-	-
Fibroscan score (per 0.9 kPa higher)	0.93 (0.76, 1.13)	0.47	-	-
ALT (per 28 IU/l higher)	0.95 (0.85, 1.07)	0.38	-	-
AST (per 13 IU/l higher)	0.96 (0.85, 1.08)	0.51	-	-
ALP (per 14 IU/ higher l)	0.95 (0.81, 1.12)	0.56	-	-
Albumin (per 2 g/l higher)	0.96 (0.84, 1.10)	0.56	-	-
Creatinine (per 10 µmol/l higher)	1.08 (0.93, 1.26)	0.31	-	-
Bilirubin (per 2 µmol/l higher)	0.89 (0.78, 1.03)	0.11	-	-
FIB-4 (per 0.35 higher)	1.01 (0.87, 1.19)	0.87	-	-

Univariably, there was no evidence of a difference in time to treatment failure between the two trials (p=0.58), but after adjusting for other factors, being enrolled into SEARCH-1 was strongly associated with a later time to treatment failure (p=0.008) despite the visit schedule being more intense than that in STOP-HCV-1 with more opportunities to detect treatment

failure earlier. Increasing the length of DAA treatment by 1 day was very strongly associated with a 13% (HR 0.87 (95% CI 0.84, 0.91)) reduction in the hazard of treatment failure, with a corresponding reduction of 98% (HR 0.02 (95% CI 0.01, 0.06)) in those who took the longest length of treatment compared to those who took the shortest length (56 vs 28 days). Having a higher baseline HCV VL was very strongly associated ($p<0.001$) with a faster time to failure with an increased hazard of 49% (HR 1.49 (95% CI 1.28, 1.81)) for every 0.33 $\log_{10}$ IU/ml higher baseline VL. Despite baseline VL directly determining the length of DAA treatment for those randomised to variable length treatment in STOP-HCV-1 and D2 VL to the length of treatment for SEARCH-1 participants, there was at most weak correlation between baseline VL and length of treatment in all participants (spearman rho=0.19) and there was no evidence of an interaction between baseline VL and days of treatment in the model (p=0.61).

Although being infected with HCV genotype 1b was strongly associated with a much lower hazard of treatment failure compared to 1a (HR 0.36 (95% CI 0.15, 0.86); p=0.02), there was no evidence of a difference between those infected with HCV genotype 6 and 1a (p=0.39). However, there was also no evidence of a difference between genotypes 1b and 6 (p=0.49). Older participants were also more likely to have an earlier treatment failure with a 21% increase (HR 1.21 (95% CI 1.04, 1.46)) in the hazard for every 8 years older.

3.4.3 VL trajectories

Figure 3.4 shows the VL trajectories of individual participants from (a) the VL measurement before the visit where treatment failure was identified following the definition above and (b) from this visit itself. The majority of the last VLs prior to treatment failure took place ~4 weeks before with 45 (65%) participants having a gap of 25-31 days (median 28 (IQR 25, 28) [range 3, 83]) between the two VLs (Figure 3.5). Seven (10%) participants had their previous visit within 7 days of their treatment failure visit. At the VL measurement immediately preceding treatment failure, only 9 (13%) participants had a VL >LLOQ: 7 participants had a VL between LLOQ and 58 IU/ml and the remaining two participants had VLs of 200 and 755 IU/ml. In general, the closer these two visits, the steeper the VL trajectory between the two time points, however there was also a group of participants with ~4 weeks between visits who had a slower incline than other participants with a similar time between visits.

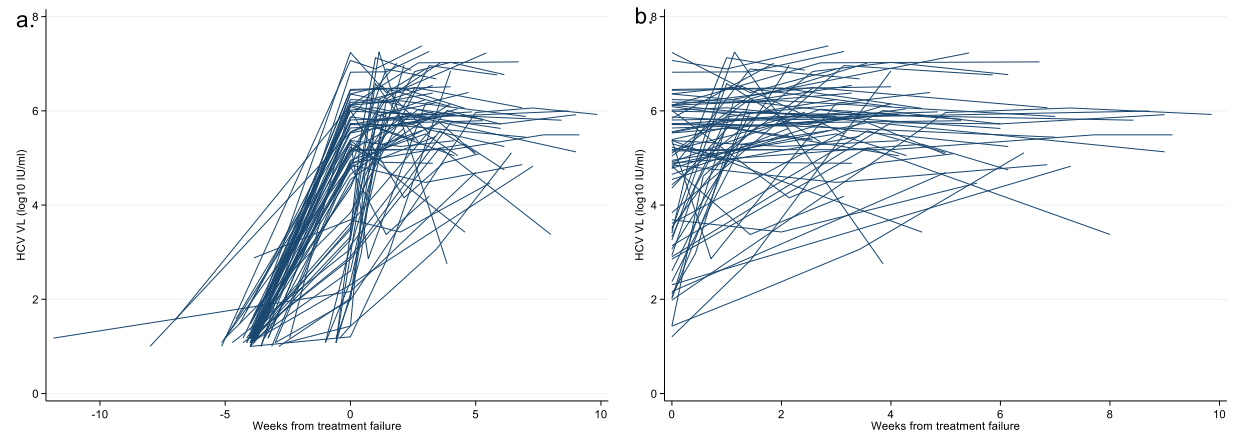


Figure 3.4: Individual viral load trajectories (N=69)

(a) Viral load trajectories including the VL measurement prior to the first visit meeting the definition of treatment failure; (b) Viral load trajectories from the first visit meeting the definition of treatment failure

Note: In both plots time zero is the time of treatment failure.

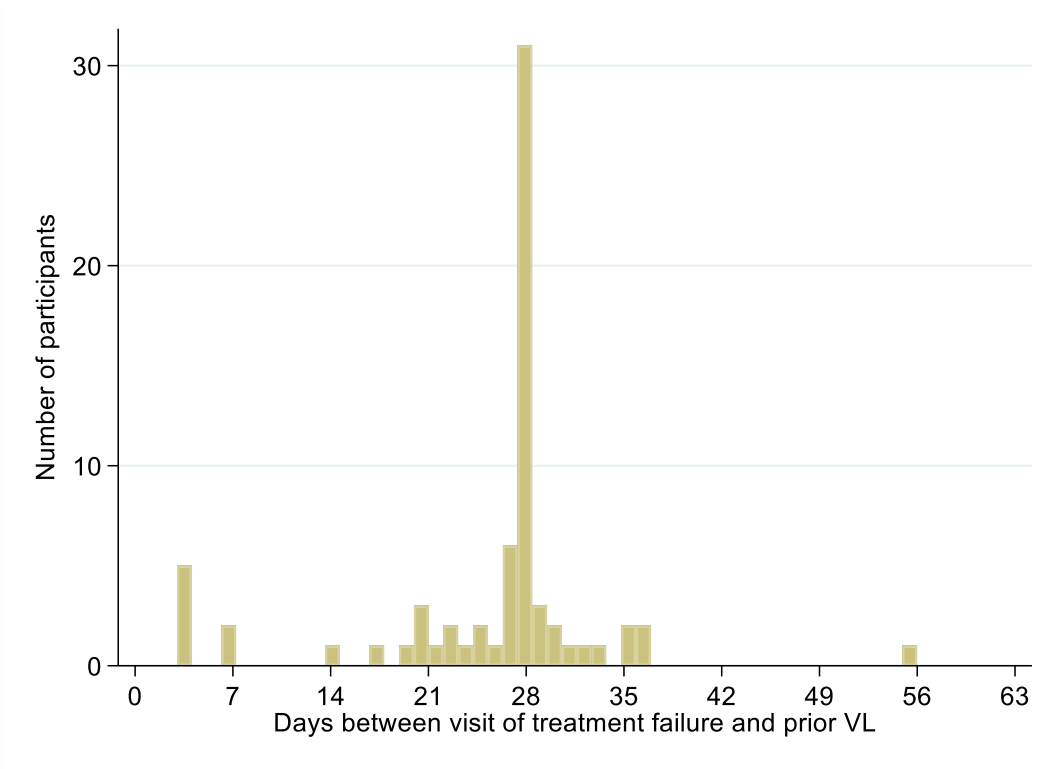


Figure 3.5: Distribution of time between treatment failure and the immediately preceding VL measurement

From the visit at which treatment failure occurred there appeared to be three main trends, based on visual examination, in the VL trajectories (Figure 3.4b, Figure 3.6): one group of

participants whose initial VL was high and either at or close to the peak of the rebound VL that changed little over the following visits; a second group of participants that had a lower initial VL that increased steeply over time with either a plateauing or slight decrease in VL after the confirmatory VL; and a third group that also had a lower initial VL but with a slower increase in VL often with no plateauing or decrease observed by the start of retreatment. These three groups largely had different timings for their peak rebound VLs: the first group had a peak either at treatment failure or the peak was of a similar magnitude to their treatment failure VL, the second group had their peak VL shortly after treatment failure, and the third group had their peak VL a few weeks after treatment failure.

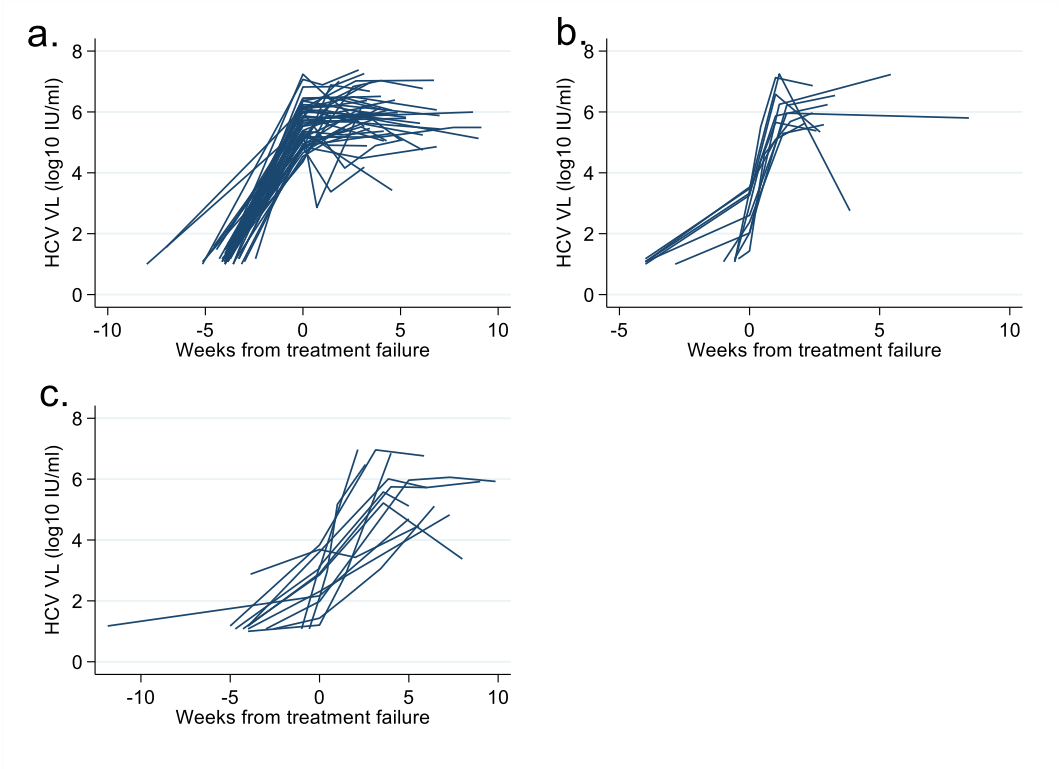


Figure 3.6: Individual viral load trajectories split by viral load behaviour
(a) Treatment failure VL is high and is close to the peak VL; (b) Treatment failure VL is low, but reaches peak quickly; (c) Treatment failure VL is low and reaches peak slowly

Note: In all plots time zero is the time of treatment failure.

3.4.4 The kinetics of HCV VL after rebound and associated factors

From a linear mixed model including only time from rebound (no covariates), the mean HCV VL at the time of treatment failure was 4.74 (95% CI 4.42, 5.07) $\log_{10}$ IU/ml, increasing up to a peak of 6.02 (95% CI 5.77, 6.28) $\log_{10}$ IU/ml 2.14 weeks later (timepoint fixed in the model) before decreasing slightly to 5.32 (95% CI 4.75, 5.88) $\log_{10}$ IU/ml 10 weeks after treatment failure (Table 3.3, Figure 3.7). While there was strong evidence that VL increased following the

first treatment failure VL up to the estimated peak at 2.14 weeks (p<0.001), there was much weaker evidence for a subsequent decline after the estimated peak (p=0.04). There was a similar amount of variation between participants in the peak VL and the rate of change in VL from the initial treatment failure VL to the peak (0.39 and 0.34 respectively), but there was only a moderate amount of positive correlation between these (0.45 (95% CI 0.07, 0.72)) indicating that those with higher peak rebound VLs tended to have faster increases in VL to this peak.

Table 3.3: Overall viral load kinetics (Participant N=69, VL N=210)

	Estimate (log_{10} IU/ml) (95% CI)	**p-value**
Fixed effects		
Peak/model intercept (log_{10} IU/ml)	6.02 (5.77, 6.28)	<0.001
First slope (increase per week)	0.60 (0.40, 0.80)	<0.001
Second slope (increase per week)	-0.09 (-0.18, -0.003)	0.04
Random effects		
Variance of the peak/intercept	0.34 (0.16, 0.72)	
Variance of the first slope	0.39 (0.22, 0.67)	
Correlation between the peak/intercept and the first slope	0.45 (0.07, 0.72)	

Univariably, peak rebound VL was higher in those with higher baseline HCV VL, higher Fibroscan score, younger individuals and those who reported current depression, but lower in those with higher ALP (Table 3.4). The rate of increase in VL after the initial treatment failure VLs was faster in those enrolled in SEARCH-1, HCV genotype 6, IL28B genotype CC, those with higher baseline HCV VL and slower in those with a longer time in between visits and higher ALT. In addition to their lower peak rebound VL, after this peak older participants had a slower decline in VL, or even an increase for those >58 years, suggesting that even though peak VL was lower, it may ultimately revert to a VL more similar to a younger participant. Declines after peak were also faster in those with baseline DAA resistance.

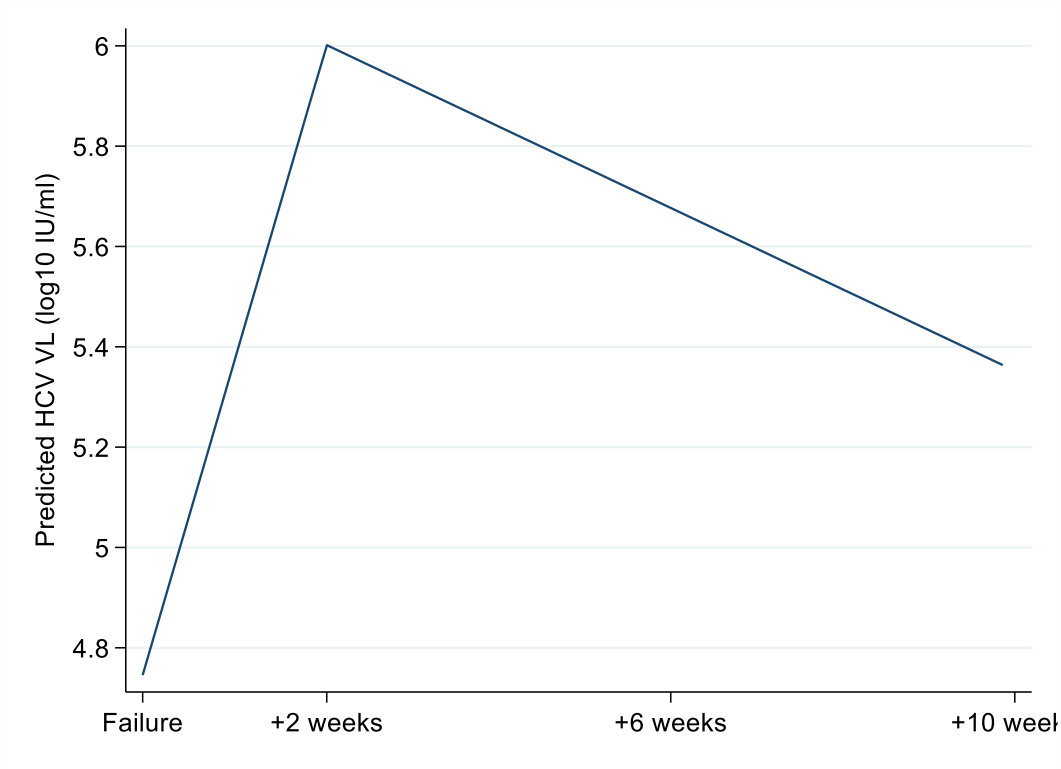


Figure 3.7: Mean rebound viral load trajectory

In the fully adjusted multivariable model, being enrolled into SEARCH-1, taking ribavirin, having a VL <LLOQ at EOT, a higher baseline HCV VL, not having baseline resistance to prescribed DAAs and a higher baseline Fibroscan score were all independently associated with a higher peak rebound VL (Table 3.5). Trial and DAA resistance had the largest impact on peak VL with each altering the peak by ~1 log_{10} IU/ml. Interestingly, for each 0.33 log_{10} IU/ml higher baseline HCV VL, there was only a 0.26 (95% CI 0.19, 0.32) log_{10} IU/ml higher peak rebound VL indicating that, all other factors being equal, there was a greater absolute difference in baseline and peak rebound VL in those with higher baseline VLs than those with smaller baseline VLs (although 95% CI were relatively large). The variance in the peak between participants decreased considerably from 0.34 in the model without any covariates to 0.01 (95% CI 0.0003, 0.20) in the fully adjusted model, therefore the selected model and factors successfully explain most of the heterogeneity in the peak rebound VL.

Table 3.4: Univariable analysis of factors potentially associated with viral load kinetics

	Change in peak (log_{10} IU/ml) (95% CI)	p-value	Change in first slope per week (log_{10} IU/ml) (95% CI)	p-value	Change in second slope per week (log_{10} IU/ml) (95% CI)	p-value
Trial: SEARCH-1 vs STOP-HCV-1	0.58 (-0.10, 1.27)	0.10	**0.67 (0.16, 1.19)**	**0.01**	-0.40 (-1.00, 0.19)	0.19
Received ribavirin	0.03 (-0.49, 0.55)	0.91	-0.23 (-0.64, 0.17)	0.26	-0.02 (-0.19, 0.16)	0.86
Timing of previous visit (per extra day)	-0.03 (-0.06, 0.003)	0.07	**-0.05 (-0.07, -0.02)**	**<0.001**	0.004 (-0.01, 0.02)	0.62
>LLOQ HCV VL at EOT	0.55 (-0.15, 1.25)	0.13	0.41 (-0.16, 0.97)	0.16	-0.13 (-0.35, 0.08)	0.22
DAA days (per 1 day more)	-0.003 (-0.03, 0.04)	0.83	-0.02 (-0.04, 0.01)	0.20	0.01 (-0.01, 0.02)	0.41
Genotype: 1b vs 1a	-0.57 (-1.31, 0.18)	0.14	-0.19 (-0.79, 0.41)	0.54	0.15 (-0.15, 0.45)	0.33
6 vs 1a	0.62 (-0.32, 1.56)	0.20	**0.71 (0.01, 1.41)**	**0.046**	-0.64 (-1.86, 0.58)	0.31
IL28B: CC vs CT/TT	0.42 (-0.12, 0.97)	0.13	**0.64 (0.23, 1.06)**	**0.002**	-0.09 (-0.27, 0.09)	0.33
Baseline HCV viral load (per 0.33 log_{10} IU/ml higher)	**0.34 (0.22, 0.47)**	**<0.001**	**0.11 (-0.001, 0.23)**	**0.051**	-0.01 (-0.06, 0.05)	0.82
Age (per 8 years older)	**-0.19 (-0.37, -0.02)**	**0.03**	-0.06 (-0.20, 0.09)	0.45	**0.07 (0.004, 0.14)**	**0.04**
Female	-0.25 (-0.78, 0.27)	0.35	0.17 (-0.25, 0.59)	0.43	-0.10 (-0.32, 0.12)	0.38
BMI (per 2 kg/m^2 higher)	-0.09 (-0.23, 0.05)	0.21	0.01 (-0.11, 0.12)	0.93	0.04 (-0.005, 0.09)	0.08
Baseline DAA resistance	-0.05 (-0.54, 0.45)	0.85	0.33 (-0.09, 0.75)	0.12	**-0.27 (-0.47, -0.06)**	**0.009**
HIV positive	0.17 (-0.39, 0.74)	0.55	-0.32 (-0.76, 0.12)	0.15	-0.03 (-0.21, 0.16)	0.79
Current substance abuse	0.59 (-0.23, 1.41)	0.16	0.16 (-0.44, 0.76)	0.61	-0.15 (-0.47, 0.17)	0.37
Current depression	**1.05 (0.25, 1.85)**	**0.01**	0.50 (-0.10, 1.10)	0.10	-0.25 (-0.54, 0.04)	0.09
DAA resistance at failure	-0.36 (-0.86, 0.14)	0.15	0.22 (-0.19, 0.63)	0.30	-0.06 (-0.23, 0.11)	0.52
Fibroscan score (per 0.9 kPa higher)	**0.22 (0.02, 0.43)**	**0.04**	0.12 (-0.04, 0.29)	0.16	-0.04 (-0.11, 0.04)	0.38
ALT (per 28 IU/l higher)	-0.09 (-0.22, 0.04)	0.17	**-0.11 (-0.21, 0.01)**	**0.03**	-0.01 (-0.06, 0.05)	0.79
AST (per 13 IU/l higher)	-0.03 (-0.16, 0.10)	0.62	-0.06 (-0.16, 0.05)	0.29	-0.02 (-0.08, 0.03)	0.36
ALP (per 14 IU/l higher)	**-0.24 (-0.43, -0.06)**	**0.009**	-0.01 (-0.16, 0.13)	0.88	0.06 (-0.01, 0.13)	0.12
Albumin (per 2 g/l higher)	0.09 (-0.05, 0.24)	0.21	0.001 (-0.11, 0.1)	0.99	-0.03 (-0.08, 0.02)	0.29
Creatinine (per 10 µmol/l higher)	0.12 (-0.02, 0.27)	0.10	-0.03 (-0.15, 0.09)	0.59	-0.01 (-0.06, 0.05)	0.85
Bilirubin (per 2 µmol/l higher)	0.02 (-0.15, 0.19)	0.80	-0.08 (-0.21, 0.05)	0.23	-0.02 (-0.07, 0.04)	0.58
FIB-4 (per 0.35 higher)	-0.13 (-0.30, 0.04)	0.15	-0.03 (-0.18, 0.11)	0.64	0.03 (-0.05, 0.11)	0.44

Table 3.5: Multivariable associations between factors and viral load kinetics (Participant N=65, VL N=198)

	Peak/slopes (log_{10} IU/ml) (95% CI)	**p-value**		
Fixed effects at the reference category				
Peak/model intercept	6.00 (5.76, 6.23)	<0.001	-	-
First slope (increase per week)	0.35 (0.14, 0.55)	<0.001	-	-
Second slope (increase per week)	-0.05 (-0.12, 0.01)	0.12	-	-

Associated factors	**Change in peak (log_{10} IU/ml) (95% CI)**	**p-value**	**Change in first slope (log_{10} IU/ml) (95% CI)**	**p-value**
Trial: SEARCH-1 vs STOP-HCV-1	1.09 (0.61, 1.58)	<0.001	-	-
Received ribavirin	0.28 (0.04, 0.51)	0.02	-	-
Timing of previous visit (per extra day, at 28 days)	-0.01 (-0.03, 0.01)	0.39	-0.04 (-0.06, -0.02)	<0.001
>LLOQ HCV VL at EOT	-0.50 (-0.81, -0.19)	0.002	-	-
IL28B: CC vs CT/TT	-0.03 (-0.32, 0.26)	0.83	0.39 (0.04, 0.74)	0.03
Baseline HCV viral load (per 0.33 log_{10} IU/ml higher, vs 6.04 log_{10} IU/ml)	0.26 (0.19, 0.32)	<0.001	-	-
Baseline DAA resistance	-0.97 (-1.30, -0.63)	<0.001	-	-
Fibroscan score (per 0.9 kPa higher, vs 4.8 kPa)	0.14 (0.05, 0.24)	0.003	-	-

Random effects	**Variance/covariance (95% CI)**
Variance of the peak/intercept	0.01 (0.0003, 0.20)
Variance of the first slope	0.27 (0.16, 0.46)
Correlation between the peak/intercept and the first slope	1 (-1, 1)*

*Correlation and wide confidence interval due to large decrease in variability in the intercept (estimate of covariance 0.05 (95% CI -0.04, 0.13)). Estimates for fixed effects similar to those in model without random effect for the intercept.

Note: For continuous factors, the reference is the median value of the factor in the dataset.

Those who had a shorter time between the treatment failure visit and the preceding VL measurement had a faster rate of increase in their VL between the treatment failure visit and the estimated peak at 2.14 weeks: for those with a previous VL 7 days before their treatment failure visit VL increased by 1.19 (95% CI 0.74, 1.63) log_{10} IU/ml per week, for those with a VL 28 days earlier by 0.35 (95% CI 0.14, 0.55) log_{10} IU/ml per week and for those with a VL 35 days earlier by 0.07 (95% CI -0.20, 0.33) log_{10} IU/ml per week. VL increased in those with IL28B genotype CC by 0.39 (95% CI 0.04, 0.74) log_{10} IU/ml per week faster than those with TT/CT. Neither factor was associated with peak rebound VL and no factor was associated with a change in VL after the peak VL. There was still a large amount of variance in the slope from the initial to the fixed peak VL between the participants (0.27 (95% CI 0.16 0.46) from 0.39 in the model without covariates) so although some heterogeneity can be explained by the timing of the preceding VL measurement before failure and IL28B genotype, there are other factors not tested for or selected in the model that would also explain the variation between participants. After adjusting for all factors, there was no longer any evidence of a change in VL after it had rebounded to its peak (p=0.12).

Comparing observed baseline HCV VL with estimated peak rebound VL in each participant (Figure 3.8), there was strong agreement between the two values (spearman's rho = 0.76). Where there was a larger disagreement between the values, participants generally had a lower predicted peak rebound VL than their baseline VL.

As well as not including ribavirin in the treatment regimen, SEARCH-1 differed from STOP-HCV-1 by excluding those coinfected with HIV but allowing for the inclusion of those with HCV genotype 6. Including HIV in the adjusted model did not alter the estimate for being in SEARCH-1 (1.12 vs 1.09) and there was no evidence of an association between HIV and HCV peak rebound VL (p=0.45). However, including HCV genotype did alter the association between trial and peak VL reducing the effect with slightly weaker evidence for the effect of trial (0.73 (95% CI 0.14, 1.31) p=0.02) and there was weak evidence of an association with genotype 6 and peak rebound VL (0.56 (95% CI 0.01, 1.11) p=0.047 vs 1a; no evidence of an effect for 1b vs 1a -0.18 (95% CI -0.59, 0.23) p=0.38).

In the STOP-HCV-1 trial, taking ribavirin was associated with fewer emergent resistance associated substitutions in those who failed and this association may explain the effect of ribavirin on peak VL. When ribavirin was replaced with having resistance to prescribed DAAs at treatment failure, there was a trend towards a lower peak rebound VL in those with resistance mutations of a similar magnitude to the effect of ribavirin (-0.30 (95% CI -0.65, 0.04) p=0.09).

Of the laboratory results that were tested for inclusion in addition to the model in Table 3.5, only baseline pre-treatment ALT and albumin were selected. Those with higher ALT had slower initial increases from initial treatment failure to peak (-0.10 (95% CI -0.19, -0.02) for every 28 IU/l higher, p=0.02), but there was no evidence of a difference in the peak rebound VL (p=0.32). Baseline albumin only affected the peak rebound VL with a 0.07 (95% CI 0.01, 0.14) log_{10} IU/ml higher peak for every 2 g/l higher albumin. Adding ALT and albumin to the model did not change the estimates or strength of evidence of any other factor substantively. Comparing the models overall, adding ALT and albumin only improved the model slightly (AIC 497.70 vs 502.68).

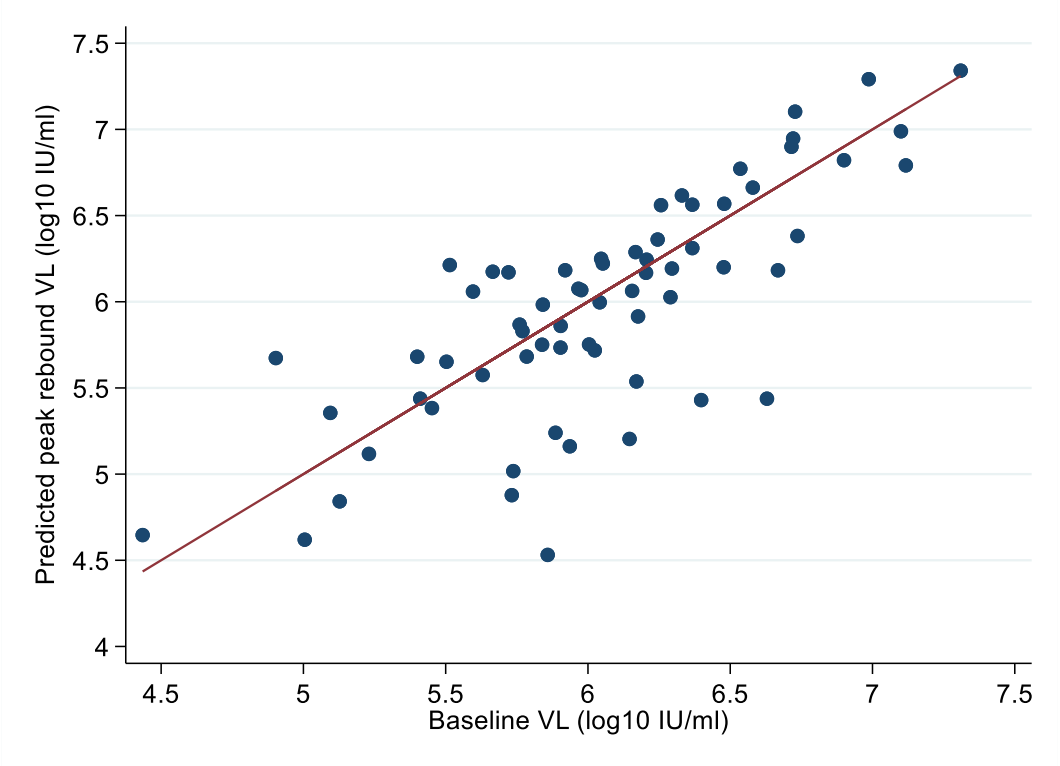


Figure 3.8: Observed baseline viral load and estimated peak rebound viral load

Note: Red line is y=x, when baseline VL and rebound peak VL are equal.

3.5 Discussion

Only one patient characteristic was associated with both rebound kinetics and time to failure: participants with a higher baseline HCV VL were more likely to fail and/or failed faster and had a higher peak rebound VL when they did have a treatment failure. Factors that had no evidence of effect on viral kinetics but were associated with a longer time to treatment failure were being younger and having HCV genotype 1b. Peak rebound VL was higher in those with no baseline DAA resistance, a higher Fibroscan score and who had <LLOQ HCV RNA at EOT.

Participants with IL28B genotype CC had faster viral rebounds than those with genotypes CT or TT. Treatment options also had an impact on the kinetics and timing of rebound: participants who took ribavirin had a higher peak rebound VL and participants who took more days of treatment had a longer time until treatment failure. Another aspect of trial design, the time between visits before treatment failure was detected, also had an effect on the speed of viral rebound.

The trial a participant was enrolled into was identified as a factor affecting both viral load kinetics and time to treatment failure. As well as the different trial designs which were partly included in the models (ribavirin in the kinetics model, days of DAA treatment in the time to failure model), this also represented potentially many other factors that were either measured and not included (for example adherence to trial drug which was higher in SEARCH-1 (10% of participants reported missing at least one dose vs 28% in STOP-HCV-1, unpublished data) or which DAAs the participant received) or were not able to be measured. It may also at least partly represent other factors specific to each trial that were part of the model selection process that were not chosen in the final multivariable model, such as HIV and HCV genotype for the VL kinetics model. For HCV genotype specifically, when this was added to the model of viral load kinetics, the effect of being enrolled in SEARCH-1 on peak rebound VL was smaller indicating that in the selected model the effect of genotype 6 may be being included within the effect of the trial, potentially due to small numbers and limited power.

It is well known that short-course treatment has lower cure rates than full length treatment (78) and within this dataset almost all participants who took 56 days of treatment cured (108/114, 95%); the reasons for this also explain why a longer treatment length was associated with less and/or later treatment failure. As the first phase of viral load decline happens so quickly with DAA treatment (67), it is only the second phase that is limited by the use of short-course treatment. As the second phase is when infected cells die, shortening the time for this process means that an inadequate number of cells are removed to prevent viral rebound from happening. The number of infected cells remaining will then influence when rebound occurs. The fewer infected cells remain, the longer it will take for the number of infected cells to increase such that the levels of virions they produce to become detectable in laboratory tests and for treatment failure to be observed.

Being infected with HCV genotype 1b was associated with a lower failure rate and longer time until treatment failure in the analysis and was also associated with SVR12 in Chapter 2. In addition, previous research has shown an association with SVR12 (45) and in another study limited to those with genotype 1b and <500 IU/ml VL on day 2 of treatment, there was a 100% SVR12 rate after only 3 weeks of treatment (41). It is likely that patients infected with HCV

genotype 1b need fewer days of treatment than those with 1a and so when given short-course treatment are either more likely to have cured or be closer to cure and have fewer infected cells from which the VL can rebound from, thus lengthening the time until detectable treatment failure, compared to similar patients with genotype 1a. There was no evidence of an association between genotype 1b and rebound kinetics: this may be due to a genuine lack of difference in rebound kinetics compared to genotype 1a, but may also be due to a lack of power as fewer participants were included in the analysis of rebound kinetics compared to SVR12 and time to treatment failure.

The association between being older and a shorter time until treatment failure/more likely to fail may be explained by the effect of age on the immune system. As patients age, the ability of T cells to respond to infection and remove infected cells declines. In patients with chronic HCV infection, T-cells are also impaired by immune exhaustion, which occurs due to prolonged exposure to HCV antigens; during DAA treatment this exhaustion is reversed and the T-cells contribute to viral load decline (79). If the patient's T-cells are additionally impaired by age, then functionality will not return to that of a younger patient and viral load decline will be slower and longer treatment will be needed to ensure cure.

It would be expected that those with higher baseline VLs had higher peak rebound VLs. While there was not an exact correspondence between baseline and peak (a 0.33 $\log_{10}$ IU/ml increase at baseline led to a 0.26 $\log_{10}$ IU/ml increase at the peak), the difference was not large. It may be that the timing of when the peak occurs is later than the 2.14 weeks fixed in the model in those with higher baseline VLs, as the VL needs to rebound higher to reach the pre-treatment level, and if a later peak was assumed for these participants than a larger association with peak VL would be estimated. However, if the peak was later in those with higher baseline VLs, it would be expected to see a difference in the change in VL post the fixed peak at 2.14 weeks compared to those with lower baseline VLs, but there was no evidence of this ($p=0.79$). It may also be the case that peak VL is smaller than baseline VL and that treatment impairs the ability of the virus to reach the same levels of VL, even if ultimately the participant did not cure on treatment. It might also be expected that those with higher baseline VLs would experience treatment failure earlier after EOT: those with higher baseline VLs are known to need longer lengths of treatment and all participants received treatment shorter than the recommended length (maximum length of 8 weeks compared to standard length 12 weeks).

Participants with resistance to prescribed DAAs at the initiation of treatment had lower peak rebound VLs than those who did not. Although the mutations made the virus less susceptible to DAAs, this may have come with a fitness cost that limited the ability of the virus to replicate,

which would lead to lower levels of virus and a lower peak rebound VL. This may also explain the association with ribavirin, which was associated with higher peak rebound VL. In the STOP-HCV-1 results, participants who took ribavirin were significantly less likely to develop a new resistance mutation to DAAs, and NS5a inhibitors in particular, than those who did not (27), and so by the time of treatment failure it is expected that those taking ribavirin were less likely to have resistance to prescribed DAAs. When ribavirin was replaced in the model with resistance to DAAs at treatment failure, there was a trend towards an effect of similar magnitude, but a decrease in peak VL instead of an increase. This strengthens the hypothesis that the effect of ribavirin might be related to lower rates of DAA resistance, and the fact that the model happened to select ribavirin rather than DAA resistance related to chance. Although baseline DAA resistance was also associated with treatment failure in the model in Chapter 2, it was not associated with a higher failure rate/time to treatment failure here. This may be because of the inclusion of participants in STOP-HCV-1 who took 8 weeks of treatment and SEARCH-1 participants: while the prevalence of DAA resistance was only slightly higher in those receiving 8 weeks compared to 4-7 weeks (18% vs 11%), there was far higher levels of DAA resistance in those enrolled into SEARCH-1 (92% vs 14% overall in STOP-HCV-1).

A higher peak rebound VL was also associated with a higher Fibroscan score. Fibroscan measures the level of inflammation or fibrosis in the liver and a higher peak VL suggests that in these participants it was more likely to be a measure of inflammation, as fibrosis may mean less tissue in the liver for the virus to infect and replicate within, which may have led to lower peak rebound VLs. Higher levels of inflammation would indicate higher levels of infection within the liver and therefore the VL is more likely to rebound to a higher level than those with lower levels of inflammation.

Although having a >LLOQ VL at EOT was associated with a lower peak rebound VL, it is more plausible that the estimated time of the peak rebound VL is incorrect in those patients who were >LLOQ at EOT and so the VL being estimated at that time point is a different point in the rebound trajectory than for those with <LLOQ VL at EOT.

Having a shorter time between the first visit indicating a treatment failure and the preceding VL measurement was associated with a faster rebound to peak VL, which is likely due to when the treatment failure was detected within the rebound trajectory. Although log viral load rebound was modelled as a linear change over time, it is unlikely to follow a constant, linear increase from the initial time of treatment failure to the peak rebound VL. If the time between VL measurements is shorter, treatment failures may be more likely to be detected in their initial stage, whereas for those with longer times between VL measurements the treatment failure may be more likely to be detected close to, or possibly even at, the peak of its rebound

and it is reasonable to assume that the rate of increase in VL at these two scenarios is different.

The difference in the rate of increase in VL between those with IL28B genotype CC and non-CC may also be explained by treatment failure being detected in a different part of the rebound trajectory and that the estimated time of the peak occurring 2.14 weeks after the initial failure VL may be incorrect for those with genotype CC. As those with genotype CC have a stronger immune response to HCV, a faster rebound appears counterintuitive. However, this stronger immune response may also alter the timings of the rebound trajectory and so the change in VL from treatment failure to 2.14 weeks is not covering the same period in the rebound trajectory in both genotype groups. Although IL28B was associated with SVR12 in Chapter 2 and with viral load kinetics here, it was not associated with time to treatment failure. The significance threshold for selection in the kinetics and time to treatment failure models was lower than that for the binary failure outcome (p=0.157 vs 0.05). When the higher significance threshold was used for selection for the time to treatment failure model, IL28B genotype CC was associated with a lower risk of failure or longer time until treatment failure (HR 0.63 (95% CI 0.36, 1.11) p=0.11).

Only two laboratory measures were selected for inclusion into the previously chosen model for viral load rebound. Higher levels of albumin were associated with higher peak rebound VL and higher ALT levels were associated with a slower change in VL from treatment failure to peak rebound VL. Low levels of albumin can indicate liver disease and, in these patients, a lower rebound VL might be expected due to fibrosis. However, albumin levels in the trial participants were within the normal range and, by inclusion criteria, all participants had mild liver disease so this is unlikely to be the case here. Higher ALT levels indicate liver inflammation, but it is unclear how this would lead to a slower rebound in VL. As a large number of factors were tested and already previously chosen for the model, it may be that these laboratory measures were selected by chance or that with the inclusion of these factors the model is overfitting to the small sample size.

Although no factors were associated with a change in VL after peak rebound VL, the inclusion of the time between the treatment failure VL and the preceding VL measurement in the model reduced the size of the estimated change in VL per week after the peak and the strength of the evidence of an effect (p=0.04, 0.12 in the overall, unadjusted model and fully adjusted model, respectively). It is therefore likely that the effect in the unadjusted model is due to differences in the timing of the peak VL and once this has been partially adjusted for by the timing of the preceding VL measurement there is no evidence of change in VL post peak.

One strength of this analysis is that for the VL kinetics model, most of the variability in peak rebound VL between participants was explained by the factors selected in the model suggesting that the key factors have been identified. Although the reduction in variability in the slope from treatment failure to peak was smaller, this aspect of the rebound in VL is much more susceptible to the timing and frequency of VL measurements and would require additional analysis to account for these differences. Future work on the kinetics model could utilise a Bayesian approach that considers the timing of the peak rebound VL as a random effect, which would allow for both an estimate of the timing of the peak VL after initial failure VL as well as allowing it to vary between participants. An additional strength is that for both models, most factors were either expected from previous research or have been identified in other models within this thesis.

In common with the previous chapter, one weakness of this analysis is the small sample size, particularly for the analysis of the VL kinetics. However, all published trials examining short-course HCV treatment to date have been small and, comparatively, this sample is of a reasonable size and has been able to identify at least some associated factors. Future work could incorporate data from the VIETNARMS trial (described in Chapter 5), which is currently in progress with follow-up due to complete Autumn 2023, to increase the sample size and allow for the development of a more stable model, if enough treatment failures occur.

Another weakness is the sparse sampling through the 4-weekly post-EOT visit schedule in the STOP-HCV-1 trial, and similarly from 4 weeks post EOT in the SEARCH-1 trial. This means for some participants their treatment failure was detected later in their rebound trajectory than for others which adds complexity to the interpretation of the models. However, to detect treatment failure close to the actual time point would require intense visit scheduling and testing which may not be palatable for trial participants.

A potential limitation is that this analysis was performed on the complete case data only and multiple imputation was not used due to the complexity of the models and the lack of standard methods; any use of multiple imputation would likely require a separate piece of methodological work to verify. Of the models selected, there was no missing data in the factors selected for the time to treatment failure model and data was not included from only four participants in the kinetics model due to missing data. Therefore, estimates using multiple imputation would either be the same or likely to be similar to those from the complete case data. However, it may have had a larger impact on factor selection. Two factors that were *a priori* important that also had large amounts of missing data were baseline resistance to DAAs and IL28B genotype: these two factors were selected for the kinetics model so using MI data would likely not impact factor selection for this model, but as they were not selected for the

time to treatment failure model, it is possible they may have been included if data was not missing.

A weakness of the viral load rebound model is that it assumed a linear increase in VL rebound. Further analysis could explore models that account for non-linear increases in VL.

In the analysis of time until treatment failure, I chose to measure the intervals in which the failure occurred from the last known VL and in the kinetics model I chose to model the VL rebound from treatment failure, but for both models I could have used the most recent undetectable VL measurement instead. For participants who met the first definition of treatment failure (listed in Section 1.5.1), which defines treatment failure as having two consecutive VL>LLOQ after two consecutive visits of VL<LLOQ, there are no intermediate detectable results between the last undetectable result and treatment failure and the analysis is exactly the same. For participants who met the second definition, two consecutive VL measurements that are >1$\log_{10}$ increase above nadir on treatment *and* >2000 IU/ml, it is possible that they did have an intermediate detectable VL <2000 IU/ml between their last undetectable VL and treatment failure. There were 18/69 (20%) participants who met the second definition, of which 8 (12% of total number of treatment failures) who did have one intermediate VL and 10 who did not, so overall there may not have been a large effect on the estimates.

For the time to treatment failure analysis, measuring the intervals from the last undetectable VL rather than the last known VL would have widened the intervals in the 8 participants with an intermediate VL measurement so any change would have led to less precise results. In the kinetics model, it could be argued that, although participants had not yet met the definition for treatment failure, they had started to rebound so these measurements should be included; an analysis including these measurements could be performed to see how influential they are. However, once participants had a VL that could indicate potential treatment failure, they could be recalled by their study site earlier than expected for the confirmatory VL, whereas the visit schedule would continue as planned if a participant had a detectable VL <2000 IU/ml. So including these intermediate VLs could lead to even greater variability in the timing of peak rebound VL from the first VL included in the model and would need to be modelled using Bayesian techniques as described above.

I also chose to censor viral load trajectories after the first retreatment visit or 10 weeks from the treatment failure VL, whichever came first, but I could have included all VLs measurements up to the retreatment D0 result. It is possible that the VL trajectory would have behaved differently past the 10-week time point and that may have provided more information about

the trajectory overall. However, the number of VL measurements past this time point was small so any results would be more susceptible to outliers and may not provide accurate or generalisable results. Additionally, if a participant had results past 10 weeks, this was due to delayed clinic visits, and not due to having more VL tests, and so these participants lacked data earlier in their VL trajectory.

Despite the weaknesses, this is an important analysis as it is the first to look at viral load timing and rebound kinetics. Further analyses with data from other trials, either by building new models to compare against those developed here or validating these models will be important in helping to identify those patients suitable for short-course treatment and how short it should be.

3.6 Key findings

- Taking longer treatment, HCV genotype 1b, a lower baseline HCV VL and being younger were independently associated with a longer time from EOT until VL rebound and treatment failure combining data from the STOP-HCV-1 and SEARCH trials.
- Taking ribavirin, having HCV RNA <LLOQ at EOT, a higher baseline HCV VL, no baseline resistance to prescribed DAAs and a higher Fibroscan score were independently associated with a higher peak rebound VL, while having IL28B genotype CC increased the speed of rebounding HCV VL.
- Peak rebound VLs were largely similar or if anything slightly lower than baseline VLs.

4 The quality of life of participants in a short-course treatment trial

4.1 Introduction and aims

The overall goal of this chapter was to explore how quality of life (QoL) was affected by DAA treatment within the STOP-HCV-1 trial. Participants completed questionnaires assessing QoL at enrolment/day 0 (D0), end of treatment (EOT) and 12 weeks after completing treatment (EOT+12) visits on both first-line and any retreatment and this allowed for an examination of QoL at different stages within the trial.

The first aim was to investigate the impact of different lengths of DAA treatment, adjunctive ribavirin and the impact of cure itself, on QoL.

Ribavirin was an important factor to consider as it is known to cause adverse events and not be particularly well tolerated by patients when used with PEG-IFN. However, within short-course treatment it may be better tolerated due to the shorter treatment length. Although DAAs are generally well tolerated by patients, shorter treatments may still be associated with better QoL than standard length treatments, which could encourage treatment adherence in hard-to-treat populations.

QoL was examined at both EOT and EOT+12/day 0 of retreatment (RT D0): at EOT a large proportion of participants were virologically suppressed and may have started to see physical improvements in their QoL from the treatment; at EOT+12/RT D0 participants would have either known they had failed or been largely confident they had cured and this may have impacted their mental QoL negatively or positively, respectively, as well as any potential continuing physical QoL improvements from treatment failure or being cured.

The second specific aim was to assess the impact of anaemia on QoL within the trial. Anaemia is a common adverse event with ribavirin that is also known to reduce QoL and may have mediated or explained any overall impact of ribavirin on QoL.

4.2 Background

4.2.1 Methods for measuring quality of life

Three methods for assessing QoL were used in the STOP-HCV-1 trial: the EQ-5D-5L, the MOS-Cog and the SF-12. The EQ-5D-5L (80) has a descriptive system that measures health-related QoL in five areas (or dimensions) of health (mobility, self-care, usual activities, pain, anxiety) on a 5 point scale (no problems, slight problems, moderate problems, severe problems, unable to/extreme) with an additional visual analogue scale (VAS) where participants are asked to rate how their health feels today on a scale from 1-100 (100 = perfect health).

The Medical Outcomes Study Cognitive Functioning Scale (MOS-Cog) (81) assesses the frequency of cognitive issues that affect quality of life. The six questions cover a range of cognitive behaviours (reasoning, memory, attention, concentration, confusion, reacting) on a six point scale (all, most, a good bit, some, a little, or none of the time).

The 12-Item Short Form Health Survey (SF-12) (82) consists of 12 questions covering 8 health concepts (physical functioning, role-physical, bodily pain, general health, energy/fatigue, social functioning, role-emotional, mental health) with a variety of scales used, depending on the question being asked. Two summary scores can be generated, a Physical Component Summary (PCS) and a Mental Component Summary (MCS), using the weights recommended in the physical and mental scales (83).

4.2.2 Impact of treatment and cure on quality of life measures

It is well documented that patients living with HCV report having clinically significant lower QoL than healthy individuals (84). Evidence from studies looking at treatment with either PEG-IFN or DAAs show that, overall, patients who receive treatment, and particularly those who cure, have better QoL than those HCV patients who do not, and that the former can have QoL comparable to healthy controls (85-88). Even though treatment with PEG-IFN is associated with an improvement in QoL after treatment has ended, there may be lasting adverse effects from PEG-IFN. Treatment with DAAs is now standard of care with shorter duration of treatment than with PEG-IFN, lower toxicity and better safety profiles. DAA therapy has been associated with better QoL than those receiving PEG-IFN (87, 89), particularly during treatment, and the rest of this section will focus on studies using DAAs. My search of the literature included both observational studies as well as clinical trials and was not limited to settings or calendar years. The only restrictions on the search were to only look for studies using DAAs and then this was limited to short-course DAA therapy only.

Assessing QoL using the SF-36 (an expanded version of the SF-12 consisting of 36 questions), nearly all studies saw a significant increase associated with successful treatment in the PCS score (only one did not (90)). However, the evidence for an improvement in MCS and, in studies with more detailed analysis, in each of the eight areas that make up the survey was mixed. For the MCS score, all studies reported a decrease during treatment and at EOT compared with baseline, although this was not always significant; two studies reported a significantly improved MCS at EOT+12 compared to baseline (91, 92). Three studies reported improvements in all eight areas comprising the MCS and PCS either at EOT or at follow-up post EOT (91, 93, 94), two reported improvements in all but one or two areas (95, 96) and two reported an improvement in only a small number of areas (97, 98). The most common area not to see an improvement was pain, with three studies either reporting no change or a significant

worsening from baseline. All studies reported a significant improvement in energy and all but one reported a significant improvement in general health. Although most studies have reported results based on statistical significance, studies using data from the German Hepatitis C-Registry defined a clinically meaningful improvement in the PCS and MCS scores as an increase of 2.5 points. They reported that 41.2-41.7% of patients had a clinically meaningful increase in PCS and 53.8-54.7% in MCS, but that these percentages were higher in patients with active drug use and psychiatric disorders (99, 100).

Assessing QoL after achieving cure using the EQ-5D-5L, all studies saw significant improvements from baseline in the mobility, pain, anxiety or depression components and the related summary index, with no study reporting a significant change in problems with self-care. Two of the studies also saw significant improvements in carrying out usual activities (101, 102), while the third also saw a significant improvement in the VAS (103). In studies examining the effect of cirrhosis, most found evidence of a difference in QoL scores between those with and without cirrhosis with larger improvements in QoL scores in cirrhotic patients who had lower baseline QoL scores (91, 92, 94, 101, 103).

The positive effect of treatment and SVR on QoL has also been shown to occur in key sub-groups of HCV patients, including in patients coinfected with HIV (91), people who inject drugs (102) and in patients with liver transplants (93).

Not all studies included ribavirin in the treatment regimen or reported its effect, but those that did showed a decline in QoL during the initial weeks of ribavirin-containing treatment; while some studies found that QoL scores returned to baseline or higher before EOT (92, 104), others showed a decline in QoL scores that persisted until EOT (90, 105) or beyond EOT (ribavirin data was not reported by SVR12 status) (100). In all cases, QoL scores in those taking regimens including ribavirin were significantly lower than those taking ribavirin-free regimens at both EOT and at any post-EOT measurements. No study controlled for anaemia in the analysis of ribavirin, so it is unclear whether any negative impacts are simply due to anaemia or other adverse effects from the drug. However, one study did not find an association between decline in QoL score and decrease in haemoglobin (90). The negative effects associated with ribavirin occurred both when taken with PEG-IFN and with DAAs. Four of the five studies testing the effect of ribavirin found it impacted the MCS (90, 100, 104, 105) and three found an association with a decreased PCS (100, 104, 105).

I could not identify any literature to date examining the effect of short-course DAA treatment on QoL compared to standard length DAA treatment.

4.3 Methods

In STOP-HCV-1, participants who experienced a treatment failure before EOT+12 weeks were not expected to attend the EOT+12 visit and so QoL data is missing at this timepoint. For participants with treatment failure, the data from their RT D0 visit can be compared against the EOT+12 data for those who cured on first-line treatment, but it is important to note that both groups would have either known or been reasonably confident of their status (failed vs cured respectively) at the point they completed these questionnaires. Specifically, although cure will not have been confirmed at the time participants completed the QoL questionnaires at EOT+12, they will have attended visits at EOT+4 and EOT+8 and been aware that their HCV viral loads at these visits will have been suppressed and so they were likely to be cured. Of the participants who attended the EOT+12 visit, only two experienced treatment failure at EOT+12 with another two failing first-line treatment at EOT+24 (who therefore achieved SVR12 and are considered as cured in this analysis, both randomised to 8 weeks fixed duration treatment without ribavirin under VUS1). This is in contrast to participants at their EOT visit who would not have even been aware of their final treatment viral load and so would have no knowledge about their chance of ultimately being cured.

For the analysis of the EQ-5D-5L questionnaires, a single summary value covering the five dimensions was created using the EQ-5D-5L value set for England (106). In the descriptive system, the questions are ranked such that a higher score is worse QoL, but for this analysis the scales have been inverted to allow for easier comparisons with the other questionnaires (so higher score is better QoL consistently throughout). For analysis of the MOS-Cog questionnaires, the responses were transformed to an overall summary score on the range 1-100 (100 = perfect health) by ranking the scales from 1-6 (6 = none of the time), converting this rank to a score by the formula (rank – 1)*20 for each question and then summing the scores of all questions, as recommended by the User Manual. Due to the large number of questions on the SF-12, only the physical (PCS) and mental (MCS) summary scores have been analysed post-baseline here.

Analysis of QoL scores at both EOT and EOT+12/RT D0 was performed with data from all available participants apart from three participants in the fixed duration arm who took substantially shorter treatment than randomised (27, 32, 40 days rather than 56 days) and subsequently failed on first-line treatment. The maximum number of participants included in a comparison was 192. Only 3 participants were LTFU before their EOT+12 visit and all participants with a recorded treatment failure returned to start retreatment and so the impact of missing data on the results was considered minimal and data was performed on complete cases.

Univariable analyses compared answers to categorical questions using chi-squared tests or Fisher exact tests as appropriate; continuous summary scores were analysed using rank sum tests. For multivariable analyses, ranked scores from the EQ-5D-5L and MOS-Cog questionnaires were analysed using ordinal logistic regression because of the limited number of values taken. The summary scores of the EQ-5D-5L and MOS-Cog questionnaires were distributed such that there was a larger proportion of participants scoring perfect QoL than expected from the rest of the distribution. To accommodate this, these scores were analysed with two-part regression models consisting of logistic regression for whether a participant reported perfect QoL vs not and beta regression for analysing the summary score of those not reporting perfect QoL. The EQ-5D-5L patient reported VAS and the MOS-Cog summary scales were converted from a 0-100 scale to a 0-1 scale to allow for modelling using the beta distribution, with 1 indicating perfect health. The summary scores from the SF-12 were normally distributed and so were analysed using linear regression. Each model was adjusted for ribavirin randomisation, duration randomisation, VUS strategy, whether the participant achieved SVR12 and the baseline value of the QoL score.

As anaemia is a common adverse event from taking ribavirin, the effect of having anaemia on QoL was also investigated by including a term for anaemia in the model for QoL at EOT described above. Participants were considered having anaemia if their EOT haemoglobin was <12g/dl for women and <13g/dl for men. Sixteen further participants had anaemia using this definition at a timepoint before EOT, without also having anaemia at EOT (all grade 1). The effect of having anaemia was not assessed at EOT+12/RT D0 as any anaemia at this time point is unlikely to have been caused by treatment and in analysis of the trial there was no difference in haemoglobin levels at EOT+12 between those randomised to receive ribavirin and those not, whereas there was an -1.67g/dl (95% CI -1.92, -1.43) difference at EOT (27). This model estimates the direct effect of having anaemia at EOT on QoL at EOT; in this model any effect of ribavirin is the residual effect not mediated through the presence of anaemia.

There were a large number of tests performed; on each of the 19 QoL measures (28 at baseline due to comparing the separate components of the SF-12), I compared at three time points:

- Fixed vs variable (VUS1/VUS2 combined) (randomised comparison so any differences at baseline must be due to chance)
- VUS1 vs VUS2
- Ribavirin vs no ribavirin (randomised comparison)
- Cured vs failed (SVR12)
- Having anaemia vs not having anaemia (at EOT only)

Despite the large number of tests, a significance level of 5% was used to present results, rather than an artificial p-value adjustment which may be conservative as it is highly likely that some QoL measures were correlated. As the maximum number of participants in each comparison was 192, power was low. Low power combined with a lack of previous research to provide information about the effects of short-course treatment on QoL is another reason for p=0.05 as an initial threshold for statistical significance. As power was low, marginal findings (where 0.05>p≥0.1) and any relative differences >1.5 or <0.5 in magnitude, regardless of statistical significance, were also highlighted (and labelled as numeric differences).

All analyses were performed in Stata v16.1.

4.4 Results

4.4.1 Descriptive summary of questionnaire responses (univariable comparisons)

Overall, on the EQ-5D-5L questionnaire, in each of the 5 domains most participants reported no problems with their health; 42-45% of participants reported no problems in all 5 domains and therefore had a perfect index value of 1 at each visit (Table 8.1, Figure 4.1). When asked to rate their health on the VAS, the health score was lower with only 8-13% reporting a perfect score of 100 and median (IQR) VAS score of 85 (75, 95) at baseline and at EOT+12/RT D0 and 90 (75, 90) at EOT (Figure 4.2). However, the percentage reporting no problems varied substantially across the five domains (Figure 4.3): at baseline 93% reported no problems with self-care whereas only 59% reported no problems with anxiety/depression; similar patterns of reporting were seen at EOT and EOT+12/RT D0.

There was no evidence of a difference in the responses to the domain questions and the VAS score reported by participants at all three visits between those who cured on first-line treatment and those who experienced treatment failure and between those who received 8 week fixed length treatment and those who received variable 4-7 week treatment (p>0.12). Given the number of tests performed, there was weak evidence that participants randomised to VUS2 were more likely to report no problems with self-care and pain at EOT+12/RT D0 than those who received VUS1 (p=0.03 and p=0.04 respectively). Similarly, those who received ribavirin were more likely to report problems with performing their usual activities at EOT (p=0.02), and less likely to report problems with mobility and pain at EOT+12/RT D0 than those who did not receive ribavirin (p=0.03 and p=0.04 respectively). There was no evidence of differences between those randomised under VUS1 or VUS1, and those receiving ribavirin or not in other domains or on the VAS (p>0.13).

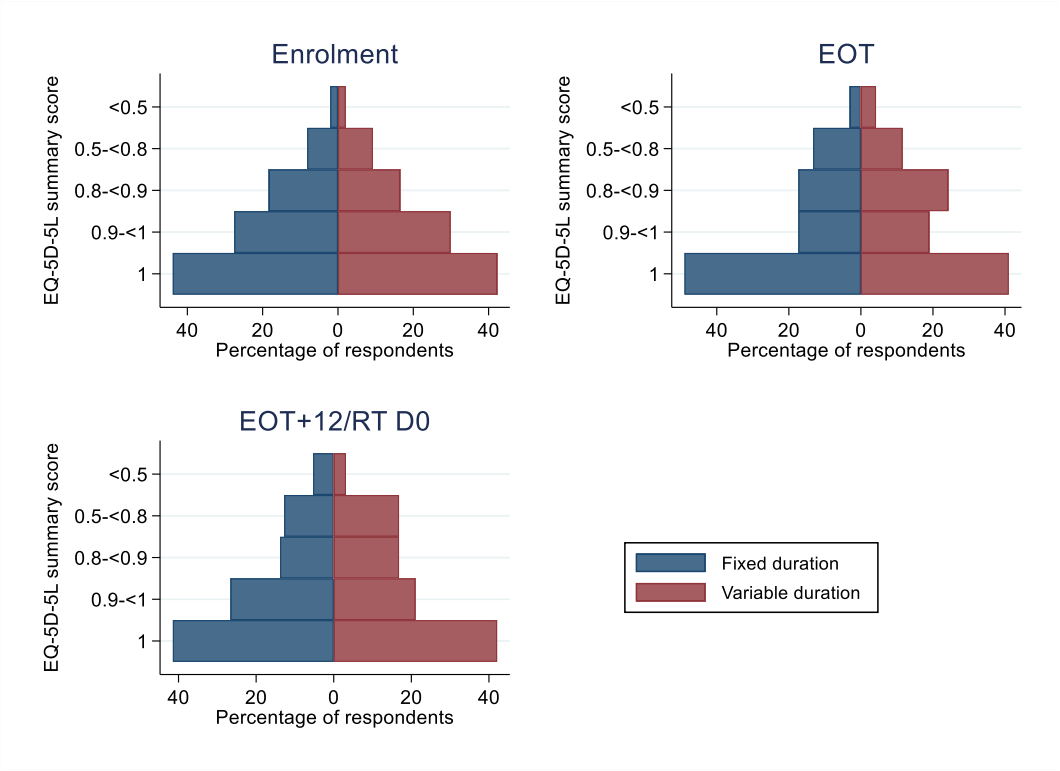


Figure 4.1: Distribution of EQ-5D-5L index scores by treatment duration randomisation

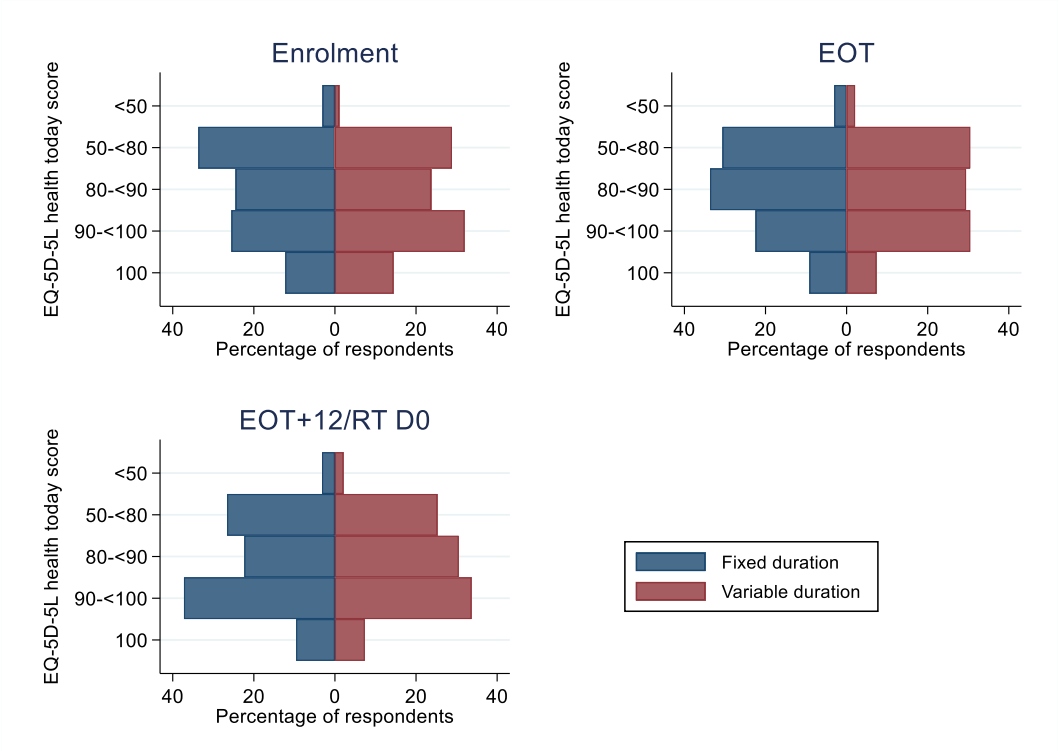


Figure 4.2: Distribution of EQ-5D-5L visual analogue scores by treatment duration randomisation

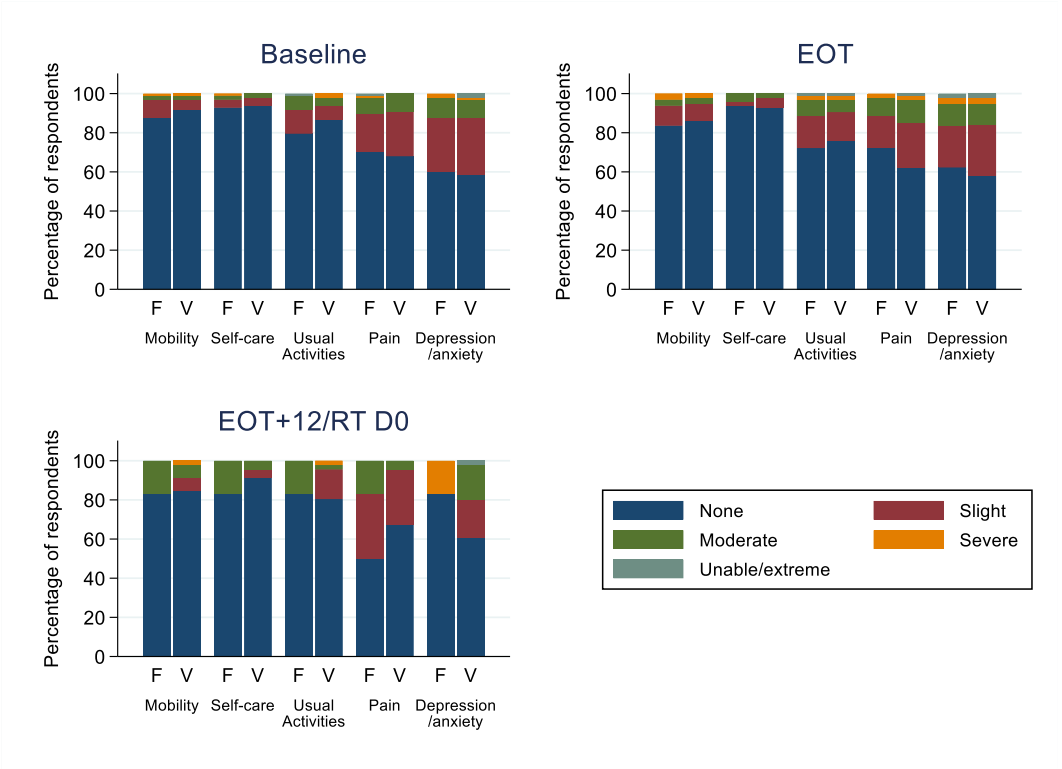


Figure 4.3: EQ-5D-5L domain responses in those received fixed (F) and variable (V) DAA treatment

On the MOS-Cog questionnaire, most participants reported either having no problems or having problems a little bit of the time (Table 8.2, Figure 4.4, Figure 4.5). Fewer participants reported perfect health on the MOS-Cog than with the EQ-5D-5L with only 19-30% reporting perfect health, however QoL was still moderately high with median (IQR) scores of 90 (70, 100), 87 (73, 97) and 90 (73, 100) at baseline, EOT and EOT/RT D0 respectively. There was some variability in the responses to each of the questions; at baseline, the lowest percentage reporting no problems was 39% reporting no problems with forgetting things, and the highest was 56% reporting no problems with reacting slowly to things.

There was no evidence of a difference in responses to questions at all time points between those who cured and those who failed on first-line treatment, those who took 8 week treatment and those who took 4-7 week treatment, and those who received ribavirin and those who did not ($p>0.08$). At EOT and EOT+12/RT D0, there was no evidence of difference between those who took VUS1 and VUS2 ($p>0.11$) but there were some differences at baseline: those treated using VUS2 were more likely to report higher levels of difficulty with reasoning and reacting slowly ($p=0.049$ and $p=0.002$ respectively), although numerically similar percentages reported having no problems.

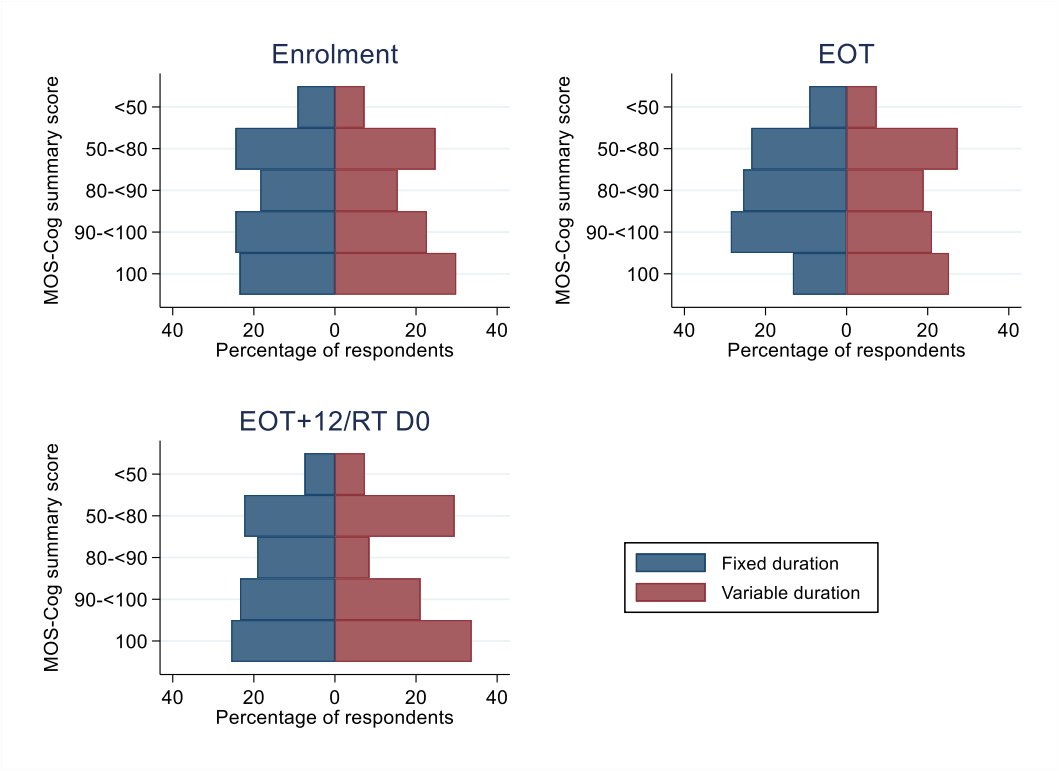


Figure 4.4: Distribution of MOS-Cog summary scores by treatment duration randomisation

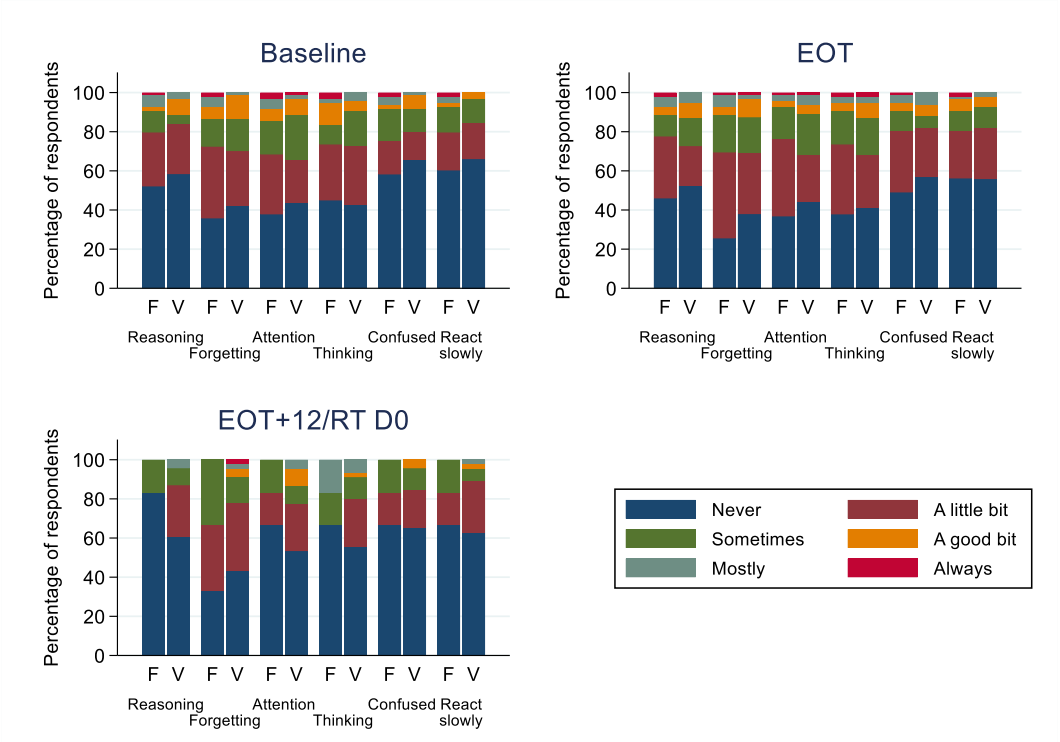


Figure 4.5: Responses to MOS-Cog questions in those who received fixed (F) and variable (V) DAA treatment

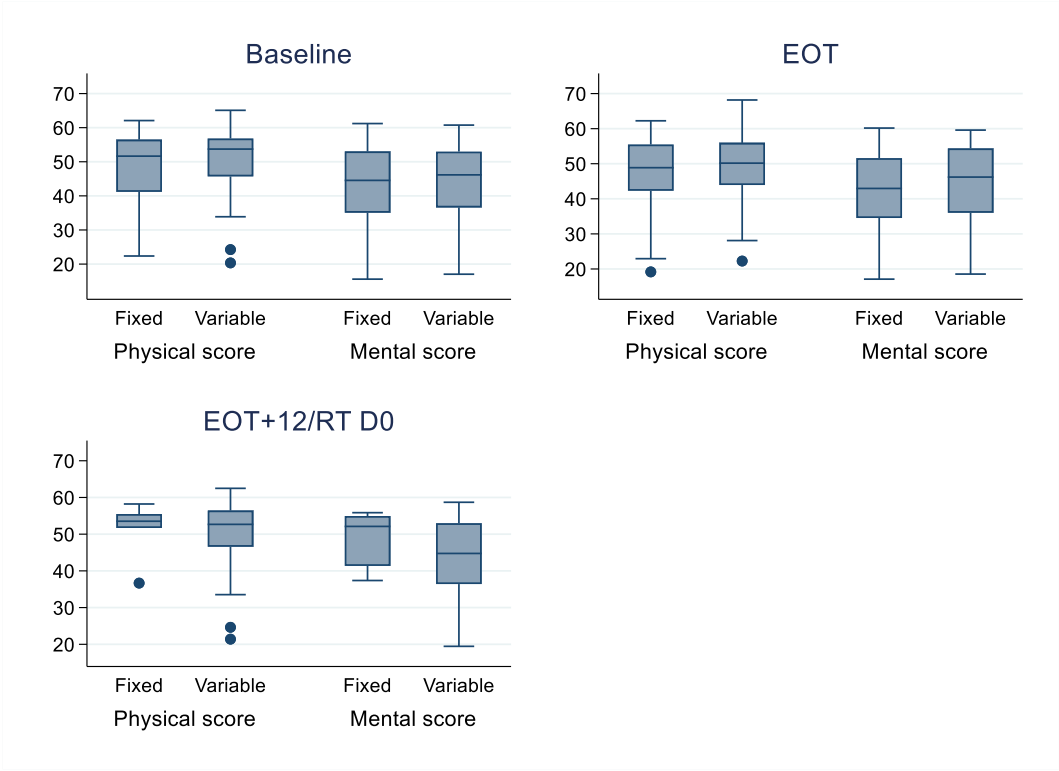


Figure 4.6: Distribution of SF-12 summary scores by treatment duration randomisation

On the SF-12 questionnaire (Table 8.3, Figure 4.6), the median (IQR) PCS score was 53 (44, 57), 50 (43, 56) and 53 (45, 56) at baseline, EOT and EOT+12/RT D0 respectively. The MCS score was slightly lower than the PCS at all visits with a median (IQR) MCS of 45 (35, 53), 45 (35, 53) and 45 (36, 53) at baseline, EOT and EOT+12/RT D0 respectively. At all three visits, there was no evidence of a difference in the MCS and PCS scores across all comparisons.

4.4.2 Effect of short-course treatment on QoL at EOT and EOT+12/RT D0 (multivariable comparisons)

Adjusting for VUS1/VUS2, ribavirin and SVR12, at EOT, QoL scores were numerically similar or slightly worse in most areas in those taking 4-7 weeks variable length treatment compared with those taking 8 weeks fixed length treatment (Figure 4.7a). While there was no evidence of a difference between the two groups in any area (p>0.10), there was a large numerical difference in reporting problems with self-care (#2 on figure) and in the proportion reporting perfect health on the VAS (#8 on figure) with participants receiving 4-7 weeks of treatment reporting worse QoL in both instances (proportional odds ratio (pOR) 0.34 (95% CI 0.04, 2.93) p=0.33 and odds ratio (OR) 0.42 (95% CI 0.13, 1.40) p=0.16, respectively).

In the equivalent fully adjusted models, at EOT+12/RT D0, there was no evidence of a difference (p>0.18) between the two groups in any area nor were there any large numerical differences of potential interest (Figure 4.7b). However, whereas at EOT those taking 4-7

weeks' DAAs had numerically similar or slightly worse QoL scores than those taking 8 weeks, at EOT+12/RT D0 some areas showed numerical higher scores compared with those taking 8 weeks.

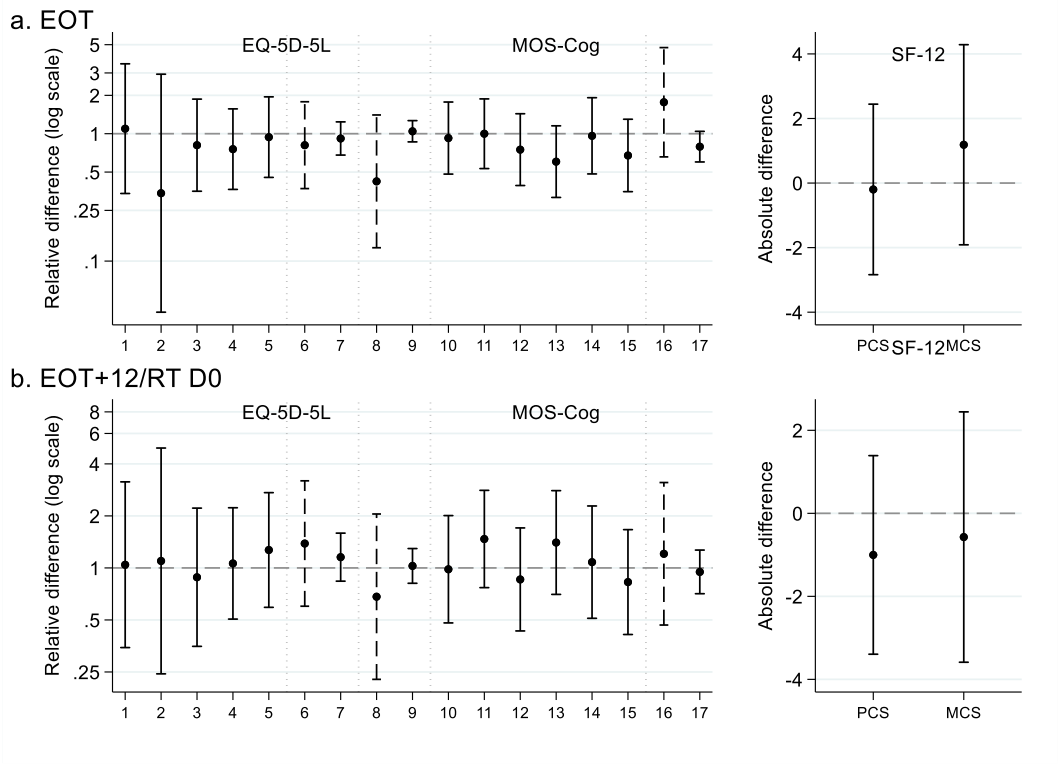


Figure 4.7: Difference (95% confidence interval) in quality of life scores at (a) EOT and (b) EOT+12/RT D0 between those taking variable 4-7 week treatment and 8 week treatment (relative difference >1 and absolute difference >0 means higher QoL at the relevant time point in those who took variable treatment)

Key: EQ-5D-5L: 1: mobility; 2: self-care; 3: usual activities; 4: pain; 5: anxiety/depression; 6: index score perfect health vs not; 7: index score continuous scale <1; 8: VAS score perfect health vs not; 9: VAS score continuous scale <100; reasoning. MOS-Cog: 10: reasoning; 11: forgetting; 12: attention; 13: thinking; 14: confusion; 15: reacting slowly; 16: summary score perfect health vs not; 17: summary score continuous scale <100.
Note: For 1-5 and 10-15, relative differences are proportional odds ratios, for 6 and 16 are odds ratios and for 7 and 17 are exponentiated estimates from beta regression (denoted "beta ratios" in main text). Differences greater than 1 for the EQ-5D-5L and MOS-Cog scores and greater than 0 for SF-12 scores represent better QoL for those who took variable length treatment.

4.4.3 Effect of VUS1/VUS2 on QoL at EOT and EOT+12/RT D0 (multivariable comparisons)

Using the same fully adjusted models, at EOT, QoL scores were broadly numerically similar between those randomised under VUS2 and those randomised under VUS1: there was no evidence of a difference in any area (p>0.10) and numerical differences did not generally favour one group (Figure 4.8a). There was a large numerical difference in the proportion reporting perfect health on the VAS (#8 on figure), with those randomised under VUS2 having

over double the odds of reporting perfect health than those randomised to VUS1 (OR 2.19 (95% CI 0.73, 6.95) p=0.19).

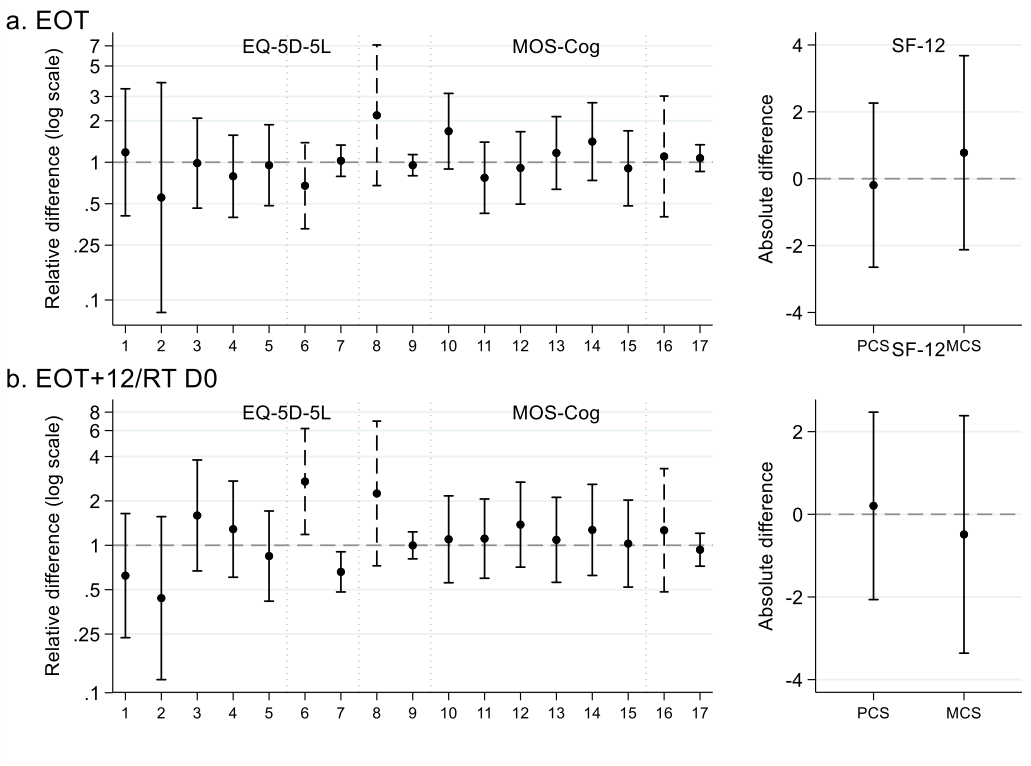


Figure 4.8: Difference (95% confidence interval) in quality of life scores at (a) EOT and (b) EOT+12/RT D0 between those randomised under VUS1 and those randomised under VUS2 (relative difference >1 and absolute difference >0 means higher QoL at the relevant time point in those randomised under VUS1)

Key: EQ-5D-5L: 1: mobility; 2: self-care; 3: usual activities; 4: pain; 5: anxiety/depression; 6: index score perfect health vs not; 7: index score continuous scale <1; 8: VAS score perfect health vs not; 9: VAS score continuous scale <100; reasoning. MOS-Cog: 10: reasoning; 11: forgetting; 12: attention; 13: thinking; 14: confusion; 15: reacting slowly; 16: summary score perfect health vs not; 17: summary score continuous scale <100.
Note: For 1-5 and 10-15, relative differences are proportional odds ratios, for 6 and 16 are odds ratios and for 7 and 17 are exponentiated estimates from beta regression (denoted "beta ratios" in main text). Differences greater than 1 for the EQ-5D-5L and MOS-Cog scores and greater than 0 for SF-12 scores represent better QoL for those who were randomised under VUS2.

In the equivalent fully adjusted models, at EOT+12/RT D0, QoL scores were either numerically similar or higher in those receiving VUS2, apart from some parts of the EQ-5D-5L (Figure 4.8b). There was evidence of a difference between VUS1 and VUS2 in the EQ-5D-5L index score, which summarises the 5 domains, with those receiving VUS2 more likely to have a perfect health index score of 100 than those receiving VUS1 (#6 on figure; OR 2.71 (95% CI 1.18, 6.19) p=0.02), but in those that did not have a perfect health score, VUS2 were more likely to have a lower score (#7 on figure; beta ratio 0.66 (95% CI 0.48, 0.90) p=0.01). Overall this indicates that

those receiving VUS2 were more likely to have an extreme score compared to VUS1 participants. Similarly, a numerically higher proportion of participants receiving VUS2 reported perfect health on the VAS than those receiving VUS1 (#8 on figure; OR 2.25 (95% CI 0.73, 6.95) p=0.16). There was also a large numerical difference in problems with self-care (#2 on figure), with those receiving VUS2 having less than half the odds of reporting fewer or no problems than those receiving VUS1 (pOR 0.44 (95% CI 0.12, 1.56) p=0.20).

4.4.4 Effect of ribavirin on QoL at EOT and EOT+12/RT D0 (multivariable comparisons)

Using the same fully adjusted models, at EOT, the QoL scores between those who took ribavirin and those that did not were numerically similar in most areas with no clear distinction between the two groups (Figure 4.9a). The only area in which there was evidence of a difference was in performing usual activities (#3 on figure), where those who took ribavirin had over 60% lower odds of reporting low levels of or no problems (pOR 0.39 (95% CI 0.19, 0.81) p=0.01) than those who did not. However, there was a trend for participants who took ribavirin to report fewer problems with forgetting things (#11 on figure; pOR 1.67 (95% CI 0.97, 2.89) p=0.07). There was also a trend for the PCS and MCS scores on the SF-12 to be lower in those who took ribavirin with a PCS score -1.91 points lower (95% CI -4.16, 0.35; p=0.10) and a MCS score -2.44 points lower (95% CI -5.10, 0.21; p=0.07) in those taking ribavirin.

In the equivalent fully adjusted models, at EOT+12/RT D0, there was no evidence of a difference in QoL between those who took ribavirin and those that did not in any area (Figure 4.9b). However, there was a trend for those who received ribavirin to be more likely to report fewer problems with anxiety or depression than those who did not (#5 on figure; pOR 1.76 (95% CI 0.91, 3.37) p=0.09) and for all other domains of the EQ-5D-5L (#1-4 on the figure) there were large numeric differences between the two groups, whereas the QoL as reported on the MOS-Cog and SF-12 questionnaires was numerically more similar in both groups. There was a particularly large difference in those reporting problems with self-care (#2 on figure), with participants who took ribavirin having over three times the odds (pOR 2.94 (95% CI 0.79, 10.99) p=0.11) of reporting fewer or no problems than those who did not take ribavirin. Also of note, although taking ribavirin was associated with more problems with performing usual activities at EOT, the effect of ribavirin had reversed at EOT+12/RT D0 and those who took ribavirin reported fewer problems performing usual activities (pOR 1.87 (95% CI 0.88, 3.98) p=0.11), although this was only a numeric difference.

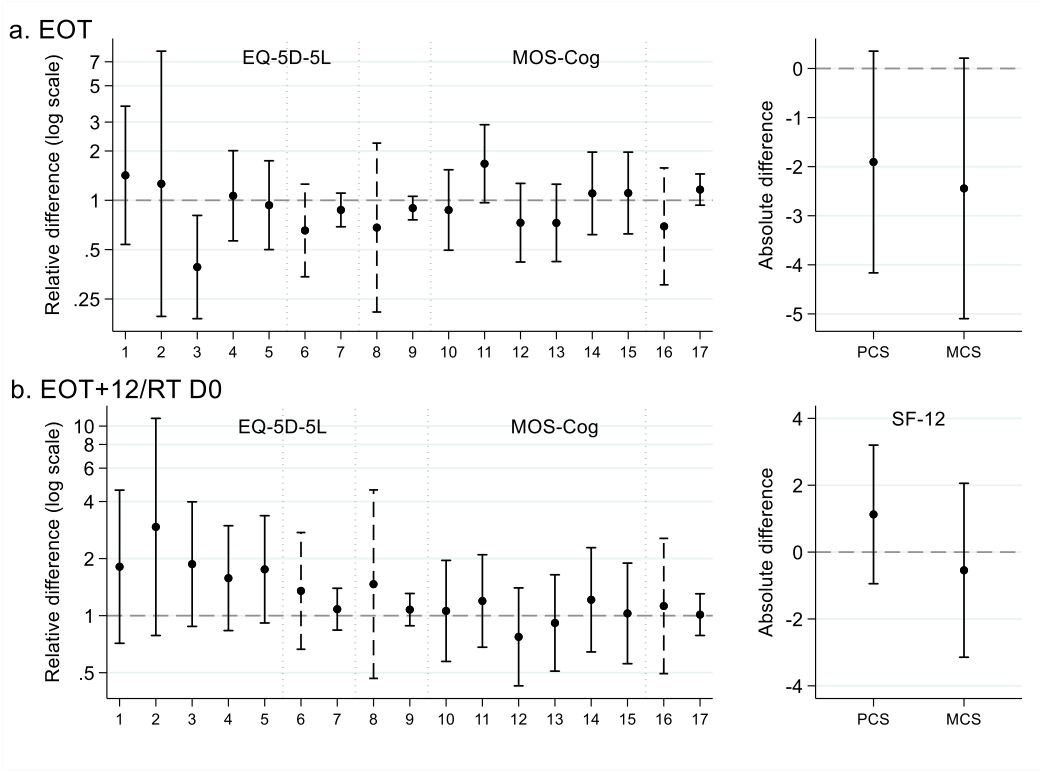


Figure 4.9: Difference (95% confidence interval) in quality of life scores at (a) EOT and (b) EOT+12/RT D0 between those who received ribavirin and those that did not (relative difference >1 and absolute difference >0 means higher QoL at the relevant time point in those who took ribavirin)

Key: EQ-5D-5L: 1: mobility; 2: self-care; 3: usual activities; 4: pain; 5: anxiety/depression; 6: index score perfect health vs not; 7: index score continuous scale <1; 8: VAS score perfect health vs not; 9: VAS score continuous scale <100; reasoning. MOS-Cog: 10: reasoning; 11: forgetting; 12: attention; 13: thinking; 14: confusion; 15: reacting slowly; 16: summary score perfect health vs not; 17: summary score continuous scale <100.
Note: For 1-5 and 10-15, relative differences are proportional odds ratios, for 6 and 16 are odds ratios and for 7 and 17 are exponentiated estimates from beta regression (denoted "beta ratios" in main text). Differences greater than 1 for the EQ-5D-5L and MOS-Cog scores and greater than 0 for SF-12 scores represent better QoL for those who took ribavirin.

4.4.5 Effect of cure on QoL at EOT and EOT+12/RT D0 (multivariable comparisons)

Using the same fully adjusted models, at EOT, those who cured had either a numerically similar or worse QoL, albeit with only a numerical difference rather than with strong supporting statistical evidence, for most areas (Figure 4.10a). The only area with evidence of a difference between the two groups was problems with thinking (#13 on figure); those who cured were less likely to report fewer problems (OR 0.43 (95% CI 0.20, 0.93) p=0.03). There was a large numeric difference in those reporting perfect health on the EQ-5D-5L VAS; those who cured had under half the odds of reporting perfect health than those who did not (#8 on figure; OR 0.46 (95% CI 0.14, 1.54) p=0.21).

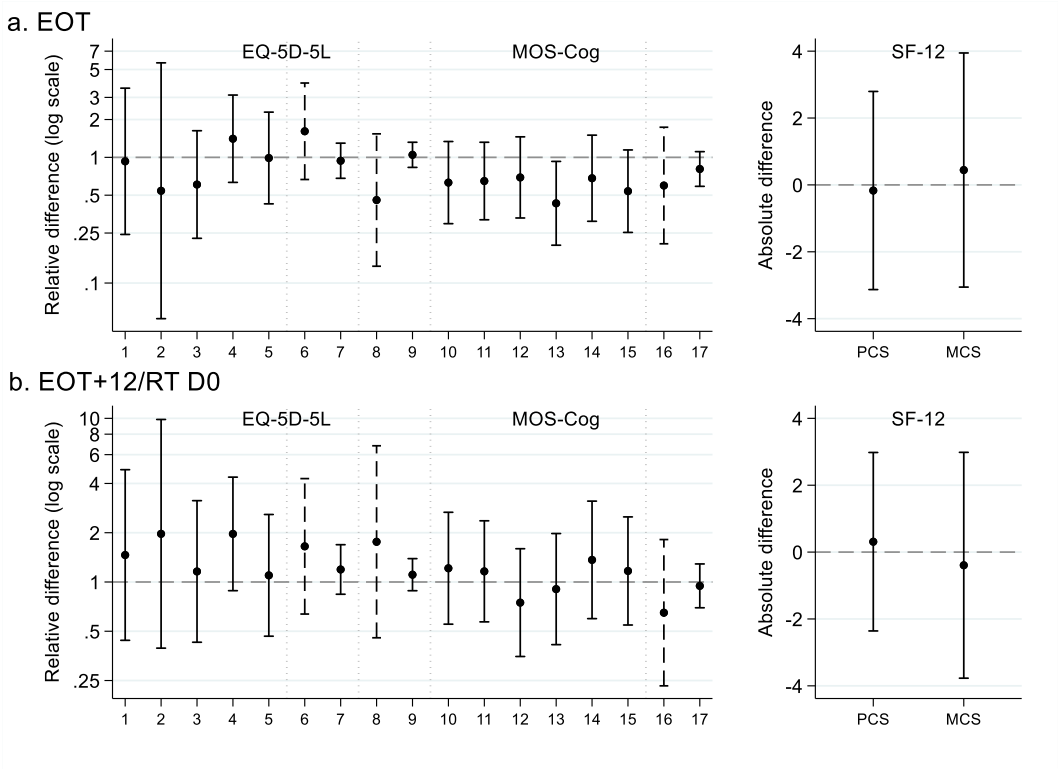


Figure 4.10: Difference (95% confidence interval) in quality of life scores at (a) EOT and (b) EOT+12/RT D0 between those who cured and those who failed first-line treatment (relative difference >1 and absolute difference >0 means higher QoL at the relevant time point in those who cured)

Key: EQ-5D-5L: 1: mobility; 2: self-care; 3: usual activities; 4: pain; 5: anxiety/depression; 6: index score perfect health vs not; 7: index score continuous scale <1; 8: VAS score perfect health vs not; 9: VAS score continuous scale <100; reasoning. MOS-Cog: 10: reasoning; 11: forgetting; 12: attention; 13: thinking; 14: confusion; 15: reacting slowly; 16: summary score perfect health vs not; 17: summary score continuous scale <100.
Note: For 1-5 and 10-15, relative differences are proportional odds ratios, for 6 and 16 are odds ratios and for 7 and 17 are exponentiated estimates from beta regression (denoted "beta ratios" in main text). Differences greater than 1 for the EQ-5D-5L and MOS-Cog scores and greater than 0 for SF-12 scores represent better QoL for those who cured.

In contrast, in equivalent adjusted models, at EOT+12/RT D0, those who cured had a numerically similar or better QoL than those who failed in most areas (Figure 4.10b). There was no evidence of a difference at EOT+12/RT D0 between those who cured and those who did not in any area (adjusted p>0.10), however there was a large numeric difference in problems with pain with those who cured having almost double the proportional odds of reporting less pain than those who did not (#4 on figure; pOR 1.97 (95% CI 0.89, 4.37) p=0.10). There were an additional two areas that had large numeric differences at EOT+12/RT D0: people who cured had almost double the proportional odds of reporting fewer problems with self-care (pOR 1.97 (95% CI 0.39, 9.84) p=0.41) and also much higher odds of reporting perfect health on the VAS (OR 1.76 (95% CI 0.46, 6.80) p=0.41).

4.4.6 Effect of anaemia on QoL

In total 49 (25%) participants had anaemia at EOT, of which 42 were taking ribavirin and 7 were not; 16 participants had grade 1 anaemia at a previous visit, but this had resolved by EOT. The majority of participants (n=45) had grade 1 anaemia (10-12g/dl or 13g/dl in women and men respectively), three had grade 2 anaemia (9.6, 9.8, 9.9g/dl) and one had grade 3 anaemia (7.4g/dl). Contrary to expectation, in adjusted models, having anaemia of any grade was associated with higher QoL in two areas of the MOS-Cog questionnaire and the corresponding summary score at EOT (Figure 4.11). Specifically, those with anaemia had over double the odds of having fewer problems with reasoning (#10 on figure; pOR 2.73 (95% CI 1.28, 5.82) p=0.01)) and with reacting slowly (#15 on figure; pOR 2.37 (95% CI 1.08, 5.23) p=0.03). There was no evidence of difference between those with and without anaemia for having a perfect MOS-Cog summary score (p=1.00), but in those with a summary score <1 (#17 on figure) having anaemia was associated with a 35% higher summary score (95% CI 1.06, 1.72; p=0.02).

Compared to the model not adjusting for anaemia, there remained no evidence of association between ribavirin and each of the QoL outcomes for which there was originally no association (p>0.1). The effect of ribavirin on performing usual activities was similar both when adjusting for anaemia and when not (pOR 0.40 (95% CI 0.18, 0.88) p=0.02 after adjustment for anaemia), suggesting that it was ribavirin and not any related anaemia that impacted participants. There was originally a trend for those who took ribavirin to report fewer problems with forgetting things, but there was no evidence for this after adjusting for anaemia (p=0.32). There was also no evidence that anaemia was associated with fewer problems with forgetting things, however the numeric difference was large (pOR 1.75 (95% CI 0.87, 3.55) p=0.12).

Having anaemia was associated with a lower PCS score at EOT, with participants with anaemia at EOT having a PCS 3.65 points lower (95% CI 0.77, 6.54; p=0.01) than those who did not. There was no evidence of a difference in PCS score between those who took ribavirin and those did not (p=0.61) after adjusting for anaemia, suggesting that the trend towards a lower PCS score in those who took ribavirin when not adjusting for anaemia is likely due to ribavirin-related anaemia and not the direct effect of ribavirin itself.

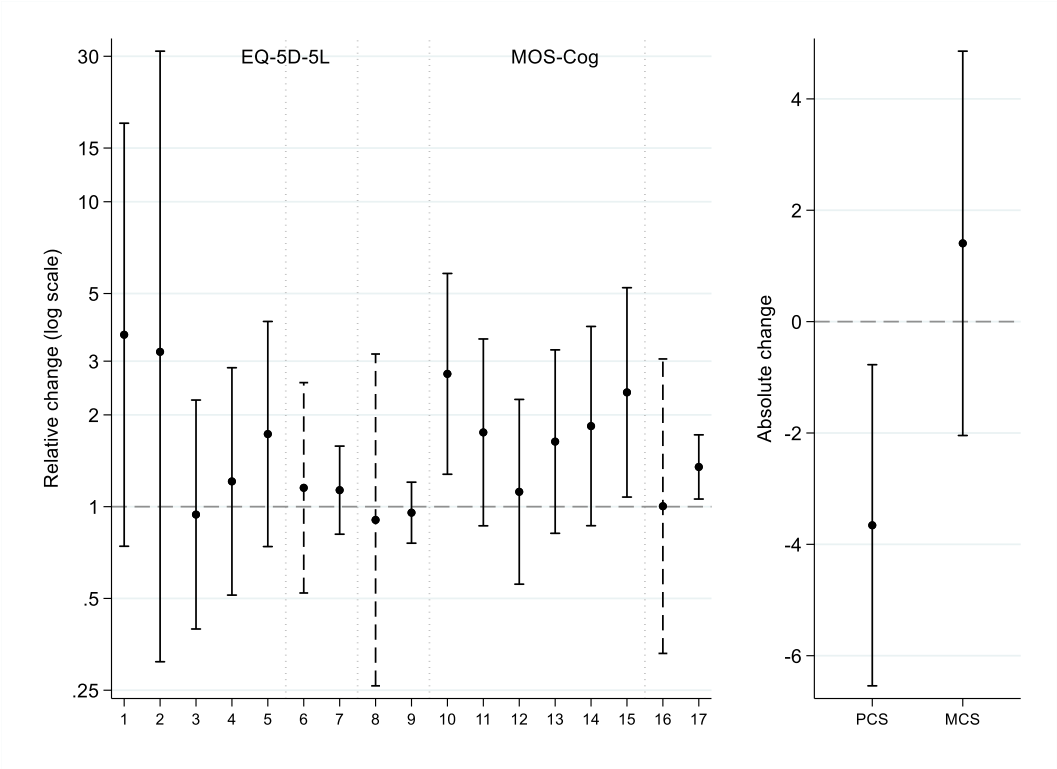


Figure 4.11:Difference (95% confidence interval) in quality of life scores at (a) EOT and (b) EOT+12/RT D0 between those with anaemia and those without (relative difference >1 and absolute difference >0 means higher QoL at the relevant time point in those with anaemia)

Key: EQ-5D-5L: 1: mobility; 2: self-care; 3: usual activities; 4: pain; 5: anxiety/depression; 6: index score perfect health vs not; 7: index score continuous scale <1; 8: VAS score perfect health vs not; 9: VAS score continuous scale <100; reasoning. MOS-Cog: 10: reasoning; 11: forgetting; 12: attention; 13: thinking; 14: confusion; 15: reacting slowly; 16: summary score perfect health vs not; 17: summary score continuous scale <100.
Note: For 1-5 and 10-15, relative differences are proportional odds ratios, for 6 and 16 are odds ratios and for 7 and 17 are exponentiated estimates from beta regression (denoted "beta ratios" in main text). Differences greater than 1 for the EQ-5D-5L and MOS-Cog scores and greater than 0 for SF-12 scores represent better QoL for those with anaemia.

While there was no evidence of association between anaemia and the MCS score (p=0.42), after controlling for anaemia at EOT the effect of taking ribavirin strengthened and there was more evidence that taking ribavirin was associated with a lower MCS (by -2.94 points (95% CI -5.86, -0.02) p=0.049), independently of any effect from anaemia.

There were also large numerical differences between those with and without anaemia at EOT in problems with mobility and self-care (#1 and 2 on figure, respectively); participants with anaemia had over three times the odds of reporting problems than those without anaemia (pOR 3.66 (95% CI 0.74, 18.09) p=0.13 and 3.22 (95% CI 0.31, 31.18) p=0.33, respectively). Despite the large differences in some EQ-5D-5L domains, there was no evidence of differences

in the summary index score between those with and without anaemia (p=0.73 and p=0.46 for having a perfect summary score and for a difference in non-perfect scores, respectively).

4.5 Discussion

Overall, there were few associations with QoL and the length of treatment, taking ribavirin and SVR12 (Table 4.1) and QoL was numerically similar between these different groups in many areas. Using a relatively generous threshold for statistical significance given the number of tests performed (p=0.05), at EOT, achieving SVR12 was associated with more problems with thinking and taking ribavirin was associated with more problems with performing usual activities. At EOT+12/RT D0, those taking VUS2 instead of VUS1 were more likely to have either perfect MOS-Cog scores or, if not perfect, lower summary scores. Additionally, for participants who took ribavirin there were trends for fewer problems with forgetting and lower PCS and MCS scores at EOT, and fewer problems with anxiety or depression at EOT+12/RT D0. However, overall power was low with only 192 participants contributing to the analysis; therefore, particularly for the EQ-5D-5L and MOS-Cog domain questions and for having a perfect health score on the summary and VAS scores, large numerical differences are also of interest.

In terms of these numeric differences which failed to reach statistical significance, at EOT, participants who took 4-7 weeks' treatment reported more problems with self-care and fewer participants reported a perfect VAS score than those who took 8 weeks' treatment, but at EOT+12/RT D0 there were no differences of note. Higher proportions of those who received VUS2 reported perfect VAS scores at both EOT and EOT+12/RT D0 than those who received VUS1, but higher numbers also reported more problems with self-care at EOT+12/RT D0. At EOT+12/RT D0, in contrast to EOT, those taking ribavirin reported fewer problems with performing usual activities than those who did not take ribavirin, and they also reported fewer problems with self-care. There were no large differences between those who cured and those who failed at EOT, but at EOT+12/RT D0 participants who cured reported fewer problems with self-care and pain, and a higher proportion reported perfect VAS scores.

Having anaemia was associated with some higher cognitive scores, but a lower PCS score at EOT. There were also some large differences in some EQ-5D-5L domains, particularly mobility and self-care, with, contrary to what might be expected, those with anaemia reporting fewer problems. When adjusted for anaemia, the effect of ribavirin on the MCS was stronger and led to a lower MCS score being associated with taking ribavirin. There was no longer any evidence of an effect of ribavirin on the PCS or having problems with forgetting after adjusting for anaemia, suggesting that anaemia may be mediating these ribavirin-associated effects, but the

effect of ribavirin on problems performing usual activities was unaltered suggesting this could have a different pathway.

Table 4.1: Summary of QoL results

Instrument	Domain	EOT: QoL worse in	EOT+12/RT D0: QoL worse in
EQ-5D-5L	Self-care	**Variable (vs fixed)**	**VUS2 (vs VUS1)** **No RBV (vs RBV)** **Failed (vs cured)**
EQ-5D-5L	Pain		VUS1 (vs VUS2) RBV (vs no RBV) **Failed (vs cured)**
EQ-5D-5L	Usual activities	RBV (vs no RBV)	**No RBV (vs RBV)**
EQ-5D-5L	Mobility		RBV (vs no RBV)
EQ-5D-5L	Anxiety/depression		*No RBV (vs RBV)*
EQ-5D-5L	VAS: perfect score	**Variable (vs fixed)** **VUS1 (vs VUS2)** **Cured (vs failed)**	**VUS1 (vs VUS2)** **Failed (vs cured)**
Mos-Cog	Thinking	Cured (vs failed)	
Mos-Cog	Forgetting	*No RBV (vs RBV)**	
Mos-Cog	Summary score: perfect health		VUS1 (vs VUS2)
MOS-Cog	Summary score: non-perfect health		VUS2 (vs VUS1)
SF-12	PCS	*RBV (vs no RBV)**	
SF-12	MCS	*RBV (vs no RBV)***	

*weakens after adjusting for anaemia at EOT
**strengthens after adjusting for anaemia
Note: grey means only in unadjusted comparisons; black means in adjusted comparisons or both unadjusted and adjusted; italics means trend (p 0.05-0.1); bold means large numeric differences (relative difference <0.5 or >1.5).

While low power limited the detection of true differences between the groups, the very large number of statistical tests carried out also increased the risk of erroneously finding an effect that does not truly exist. For the main comparisons alone, 152 tests were carried out (2 time points, 4 comparisons, 19 regressions (one for each of the different QoL areas compared)),

with another 336 (3 time points, 4 comparisons, 28 QoL scores) for the simple univariable comparisons carried out in the first part of the analysis and 95 (76 as before plus anaemia in each of the 19 regressions at EOT) for the analysis examining the effect of anaemia, giving 583 tests in total. At a 5% significance level, I would therefore expect to find 8 results from the main comparisons as significant even if there was no genuine difference between any factor assessed (=0.05*152), 29 overall (=0.05*583). The fact that I found fewer than this in total, and none significant at a 1% level, suggests that many of the findings above could be simply due to chance.

Few differences were found between those taking 4-7 weeks' variable length treatment and those taking 8 weeks' fixed length treatment, even when considering numerical differences. It is possible that differences were not observed as the treatment lengths between the two groups were too similar: while 23 (24%) participants took 4 weeks' treatment exactly, 33 (34%) took between 4 and 5 weeks, 41 (42%) took between 5 and 6 weeks and 11 (11%) took over 6 weeks. Additionally, for the fixed duration regimen, the 8 week treatment length was still shorter than the recommended treatment length of 12 weeks. It is possible that had those participants received the full 12 weeks, more differences might have been observed between the groups. Some numerical differences were seen at EOT, specifically more problems with self-care and lower levels of perfect VAS scores, which could be attributed to the burden of longer treatment, but these differences were no longer apparent at EOT+12/RT D0. At EOT+12/RT D0, regardless of the length of treatment a participant received, it had been ~12 weeks since most participants completed treatment (median [range] weeks from EOT to starting retreatment was 11 [6, 42]). Any short-term benefits of a shorter treatment length would not only have receded and likely forgotten, but participants would have been more focussed on their probable cure or start of retreatment.

Similarly, few differences in QoL were observed at EOT between participants who received the VUS1 strategy and those who received the VUS2 strategy. This is also likely to be due to the small number of days difference between the two strategies; the difference in length of an individual's treatment between VUS1 and VUS2 was between 0-7 days depending on baseline HCV VL. However, there were some differences between the groups at EOT+12/RT D0 once adjusted for other factors, including a higher proportion of perfect scores on the EQ-5D-5L index and VAS, but worse non-perfect index scores and more problems with self-care. VUS2 was implemented part way through the trial after a DMC meeting reviewing interim results so there may also have been temporal effects, such as participants being recruited from a different mix of study centres (supported by the small differences in baseline scores between

VUS2 and VUS1), that were either never measured or not have not been accounted for in the analysis and that may have had an impact on the results.

In previous studies, taking ribavirin was associated with worse QoL during treatment compared with both baseline and those not taking ribavirin, but by EOT+12 the negative effects had usually ended and QoL was back to baseline or higher (90, 100, 104, 105). The results from this analysis also show the negative QoL reported by participants at EOT no longer being reported at EOT+12, but overall QoL was not lower at EOT in most areas compared to those not taking ribavirin, the exception being usual activities (with a similar trend for PCS and MCS). Some studies investigating the impact of ribavirin did see a return to baseline QoL by EOT (90, 104), so it is possible that this was also the case in STOP-HCV-1 and more negative impacts on QoL would have been detected if the questionnaire had been completed by participants sooner after starting treatment. It is also possible that patients are able to tolerate ribavirin better if it is given for short-course treatment (4-8 weeks in STOP-HCV-1). There are concerns about patients' tolerance of ribavirin in treating HCV, but if it is combined with a short-course DAA treatment, and so limiting the time during which QoL could reduce, with a quick rebound in QoL after completing treatment, it may be more acceptable. Although the analysis showed a trend towards fewer problems with forgetting in those taking ribavirin at EOT, this is unlikely to be caused by the treatment and is most likely therefore due to chance, especially as previous research has shown that when given with PEG-IFN, ribavirin was associated with depression and neurocognitive decline (107-109). However, I was not able to find literature focusing on ribavirin alone or in conjunction with DAAs to compare against the effects when given with PEG-IFN and it is possible that these negative mental effects are due to PEG-IFN alone.

At EOT, there were almost no differences, statistical or numerical, that indicated better QoL for those who cured. Although SVR12 is known to improve QoL in previous studies compared with those who failed, the benefits of this may not have been likely to be felt at EOT as treatment length was relatively short and not all participants had virological suppression at that time point (15% of participants in each group had detectable viral load at EOT). There was also statistical evidence of an association between achieving cure and having problems with thinking, which may simply be due to chance because of the large number of statistical tests performed, but all areas of the MOS-Cog had numerically worse scores for those with SVR12. One possible explanation is that, despite similar rates of virological suppression at EOT, those who later cured had a stronger reaction to the DAAs than those with subsequent treatment failure which negatively impacted their QoL at EOT.

At EOT+12/RT D0, most numerical, and all large, differences were in favour of those who cured. At this time point, the viral load of those who cured would have been suppressed for a substantial amount of time and the health benefits of cure may have been felt, while the viral load of those with treatment failure would have rebounded. As well as the virological differences, participants in the two groups would have been in very different stages of the trial: those who cured would be expecting to hear so after the visit in which the questionnaire was completed (and only have one more visit left in the trial at EOT+24), while those who failed on first-line treatment were starting retreatment with the prospect of taking another full course of treatment and having many further follow-up visits. It is likely the difference in expectations impacted how the participants felt about their own health and QoL. While there was no statistical evidence of a difference between the two groups, this could be due to a lack of power or because the participants QoL was not initially poor at baseline, so any changes were very small.

Although the majority of previous research has shown that SVR12 does lead to improved QoL (91-98, 101-104), this is often compared to a patient's baseline and not against a similar group who did not achieve SVR12. In comparison with previous studies, baseline QoL scores tended to be higher amongst the STOP-HCV-1 participants in all or most areas, so the potential for any improvement may have been more constrained by other factors in a participant's life rather than their illness. There are likely several reasons why STOP-HCV-1 participants had higher baseline QoL scores. Participants agreed to take part in a clinical trial, whereas previous research included a mixture of clinical trials and observational studies, and an individual may have been less likely to agree to extra STOP-HCV-1 trial procedures if they did not feel well. As part of the inclusion criteria for STOP-HCV-1, participants had to have mild liver disease, which may also have contributed to higher QoL scores. As well as having higher baseline scores, participants who did not achieve SVR12 on first-line treatment were offered retreatment at the earliest opportunity (and all went on to cure on retreatment) so their QoL might not have had enough time to decline by the retreatment visit.

Anaemia, including mild anaemia, is usually associated with negative physical and cognitive symptoms in patients and the lower PCS and MCS scores at EOT observed in the participants receiving ribavirin, and the significantly lower PCS score in those with anaemia, within the trial corresponds with this. The higher cognitive scores and the large differences in mobility and self-care in those with anaemia appear to contradict this. While lack of differences between the groups may be explained by the large proportion of grade 1 anaemias that were likely asymptomatic, it is unclear why those with anaemia reported fewer problems in these areas.

After adjusting for anaemia, the effect of ribavirin on the MCS strengthened and ribavirin was associated with a lower MCS score than without adjusting for anaemia. This suggests that anaemia reduced the negative effects of ribavirin, at least partially, on mental health; this is supported by the higher MOS-Cog scores reported by those with anaemia. For the PCS, there was no evidence of an effect of ribavirin after adjustment for anaemia and, given the association between anaemia and PCS, it is likely that all or most of the effect of ribavirin is mediated through anaemia. Additionally, there was no evidence of an effect of ribavirin on problems with forgetting after adjusting for anaemia; there was also no evidence of an effect of anaemia, but there was a large numeric difference with those with anaemia reporting fewer problems. As for the effect on MCS and the higher scores on the other MOS-Cog areas, it is likely that the effect of ribavirin is at least partially mediated through anaemia, but power is too low to fully differentiate the effects between the two factors. There was no evidence that anaemia mediated the impact of ribavirin on performing usual activities as there was no change in the effect after adjustment for anaemia. This suggests that the reason participants had problems was likely due to mental health reasons and not physical health reasons as, from the effects of ribavirin and anaemia on PCS and MCS, it is likely that negative mental health effects from ribavirin are from the direct effects of the drug itself, while negative physical health effects are from ribavirin related anaemia.

This is the first known analysis investigating the impact of short-course treatment on QoL, also assessing whether the length of short-course treatment could also affect QoL by comparing VUS1 vs VUS2, and fixed vs variable length treatment overall. Previous studies of QoL in patients receiving DAA treatment were either limited to only those who achieved SVR12 or reported high levels of SVR12 (usually >95%), and so while the effect of SVR12 could be reported compared with baseline, QoL was rarely compared between those who did and did not cure. The overall SVR12 rate within this analysis was 71% so an additional strength of this work is that a direct comparison was made between those with SVR12 and those without, even if this still had limited power.

For the analysis examining the impact of anaemia, I chose to only use EOT haemoglobin in my definition of anaemia, that is, defined anaemia at the visit at which they completed the QoL surveys. Participants who were anaemic at previous visits (2-4 weeks prior to EOT depending on when the anaemia occurred and length of DAA treatment), but not by EOT, may no longer have felt any lasting impact on their QoL, particularly as all such anaemias were grade 1, and could have potentially biased the results towards those of participants with no recorded anaemia. Further work could investigate any lasting effect of anaemia while on treatment by including the anaemias that had resolved by EOT or by comparing QoL scores at EOT+12/RT D0

between those who were anaemic on treatment and those who were not. I also chose to examine the impact of anaemia as a binary condition and not haemoglobin levels. It is reasonable to assume that any effects of anaemia or haemoglobin on QoL are gradual and that the QoL between people with haemoglobin levels just above and just below the definition of anaemia might be very similar. However, few people (25%) had anaemia of any grade and an even smaller number had an anaemia of grade 2 or higher (2%). As the data is sparse for the levels of haemoglobin of most interest, any effect of haemoglobin that would be detected may not be generalisable for these higher levels of haemoglobin. To overcome this, any future work modelling of the effect of haemoglobin on QoL should consider non-linear forms or the use of splines.

One limitation of this analysis is the large number of statistical tests performed; I did not adjust for multiple testing and therefore the probability of discovering a result by chance is very high. No adjustment was made due to the relatively small sample sizes and the low risk of harm caused by finding an erroneous result.

Another potential limitation is that at EOT+12/RT D0, participants were either aware of their SVR12 status (failures) or are likely to be confident about it (cures). Interpretation of the results must take account of this, especially if interested in only the direct effects of treatment at that time point and not in how participants felt about their SVR12 status as this knowledge is likely to bias results towards a more positive or negative outlook for those who cured and failed respectively. However, it is also arguable that knowledge of SVR12 status helps in forming a participant's idea of their QoL and if the questionnaires were completed blind, this ignores the impact of achieving SVR12 or not on the participant's life.

Another limitation is with the QoL instruments themselves. Although these tools, particularly the EQ-5D-5L and SF-12, are widely used both to report the QoL of HCV patients and in other diverse areas, these instruments are not tailored towards HCV patients. Therefore, they may not reflect how HCV patients feel about their QoL, accurately detect changes that are clinically important to HCV patients or detect changes in areas that are most meaningful to them. While there are QoL tools that are specific to HCV patients (the Chronic Liver Disease Questionnaire (CLDQ) (110) and the Work Productivity and Activity Impairment: HCV (WPAI:HCV) (111), these were not used in the STOP-HCV-1 trial.

4.6 Key findings

- The length of treatment, whether 4-7 weeks vs 8 weeks or VUS1 vs VUS2, did not have a large impact on QoL.

- Participants taking ribavirin did not have substantially worse QoL at EOT than those not taking ribavirin and any problems were no longer reported at EOT+12/RT D0.
- Having anaemia (mostly grade 1) had a mixed effect on QoL scores,but mediated the negative effect of ribavirin on physical health.

5 Assessment of the statistical aspects of VIETNARMS: a complex multi-arm trial testing several drug shortening or sparing HCV treatments

5.1 Introduction and aims

In previous chapters I have investigated which factors are associated with cure and the mechanisms around treatment failure to identify which groups of patients might be most suitable for short-course therapy. I also examined QoL in participants taking different lengths of treatment and adjunctive ribavirin to assess the impact of these treatment choices, which could help determine which treatment lengths and regimens might be most acceptable for patients. From this work I have identified baseline viral load as the key factor in determining SVR12 and that QoL is not clearly impacted by length of treatment or from adjunctive ribavirin. However, while valuable insights can be gained from previously collected data, it is also important to be able to test selected drug-shortening strategies in clinical trials.

The overall aim of this chapter was to examine the statistical aspects of VIETNARMS, which planned to test multiple strategies of shortening or sparing HCV treatment. The trial design of VIETNARMS is described in more detail in the methods section, but it was a complex design due to having two factorial and one partial factorial randomisations. There was a Bayesian monitoring mechanism within the trial to allow for unsuccessful treatment strategies to be detected early and for randomisation into these strategies to be stopped.

While factorial trials and Bayesian monitoring are not new techniques, at the time of performing this research, there were no records of previous trials that employed Bayesian monitoring within a complex factorial trial. Therefore, there was no previous research into how the trial design would work in practice and whether it would be appropriate for the aims of the trial. The overall aim of this chapter was to answer these questions for the VIETNARMS trial.

My first aim of this chapter was to assess the operating characteristics of the Bayesian stopping guideline used in the trial analytically and through simulation. This was to ensure that the guideline worked as intended and that failing strategies could be detected and stopped early enough in the trial to be a benefit. It was also to examine if arms would be stopped inappropriately for successful strategies.

The second aim was to find an initial plan for the number and timing of interim analyses based on primary predictions of cure rates to ensure that it was possible to have analyses that work with both the monitoring guidelines and the time constraints of the trial. It also allowed these

to be loosely scheduled in advance, providing both the trial team and oversight committees a rough schedule of interim analyses.

The third aim was to assess the impact of the design on the overall power of the trial using simulations. The trial is complex and the sample size of the trial was estimated based on the monitoring phase of the trial, so this analysis was necessary to ensure there was sufficient power for the final analysis.

I have also developed priors to be used both in the monitoring phase and the final analysis of the trial, suggesting a way in which other researchers may also do the same.

While this work focuses on one HCV trial and this trial design is particularly suited to trials testing shortening strategies for HCV, as described in the background, the design and statistical concepts could be applied equally to various other disease areas. Therefore my work in this chapter both fits within my overall thesis aim of assessing trial designs that can be used to investigate HCV treatment shortening strategies, but can also be generalised to a wider research audience.

Work from this chapter has been presented at the October 2019 ICTMC conference in Brighton (112) and published in Trials in May 2020 (113).

5.2 Background

5.2.1 Trial designs allowing for the testing of multiple strategies

As there is very little data to inform optimal ways to shorten HCV treatment with DAAs, particularly in genotype 6, and there are many possible strategies that could be tested, trial designs that allow for the testing of multiple different options whilst allowing for the early stopping of unsuccessful treatments in order to focus on more successful treatments are essential, both for trial efficiency and to protect participants.

Two trial designs that incorporate both of these aspects are factorial trials and multi-arm multi-stage (MAMS) trials. Factorial trials allow for more than one strategy to be tested concurrently by randomising the participant to each intervention so that they may receive all, none or some combination of interventions (114). Factorial trials do not inherently have a mechanism for testing for unsuccessful treatments during the trial's progress, but accumulating trial data will be reviewed by a Data Monitoring Committee (DMC) and various stopping guidelines can be easily applied to the trial design. MAMS trials compare multiple treatment strategies simultaneously against a single control group (115). Interim analyses are pre-specified to occur when a specific number of outcomes are observed or after a pre-determined amount of follow-up time has passed, and each of the intervention groups are

tested against the control or against a historical control using an intermediate outcome measure. If an intervention group fails to meet the requirement to move onto the next stage, it is discontinued and future participants will be randomised to any remaining group carried forward into the next stage.

Both designs allow for greater efficiency in trials by reducing the number of participants required and shortening the time needed to test multiple interventions compared to sequential trials of individual interventions (116, 117). MAMS trials treating each combination of interventions as an independent arm are the preferred design when there is likely to be an interaction between the interventions, as factorial trials are not powered to detect interactions when each intervention is a single randomisation (118). However, the timing and maximum number of interim analyses within MAMS trials generally have to be pre-specified (119) or otherwise controlled with an alpha-spending function, which can be computationally difficult, particularly as the complexity of the design increases (120). Therefore, MAMS designs may be less suitable for interventions where the effect on outcomes is unknown and the interim analysis schedule may have to be altered, at least where *a priori* interactions are not expected.

5.2.2 Bayesian statistics

In frequentist statistics, the parameter of interest is a fixed and unknown constant and is estimated by finding the value of the parameter that maximises the likelihood function of the data. Uncertainty around the estimate is given by confidence intervals; estimated using the parameter and its standard deviation. The parameter will lie within the interval given a specified frequency. For example, with 95% confidence intervals the parameter will lie in the interval 95% of the time. In contrast, in Bayesian statistics the parameter is a random variable which is estimated usually by the mean of the posterior probability distribution, distributed over all possible parameter estimates, though other measures may be used. The posterior probability distribution is also used to calculate credible intervals, which are intervals that the parameter estimate lies within, given a specified probability. For example, 95% credible intervals provide the interval in which the estimate has a 95% chance of being in.

The basis of deriving the posterior probability comes from Bayes' Theorem and involves the prior distribution (the prior belief in the parameter of interest) and the likelihood. Bayes' Theorem states

$$P(H|D) = \frac{P(D|H) * P(H)}{P(D)}$$

where H is the hypothesis and D is the data. P(H) is then the probability of the hypothesis without any data, which is the prior distribution; P(D|H) is the probability of the data given the hypothesis, the likelihood; and P(H|D) is the probability of the hypothesis given the data, which is the posterior probability. P(D), the marginal distribution of the data, is a normalising constant and may not be of interest as it is constant for all hypotheses. Disregarding P(D), the theorem can then be restated as

$$P(H|D) \propto P(D|H) * P(H).$$

Posterior distributions can be complex and difficult to obtain; however if the prior and likelihood distributions are conjugate then this process is simpler and the posterior distributions will be of the same family as the prior distribution. For example, when the data generated is from a binomial distribution the conjugate prior is a beta distribution and the resulting posterior distribution will also be a beta distribution. When data are being collected, for example in an ongoing trial, the posterior probability can then become the prior distribution for the next analysis, and so the prior distribution can always be updated (121).

Priors may either be uninformative, which provide little information to the posterior distribution and are helpful when there is no previous research or other knowledge into the outcome of interest, or they may be informative based on previous similar studies or from experience of clinicians who work in the field of interest. In practice, sensitivity analyses using different priors from the one used initially in the primary analysis are recommended to determine sensitivity to the prior distribution chosen (122).

5.2.3 Bayesian monitoring

Data monitoring and stopping guidelines are most commonly framed within a frequentist framework with guidelines based on p-values or conditional power. When planning frequentist interim analyses it is generally accepted that care must be taken to control type I error, which can limit the ability to change the monitoring schedule to adapt to accumulating data, and may lead to delays in stopping unsuccessful treatments (for example, if strict guidelines such as Haybittle-Peto ($p<0.001$) are used).

Although strict-error control may not always be necessary in a trial using frequentist methods (123), a more flexible approach to monitoring can be easier to implement and justify by using a Bayesian approach, which allows for stopping guidelines that are based on directly interpretable probabilities, particularly in complex multi-arm trials (121, 124-126). However, some suggest that the frequentist type I error rate should be controlled even in Bayesian trials as it increases power (127). Common stopping guidelines used with Bayesian monitoring are the posterior probability, which is the probability that an outcome exceeds a specified value at

the point where the analysis is done, and the predictive probability, the probability of seeing a success at the end of the trial given the data accrued to the point where the analysis is done.

Bayesian monitoring has been applied retrospectively to trials conducted using a frequentist approach (128-130) and some trials have reported using Bayesian monitoring to assist with interim decisions, primarily in phase 2 studies (131-134) but also including two multi-arm trials (135, 136). However, from a review of the literature, no trials at the time of writing this chapter (Spring 2020) had reported applying Bayesian monitoring to a complicated factorial trial such as VIETNARMS. The search of the literature was not limited to a specific disease area and was both a focused search aiming to identify previous trials with a similar design and also a more general review of Bayesian monitoring in trials of any design.

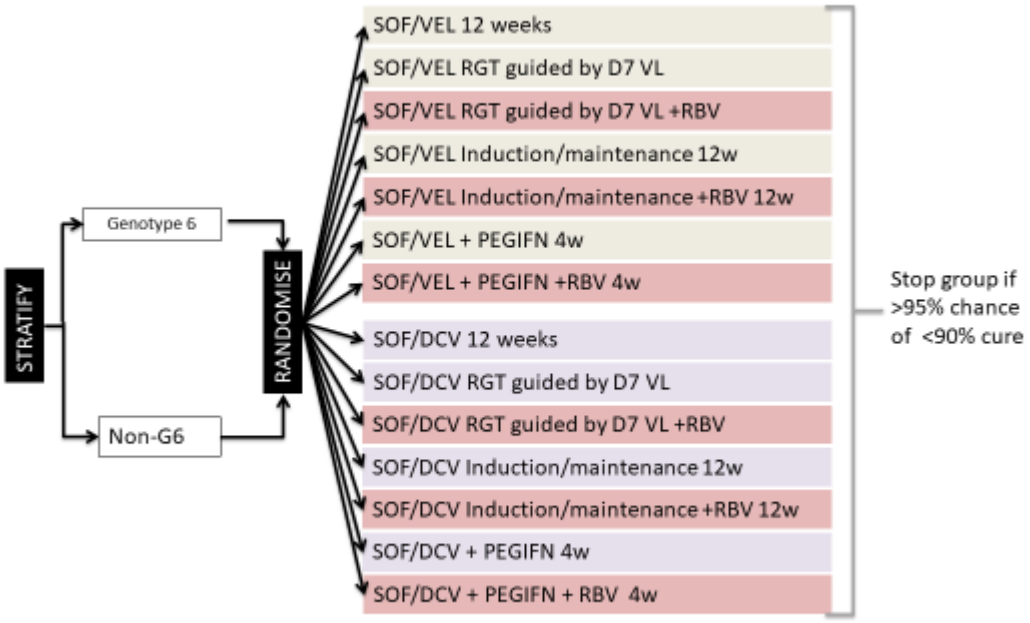
5.3 Methods

5.3.1 Trial design

VIETNARMS is a parallel-group open-label factorial trial (Figure 5.1). 1092 participants will be factorially randomised: 1:1 to two different WHO recommended dual DAA regimens (sofosbuvir/velpatasvir vs sofosbuvir/daclatasvir); 1:2:2:2 to the standard licenced 12-week treatment vs 4 weeks treatment with PEG-IFN+DAA vs 4-12 weeks response guided therapy (RGT) vs 12 weeks treatment using an induction/maintenance approach; and, if not randomised to standard 12-week treatment, 1:1 to adjunctive ribavirin vs no ribavirin for the duration of their DAA treatment. Randomisation will be stratified by genotype 6 vs all other genotypes.

Participants randomised to the PEG-IFN strategy will receive DAAs for 4 weeks with weekly PEG-IFN for 4 weeks starting at day 7. Treatment length for those randomised to RGT will be determined by HCV VL at day 7 and based on predicted viral kinetics (67): those with VL <lower limit of quantification (LLOQ) at day 7 will receive 4 weeks of treatment, those with VL LLOQ-250 IU/ml will receive 8 weeks and all others will receive 12 weeks. Participants randomised to induction/maintenance will receive 12 weeks of treatment: 2 weeks of daily treatment (induction phase) followed by 10 weeks of 5 days treatment per week taking weekends off from the first weekend following their full 2 weeks treatment. Within the trial, any participant not achieving cure with their first-line treatment will receive 12 weeks retreatment with the alternate drug regimen to the one they were originally randomised plus ribavirin.

Figure 5.1: VIETNARMS trial schema



Note: SOF/VEL=sofosbuvir/velpatasvir, SOF/DCV=sofosbuvir/daclatasvir, RBV=ribavirin, VL=viral load, RGT=response guided therapy, PEGIFN=pegylated interferon.

5.3.2 Trial analyses

The primary outcome of the trial is SVR12 on first-line treatment only, which is binary: all observed outcomes will either be SVR12 or treatment failure. The primary analysis will be Bayesian with 90% credible intervals; secondary analyses will use frequentist methods and 95% intervals. The final analysis will estimate risk differences between groups using marginal effects after logistic regression. The model will include all main randomised effects and strata, and will test interactions between all randomisations. Interactions will only be included in the final model if the 95% credible interval for the interaction term excludes no effect ($p<0.05$ for frequentist analyses).

The regimen comparison will be non-inferiority with a pre-specified margin of 5%. Both regimens are recommended by the WHO (19) and have similar cure rates and safety profiles in other studies, predominantly outside of genotype 6 (137, 138). However, there have not been any direct randomised comparisons to date. Showing non-inferiority between the two regimens will allow increased competition between DAAs and should lead to lower DAA prices. The strategy comparison will also be non-inferiority with a pre-specified margin of 10%. As the control strategy for this comparison is the standard 12 week duration, it is highly unlikely that any of the strategies will perform better than control, so a superiority analysis is not suitable. The margin of 10% is higher than for the regimen comparison because each strategy has its

own advantages and disadvantages, and the potential benefit in terms of numbers treated for the same fixed budget is much greater. Both non-inferiority margins were based on clinical judgement and the size of margins used in other trials of anti-infectives with relatively low failure rates (139).

The ribavirin comparison is a superiority comparison, powered to detect a 5% absolute difference. Ribavirin has been shown to increase SVR rates when taken with PEG-IFN (11), but its effect on SVR rates with DAA was unclear when the trial was designed (before STOP-HCV-1 reported). As there is no other expected benefit to taking ribavirin other than to increase SVR rates and as it can cause adverse effects such as anaemia (16), there is only interest in ribavirin if superiority can be shown.

5.3.3 Monitoring and stopping guideline

For the strategies to be viable outside the trial, first-line cure rates need to be high (>90%). The design of the trial therefore allows for failing groups to be stopped early at any time and subsequent participants to be randomised to more successful groups. Individual performance of groups receiving shortening strategies will be monitored during recruitment by an independent DMC who will make decisions on whether a group should be stopped. Groups receiving standard 12-week treatment will not be monitored as this is the licenced duration with cure rates >90% (137, 138). Interim analyses will not be comparative as the aim of monitoring is not to find the best strategy, but to find any strategy that meets a minimum acceptable cure rate that may also be non-inferior to standard treatment as different strategies may benefit different patient populations.

Analyses of cure rates will follow the Bayesian paradigm to allow the probability of the true cure rate being below different thresholds to be calculated: recruitment into a group will stop if there is a >0.95 posterior probability of the true cure rate being <90% (Pr(true cure rate <0.9|x)>0.95 where x is the data currently observed). The primary monitoring is combined across genotypes; if the combined group reaches the stopping guideline, each genotype stratum will be tested separately and the DMC will have the discretion to stop only those strata reaching the stopping criteria. Differences in stopping groups across strata are only likely to occur when there are extreme differences in the cure rates between the strata, which is not expected, and so the operating characteristics of the trial are based on stopping combined strata only. If neither stratum reaches the stopping criteria despite the combined strata doing so, it will be at the discretion of the DMC whether to stop recruitment into the stratum or group.

Analysis of SVR12 during the monitoring period will effectively be that of a single arm trial and not within the randomised and factorial structure of the trial. There are several benefits of this. As the treatment length varies between strategies, each strategy will have a different number of outcomes to the other strategies at any given time point. If interim analyses could only take place when the groups with the longest treatment lengths were suitable for analysis as would be needed for a comparative analysis, then this could delay the detection of failing groups with shorter treatment lengths.

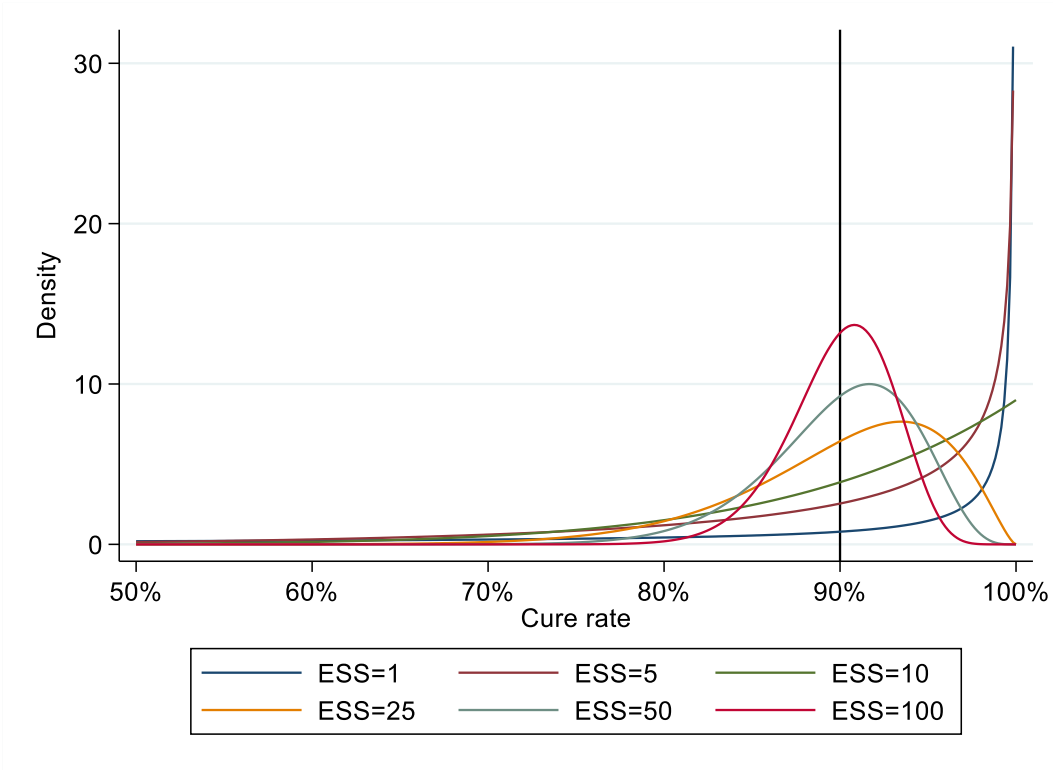
Analysing the groups separately will also allow for the detection of any potential interactions more easily during the trial and so give more confidence that there will be no interactions in the final analysis, which as this is a factorial trial is not desirable. Given that we are only interested in strategies with >90% true cure rate, it is likely that any interaction will lead to at least one group to have a cure rate below this threshold. Recruitment into the lower cure rate group(s) will then stop, removing the interaction from the trial.

5.3.4 Development of priors

As there are two types of analysis for the trial, both an interim analysis comparing single groups to a cure threshold of 90% and the final analysis which will compare groups directly, two sets of priors are needed.

For interim analyses, one initial prior is chosen and this will then become updated as outcomes are reported and information is added to the prior. As there is uncertainty about the performance of the shortening strategies, it was assumed that one strategy would fail completely such that all 4 groups receiving that strategy, of a total of 12 tested, would meet the stopping guideline. As each individual outcome is assumed to be Bernoulli-distributed, and therefore distribution of all outcomes is binomial, a beta prior was chosen as this was the conjugate prior for the binomial distribution. The mean of the prior was fixed at 0.9 and the effective sample size (a+b where beta(a, b) is the prior distribution) of the prior was varied until a distribution was found such that there was a ~0.33 probability of a cure rate <90%; ~0.33 probability was chosen as 4/12 randomised groups are expected to fail and therefore have a true cure rate <90%. The prior chosen was beta(4.5, 0.5) with mean 0.9, variance 0.015, an effective sample size of 5 and a 0.34 probability of a cure rate <90% (Figure 5.2). The relatively low precision of the prior will allow greater influence of the data in the posterior distribution.

Figure 5.2: Distribution of prior beta(a, b) with varying effective sample sizes



ESS: effective sample size.

For the end of trial analysis, the control cure rate analysis prior is beta(4.75, 0.25), which has a mean of 0.95 and the same effective sample size as the monitoring prior. The mean as derived from previous research into the trial drug regimens (137, 138). Sensitivity analyses will use a range of informative priors reflecting plausible belief in the clinical community (Table 5.1). Sceptical analysis prior distributions were chosen with means corresponding to the null hypothesis for each randomisation and enthusiastic analysis priors with means y greater than this, where y is the non-inferiority margin or absolute difference specified in the power calculations. The variances were arbitrarily set such that 90% of the prior distribution is within +/- y around the mean to reflect the strength of the belief in the mean effect. Thus for example, the risk difference for the drug regimen comparison has the sceptical analysis prior centred on -5% (the null hypothesis for the non-inferiority comparison) with 90% limits ±5%, giving a 0.05 probability that the cure rate will be 10% worse and a 0.05 probability the cure rate will increase (i.e. be >0%). The enthusiastic analysis prior is centred on 0% with 90% limits ±5%, giving a 0.05 probability that the cure rate will be 5% worse and a 0.05 probability that it will be 5% better with one regimen than the other.

Table 5.1: Priors to be used in the final analysis

	Primary analysis: uninformative	Sensitivity analyses	
		Sceptical	Enthusiastic
Control cure rate	Beta(4.75, 0.25)	Beta(4.75, 0.25)	Beta(4.75, 0.25)
Regimen comparison	N(0, 10000)	N(-0.05, 0.009)	N(0, 0.009)
Strategy comparison	N(0, 10000)	N(-0.1, 0.0036)	N(0, 0.0036)
Ribavirin comparison	N(0, 10000)	N(0, 0.009)	N(0.05, 0.009)

The beta prior for the control cure rate has mean 0.95 and variance 0.008.
The model will also be adjusted for genotype, which will have the prior N(0, 10000) in all analyses.
N(m,v) is the normal prior with mean m and variance v.

5.3.5 Statistical analysis

To define the performance characteristics of the proposed stopping guideline, posterior probabilities of cure rates and the probability of stopping groups at each number of outcomes were calculated analytically using beta and binomial distributions respectively. Timings of interim analyses were determined by applying the probabilities of stopping groups to a projected recruitment schedule. The average probability of stopping a genuinely inferior group was estimated by integrating the probability of stopping a group with respect to the monitoring prior beta(4.5, 0.5) over cure rates between 60% and 90%. The lower bound was determined from previous studies testing strategies most similar to those in VIETNARMS which have reported cure rates of >90% with lower confidence interval bounds >60% (41, 42). In studies with cure rates <60%, all participants received shortened therapies, regardless of their HCV VL, and did not receive adjunctive drugs (46, 47), therefore they were not considered relevant to this analysis – although power would be even greater to stop such a group. Cure rates above 90% were not considered as groups with these cure rates should not be stopped and so do not affect timing of the analyses.

Simulations of 5000 datasets with outcomes taken from binomial distributions were used to determine the overall probability of stopping a group, the cumulative probability of stopping groups at specified analysis time points, and to estimate power using marginal effects after fitting a logistic regression model to obtain risk differences. The model used to analyse data contained all randomised comparisons with main effects and adjustment for strata. 90% confidence intervals were used, reflecting the sample size calculation and primary analysis of the main trial results.

Predictive probabilities were calculated analytically using the beta-binomial distribution in R 3.5.1. All other analyses were performed using Stata v15.1.

5.4 Results/discussion

5.4.1 Operating characteristics of the Bayesian stopping guideline

The minimum number of failures required to satisfy the stopping criteria for the main monitoring beta(4.5, 0.5) prior and for each number of analysed participants are summarised in Table 5.2 (full details in Table 5.8). The probability of stopping a group is then the probability of observing the required number of failures in the group. When the true cure rate is 90%, the probability of incorrectly stopping a group is always <0.05, and this decreases as the true cure rate increases. It is expected from the specification of the stopping guideline that when the true cure rate is equal to the mean of the prior, then the probability of stopping a group is 0.05 so the probability of incorrectly stopping a group is always maintained below the correct level. The calculated probability is not exactly 0.05 and differs depending on the number of participants analysed due to the discrete nature of the outcome. This value is for each individual group at each analysis; the impact of multiply testing arms is covered in Section 5.4.4.

For small numbers of analysed participants, a larger proportion of failures are required to stop a group, increasing from a minimum of 17% to 100% of those analysed; therefore the probability of stopping a group incorrectly early in recruitment is also smaller as there is a smaller chance of observing greater proportions of failures regardless of the true cure rate. This protects against groups being stopped erroneously due to a high concentration of failures amongst the initial participants reaching the time point at which the primary outcome is measured (12 weeks after the end of treatment, EOT+12); if the two strata within a group share the same true cure rate then it is unlikely only one will reach the stopping threshold.

Groups with the lowest cure rates considered plausible (60%) are highly likely to reach the stopping criteria quickly (>90% chance of stopping after analysing 21-26 participants). Groups with moderately low cure rates (<80%) are also likely to be stopped before recruitment ends. However, groups with true cure rates slightly under 90% are unlikely to be stopped; the example of a true cure rate of 89% is given in Table 5.2. A low chance of stopping a group just below the target cure rate might be considered unacceptable in futility stopping guidelines in other situations, but the target of 90% is largely arbitrary and there may be interest in strategies that have a slightly lower cure rate if they are able to expand treatment access to difficult to reach populations. Increasing the probability of stopping groups with cure rates just below 90% would lead to a greater chance of incorrectly stopping groups with cure rates >90%

Table 5.2: Number of failures needed to stop a drug strategy group for each number of analysed participants

Analysed participants	Minimum number of observed failures needed to stop group*	Maximum probability of stopping group if true cure rate equals:		Minimum probability of stopping group if true cure rate equals:				
		90%	95%	90%	89%	80%	70%	60%
3-7	3	0.026	0.004	0.001	0.001	0.008	0.027	0.064
8-13	4	0.034	0.003	0.005	0.007	0.056	0.194	0.406
14-20	5	0.043	0.003	0.009	0.014	0.130	0.416	0.721
21-26	6	0.040	0.002	0.014	0.022	0.231	0.637	0.904
27-33	7	0.042	0.001	0.015	0.024	0.287	0.744	0.958
34-39**	8	0.037	0.001	0.017	0.028	0.367	0.844	0.986
40-41**	8	0.048	0.001	0.042	0.067	0.563	0.945	0.998
42-48	9	0.046	0.001	0.021	0.036	0.469	0.920	0.997
49-55	10	0.044	0.0004	0.022	0.038	0.528	0.952	0.999
56-63	11	0.047	0.0003	0.021	0.040	0.580	0.971	1.000
64-71	12	0.048	0.0002	0.023	0.045	0.648	0.985	1.000
72-78	13	0.045	0.0001	0.025	0.049	0.705	0.993	1.000

* >0.95 posterior probability of the true cure rate being <90% (Pr(true cure rate <0.9|x)>0.95, where x is the data currently observed) with the beta(4.5, 0.5) prior which has mean=0.9 and variance=0.015

**These rows not pooled despite the same number of minimum failures to provide information about the fully recruited strata (n=39).

Note: Groups will recruit a minimum of 78 participants with 39 participants in each stratum. If recruitment is stopped into one group, the maximum number of participants in each other group and the other stratum for that group will be higher.

Maximum and minimum probabilities are for the range of analysed participants.

and it is considered more important to retain these than to stop groups with slightly lower cure rates. Additionally, any other stopping guideline would similarly be unable to discriminate between these cure rates without a very large sample size. The probability of incorrectly not stopping a group rapidly decreases as the true cure rate decreases.

The overall probability of stopping a group (the proportion of simulations where a group was stopped rather than the expected proportion), and therefore making a correct decision or incorrect decision to stop recruitment into a group (again analogous to the frequentist concepts of power and type I error), show similar results to those given above (Table 5.3). For cure rates >90%, the probability of incorrectly stopping a group is always maintained <0.05 and for cure rates <90% there is a high probability of correctly stopping a group, with almost all groups being stopped when the true cure rate is ≤70%. The probability of incorrectly not stopping groups with cure rates ~80% is 12%, and for cure rates of 90% a group is far more likely not to be stopped than correctly stopped. However, as described above, for our design this is not as much of a concern as it may be for other trial designs because the aim of the monitoring is to stop clearly inferior regimens rather than those close to the arbitrary 90% threshold.

Table 5.3: Overall probability of stopping recruitment into a group

True cure rate	Probability of stopping group	Probability of making correct decision	Probability of making incorrect decision
60%	1	1	0
70%	0.999	0.999	0.001
80%	0.877	0.877	0.123
82.5%	0.740	0.740	0.260
85%	0.530	0.530	0.470
87.5%	0.312	0.312	0.688
90%	0.138	0.138	0.862
92.5%	0.040	0.960	0.040
95%	0.008	0.992	0.008
97.5%	0.001	0.999	0.001

Note: For true cure rates ≤90%, the correct action is to stop recruitment into a group; for cure rates >90% the correct action is to maintain recruitment into a group.

5.4.2 Timing of interim analyses

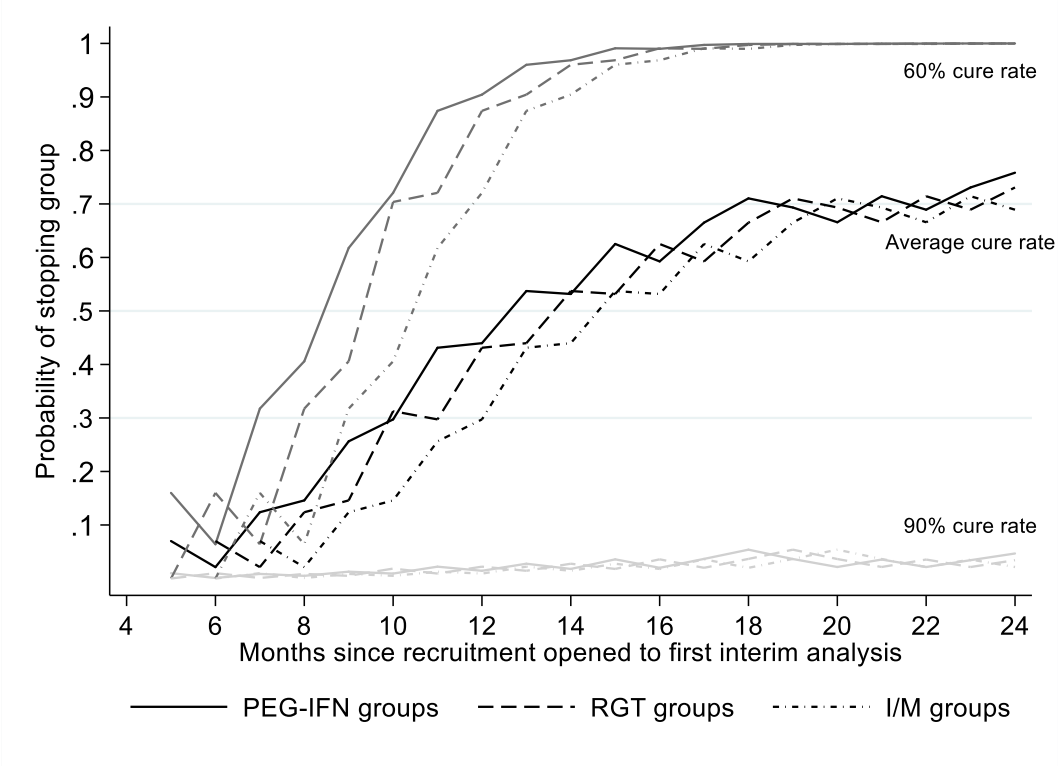
Interim analyses need sufficient numbers of participants at EOT+12 to give a reasonable probability of stopping a genuinely inferior group. It was therefore decided to perform analyses after the first month such that at least one inferior group has a 0.3, 0.5 or 0.7 average

probability of being stopped were this to be the first interim analysis, assuming cure rates are uniformly distributed on [0.6, 0.9] and given projected recruitment (Table 5.4). An average probability is used to reflect the uncertainty about the true cure rates; for low cure rates the probability of stopping a group can be substantially higher (Figure 5.3). An additional analysis before these thresholds will allow for any groups with a very low cure rate below the anticipated minimum of 60% to be detected early despite the very small probability of detecting a group with a cure rate between 60-90%.

Table 5.4: Predicted recruitment schedule

Month since enrolment opened	Number recruited in the month	Total participants recruited	Total left to recruit
1	8	8	1084
2	16	24	1068
3	24	48	1044
4	30	78	1014
5	35	113	979
6	40	153	939
7	52	205	887
8	52	257	835
9	52	309	783
10	52	361	731
11	52	413	679
12	52	465	627
13	52	517	575
14	52	569	523
15	52	621	471
16	52	673	419
17	52	725	367
18	52	777	315
19	52	829	263
20	52	881	211
21	52	933	159
22	52	985	107
23	52	1037	55
24	55	1092	0

Figure 5.3: Initial probability of stopping groups over an estimated recruitment schedule for various true cure rates



Note: The average true cure rate is uniformly distributed over 60-90%. Probabilities are calculated assuming no previous interim analysis. The total number of participants recruited at each month, of a total target of 1092, is: 113 at 5 months, 153 at 6, 205 at 7, 257 at 8, 309 at 9, 361 at 10, 413 at 11, 465 at 12, 517 at 13, 569 at 14, 621 at 15, 673 at 16, 725 at 17, 777 at 18, 829 at 19, 881 at 20, 933 at 21, 985 at 22, 1037 at 23 and 1092 at 24.

Four analysis time points were chosen to provide multiple opportunities to detect failing groups while allowing adequate time between analyses for the accrual of participants and outcome data, and preventing unnecessary burden on time and resources needed for analyses and subsequent DMC meetings. The highest probability threshold (0.7) is determined by the maximum average cure rate of the genuinely inferior groups and by the recruitment schedule, as analyses need to be performed sufficiently early enough to gain the benefit of randomising remaining participants to the other groups. The other thresholds (0.3, 0.5) were evenly spaced across the probabilities of stopping a group with an average cure rate, taking into consideration the first, early DMC meeting not based on these probabilities. Based on the underlying projected recruitment, the interim analyses are therefore expected to take place after 7, 10, 13 and 18 months of recruitment. The number of participants in each group and the probability of stopping an inferior group of each strategy type at these analyses are listed in Table 5.5.

Table 5.5: Timing of interim analyses

	First	Second	Third	Fourth
Months since recruitment started	7	10	13	18
Total recruited (out of maximum 1092)	205	361	517	777
Total reached EOT+12 weeks	44	144	286	533
At EOT+12 weeks in each:				
PEG-IFN group	5	14	24	42
RGT group	3	11	21	39
Induction/maintenance group	2	8	17	35
Average probability genuinely inferior group will be stopped*:				
PEG-IFN groups	0.124	0.297	0.537	0.710
RGT groups	0.021	0.313	0.440	0.665
Induction/maintenance groups	0.070	0.150	0.431	0.593

*Assuming true cure rate uniformly distributed over 60-90%
EOT+12: 12 weeks after the end of treatment; PEG-IFN: pegylated-interferon; RGT: response guided therapy

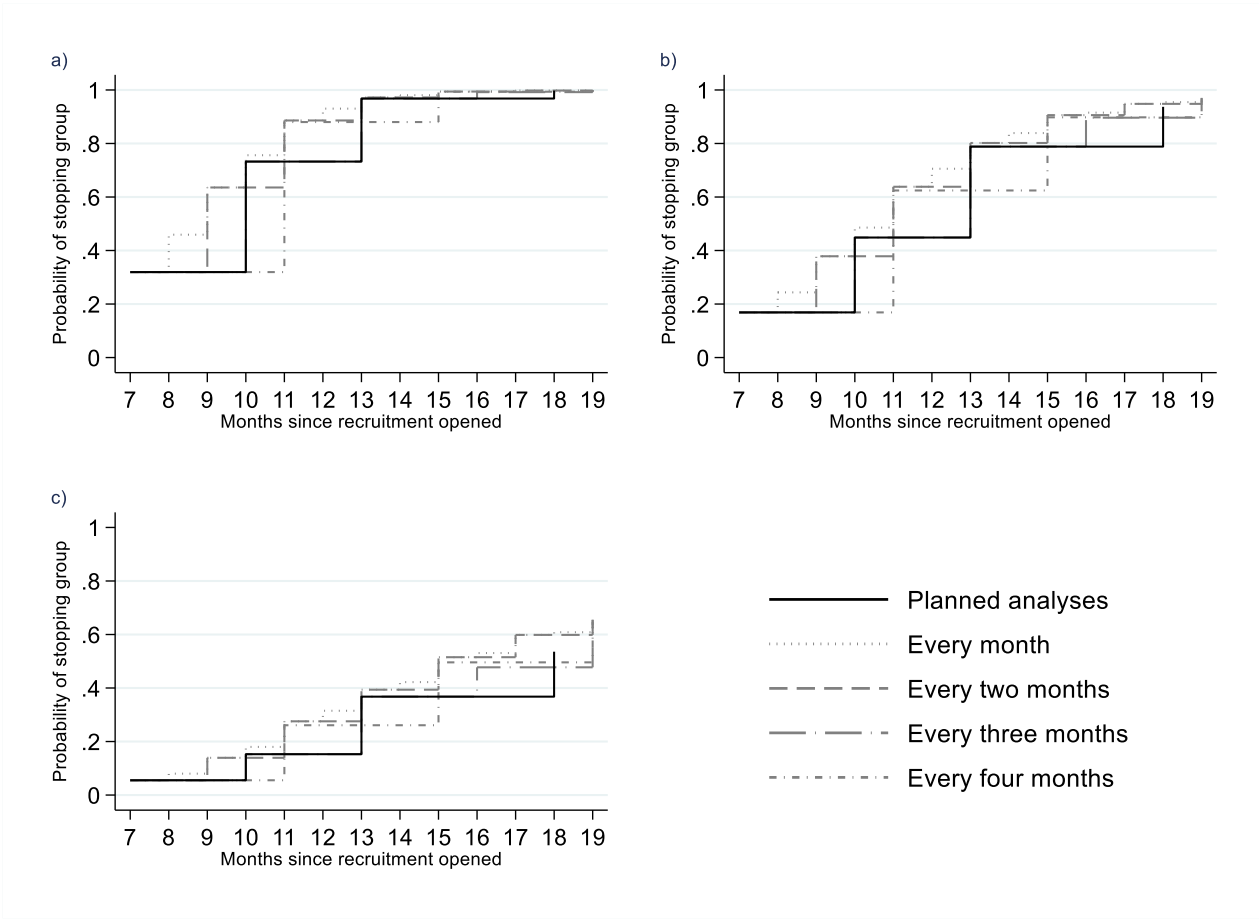
By assessing the cumulative probability of stopping a group (Figure 5.4), the schedule provides a balance between the number of analyses and the probability of stopping an arm. Performing interim analyses every two months provides more and earlier opportunities to detect a poorly performing group, however the timing of these analyses would mean that there would be a lower probability of finding such a group at each time point compared to the proposed schedule. Conversely, planning interim analyses for every four months provides a higher probability of detecting a failing group, but as analyses are further apart it may lead to too many participants being randomised into a group when it could have been stopped earlier.

The first three planned analyses follow a three month schedule, with the final analysis planned for 5 months after the third analysis. If the planned analyses took place every three months, then there would be five analyses in total: one 3 months after the third analysis at 16 months and a final analysis at 19 months after enrolment starts. However, as most participants will have been randomised into the trial by the later part of recruitment, there is limited benefit to performing more multiple analyses and stopping an arm so performing one analysis between 16 and 19 months is more efficient.

The only schedule to provide both more frequent opportunities without compromising on the probability of finding a failing group is to have an interim analysis every month. However, this

would be impractical due to the resources and time required of both the trial team and the DMC for each analysis.

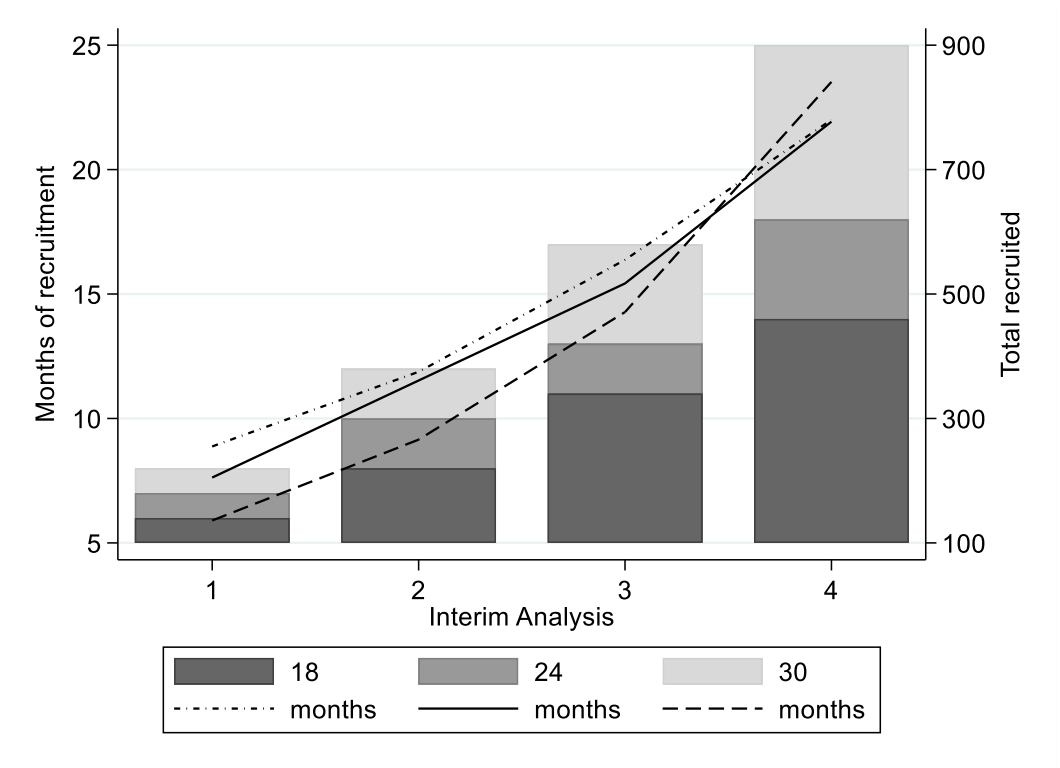
Figure 5.4: Cumulative probability of stopping interferon groups for different interim analysis schedules assuming true cure rates of a) 60% b) 70% c) 80%.



As this is only projected recruitment, sensitivity analyses were performed to examine the effect of faster or slower recruitment (Figure 5.5). Analyses closer to the start of recruitment and so more dependent on assumptions of cure rates would be at similar time points, regardless of recruitment speed. There were greater differences in the optimal timings of the last analyses, but the timing of this analysis is the most flexible and can be determined based on observed rather than assumed true cure rates. At each of the analyses, the difference in the total number recruited at each time point is less than the number recruited in one month, indicating that changes to the recruitment schedule alter only the timing of the analyses and not to the number of outcomes analysed. Therefore, if there are significant delays in recruitment, interim analyses will be timed such that they include a similar number of participants at EOT+12 to that in the expected schedule. Sensitivity analyses also explored changing the lower bound of the distribution over which cure rates of genuinely inferior groups are assumed to be distributed from 60% (Table 5.6), but this had minimal effect on the

timing of initial interim analyses. Again, the largest difference in the timing of the analyses was for the last interim analysis, which can be based on observed cure rates instead.

Figure 5.5: Sensitivity analysis of the optimal timing of interim analyses comparing recruiting over 24 months to recruiting over 18 months or 30 months



Note: assumes true cure rate in inferior arms is uniformly distributed over 60-90%. The bars show how many months after recruitment started that the numbered interim analysis would take place, the lines show how many participants would be recruited if the interim analysis took place at that time point.

Table 5.6: Sensitivity analysis of the optimal timing of interim analyses by altering the lower limit of the uniform distribution over which cure rates of genuinely inferior arms are assumed to be distributed

	Lower bound of uniform distribution of cure rates of inferior arms								
	0	0.1	0.2	0.3	0.4	0.5	**0.6**	0.7	0.8
Probability of ≥1 arm being dropped									
30%	5	5	7	7	8	9	**10**	13	18
40%	7	7	7	8	9	10	**11**	15	-
50%	7	7	8	9	10	11	**13**	17	-
60%	8	9	9	10	11	13	**15**	24	-
70%	9	10	10	11	13	15	**18**	-	-
80%	11	13	15	16	17	23	**-**	-	-
90%	18	23	22	-	-	-	**-**	-	-

Note: the lower bound used for the timing of analyses is in bold for clarity. The table shows the month after recruitment started that interim analyses would take place, varying the assumption of the lower bound of the expected average cure rate (columns) and the probability of ≥1 arm being dropped (rows). The lower bound used for determining the timing of analyses was 0.6 and the probabilities of dropping an arm were 30%, 50% and 70%.

5.4.3 Power for final analysis

There is a lot of uncertainty about the power for each comparison in the final analysis, due to the overall sample size being determined by the monitoring analyses and the lack of knowledge of the standard 12-week cure rates for this population, in which 50% of participants are expected to have genotype 6. It is hypothesised that the overall cure rate is 95%, but the cure rates in specific randomised groups may also vary depending on how the randomisations relate to each other. In particular, how the cure rates in the shortened treatment with ribavirin groups compare with shortened treatment without ribavirin groups and with standard treatment groups will have an impact on power as these are the comparisons that can vary the most against each other.

Based on previous research (137, 138), it is reasonable to assume that the cure rates between the different drug regimens will either be equal or very similar such that it will have only a very small effect on the power. If it is also assumed that the regimen null hypothesis is rejected and there is a 5% absolute difference between participants taking ribavirin and those not then, due the constraints of the regimen and ribavirin hypotheses, the cure rates in each group are completely determined by the difference in cure rates between the shortening strategies.

There are three potential comparisons that are non-inferiority with respect to the standard duration group: the pooled shortening strategy groups, the shortening strategy with ribavirin groups, or the shortening strategy without ribavirin groups. If the pooled shortening strategy groups are non-inferior to the standard duration group and there is a 5% ribavirin superiority, then the groups taking ribavirin must have cure rates 2.5% higher than the standard duration group and the groups not taking ribavirin must have cure rates 2.5% lower. If the shortening strategy without ribavirin groups are non-inferior, then these groups must have a cure rate equal to the standard groups and the shortening strategy groups taking ribavirin must have a 5% higher cure rate than the standard duration groups. Finally, if the shortening strategy with ribavirin groups are non-inferior, cure rates in these groups will be equal to those in the standard duration groups and those not taking ribavirin will have cure rates 5% below those in the standard duration groups. These alternatives, in relation to the standard duration and ribavirin groups cure rates, are shown in different columns of Table 5.7.

Power to determine non-inferiority for the regimen comparison using a 5% margin is mostly unaffected by assumptions about different values for the standard 12-week cure rate with power varying between 97-99%. Power to determine non-inferiority in the strategy comparison using a 10% margin is 100% when the ribavirin group cure rate is higher than the standard duration cure rate, but drops to 96% when the cure rates in the two groups are equal. Power is lower for the ribavirin comparison, and so for if testing for superiority in the

regimen comparison, than for the non-inferiority comparisons but remains high at >90% regardless of the standard duration cure rate and ribavirin effect.

Table 5.7: Group specific cure rates and power for 1092 participants with an overall 95% cure rate

	Ribavirin group cure rate compared to 12 week standard treatment cure rate		
	5% higher	2.5% higher	Equal
Group-specific cure rates:			
Standard 12 week treatment groups	93.3%	95%	96.7%
Shortened treatment with ribavirin groups	98.3%	97.5%	96.7%
Shortened treatment without ribavirin groups	93.3%	92.5%	91.7%
Power for a:			
5% non-inferiority margin for regimen comparison	99%	98%	97%
10% non-inferiority margin for strategy comparison	100%	100%	96%
5% absolute difference for ribavirin comparison	98%	95%	91%

Note: Assumes no groups have been stopped during recruitment.

5.4.4 Type I error

As discussed in the background, there is some dispute about whether all trials, particularly trials under the Bayesian paradigm, need to have strict control over the type I error rate. In VIETNARMS, there are many hypotheses that are tested throughout the trial, in both the monitoring and final phases of the trial, but there is no multiple-test correction.

During the monitoring phase, interim analyses will examine if each single group has a true cure rate >90%. These tests will be repeated at every interim analysis and as the trial is factorial, each group is not independent from all other groups: for example, unless there is an extreme interaction between the strategy and ribavirin randomisations, all groups taking ribavirin cannot be considered to be independent. Most trials will account for this using a less restrictive stopping guideline or using a multiple-testing correction, such as the Haybittle-Peto rule or Bonferroni correction. However, VIETNARMS does not. This is partly because not making adjustments while using Bayesian statistics is more accepted and easier to implement, and therefore also easier to justify, than frequentist methods. Additionally, multiple testing will only decrease the probability that the true cure rate of a group is <90% and therefore may

allow groups to progress to the final analysis unnecessarily. However, strict control of groups included in the final analysis is not needed as the 90% cure rate is arbitrary and any errors made during the monitoring phase can therefore be resolved at the final analysis.

For the final analysis, the hypotheses are comparative making these distinct and independent from the previous interim analyses. As all the randomisations and strategies are distinct and will be reported together, there is no need to control the type I error rate for this analysis (117).

5.4.5 Limitations of the design

A potential weakness in the design is that the sample size was not originally calculated using Bayesian principles, but primary analyses will be conducted using Bayesian methods to allow for the calculation of posterior probabilities exploring the difference in cure rates between the interventions. However, for the non-inferiority comparisons, sample size estimates obtained using Bayesian methods are similar to or smaller than those obtained using frequentist methods (140), suggesting that our design is likely to be conservative. Additionally, secondary analyses will use frequentist methods for comparison. The overall sample size has an effect only on the number of interim analyses and has no effect on the timing of them. The timing of the analyses is based on observing a given number of outcomes given an average cure rate distributed over 60-90%. It is possible that a large increase in sample size may increase the rate of recruitment, for example by having more study centres join and help recruit for the trial, but if this rate does not change then the numbers of outcomes at each time point will not change. A change in sample size is also clearly independent of the true cure rate of the strategies, and so the timing of the interim analyses are not affected.

As the analyses will be Bayesian, priors have been chosen, but as there is a lack of previous data to guide the selection of these priors they may not be appropriate. For the monitoring prior, it is assumed that one strategy will fail completely. It is likely that if a strategy performs very poorly in one arm, all four arms of that strategy will perform poorly. For those strategies that exceed the stopping criteria slightly it may be possible that not all arms of that strategy fail. Additionally, more than one strategy may fail. However, as the prior is relatively uninformative the data accrued during the trial will have a strong influence on the prior lessening the effect of the initial prior choice as the trial continues. An updated analysis may need to be performed if the choice of prior appears to be grossly incorrect. For the final analysis, the control cure rate prior suffers from the same issues as the monitoring prior and may need to be altered. For the remaining priors in the final analysis, it is planned to use uninformative priors for the primary analysis, with sceptical and enthusiastic priors for sensitivity analyses, so these are designed not to have an impact on the analysis.

The timing of and the number of participants at interim analyses are determined by at least one group, usually the 4 week treatment group with PEG-IFN since this has the shortest overall treatment duration, reaching a certain probability threshold of being stopped. This may mean delays in identifying unsuccessful groups receiving other strategies. This is unlikely if groups have cure rates lower than the average cure rate because, as discussed above, these will be detected faster than anticipated, but delays may occur if the cure rate is above the average, but <90%. As the treatment length of participants in the RGT groups is unknown until after their day 7 visit, it is not possible to stagger treatment start dates so that the length between randomisation and EOT is the same for all strategies. Staggered treatment start might also lead to drop out after randomisation but before starting treatment, leading to inefficiency and potential bias. Cure rates will be monitored and if the timing rule is found to be inappropriate it can be adjusted to change the timing or frequency of analyses with no penalty to the probability of incorrectly stopping recruitment into a randomised group, due to the use of Bayesian monitoring (121).

The power calculations for the final analysis assume that all groups will be included and that no groups have been stopped. It is possible that power will be lower if fewer groups are included due to some groups being stopped if they meet the stopping criteria, but for most comparisons with a full sample, power is very high and is likely to remain acceptable at the final analysis even with the exclusion of some participants. To help preserve power, if groups are stopped early subsequent participants will be randomised to open groups. The power calculations were also estimated using frequentist methods, though the primary analysis will use Bayesian methods. However as power is extremely high, the analogous concept to power in Bayesian analysis, that for non-inferiority comparisons the lower credible interval bound is above the non-inferiority margin, is likely to be similarly high. Additionally, due to the many possible combinations of strata and groups that could be stopped with different true and observed failure rates and at different times, examining the impact of stopping multiple combinations would require a large number of assumptions, probably also using a factorial simulation design, and hence would be a large piece of additional work in its own right separate to the work in this chapter. This is also the case for examining the impact of stopping multiple randomised groups on other aspects of the trial, such as bias.

5.4.6 Alternative designs

Alternative Bayesian designs include basing the stopping guideline on a predictive probability, the probability of achieving a success at the end of the trial, which have been used elsewhere (141). In VIETNARMS, a success during the monitoring period is stopping a genuinely inferior group, which means that there is a >0.95 posterior probability of a <90% true cure rate in that

group. A rule based on predictive probabilities would then state a group will be stopped at an interim analysis if there is a >0.95 chance of stopping a group at the end of the trial (Pr([Pr(true cure rate<0.9|z)>0.95]|x)>0.95 where x is the data currently observed and z the complete data with all outcomes observed). For the monitoring beta(4.5, 0.5) prior and a fully recruited group, the stopping criteria is met with 13 failures so equivalently the group is stopped if there is a >0.95 chance of observing ≥13 failures in the fully recruited group.

Predictive probabilities place a large emphasis on the arbitrary target cure rate of 90%, and hence were not used for VIETNARMS. The final analysis will compare strategies against control and not test cure rates in individual groups. The aim of monitoring is to detect poorly performing groups and stop them early, rather than to ultimately achieve a particular cure rate threshold within a group at the end of the trial, as there may be other advantages to strategies that have a slightly lower cure rate than 90% in specific populations or circumstances. Compared to the posterior probability based stopping guideline, using predictive probabilities requires a similar number of failures or more to stop a group (Table 5.8), so they do not offer any benefits in detecting poorly performing groups more quickly for this design, though it may for others (141). Stopping rules and guidelines based on posterior probabilities can be converted to those based on predictive probabilities (142) so interim analyses can incorporate predictive probabilities to provide more information to the DMC if needed.

Another approach would be to analyse the outcome data after every reported outcome rather than at scheduled interim analyses, which could reduce the time until a genuinely inferior group is stopped. Implementing this would be complex due to the many groups and varying treatment lengths. The small benefit in the reduction in time would not justify the additional work required to monitor outcomes intensely.

In VIETNARMS, if a group performs badly then randomisation will completely stop into that group. Alternatively, the trial could have utilised response adaptive randomisation where all groups would be retained, but the allocation ratio would alter to favour randomisation into a group that is showing the most potential. This design was not considered suitable for VIETNARMS for several reasons. Firstly, each group will be tested separately and, as the randomisation is factorial, each group is the result of multiple randomisations and so it could be unclear how to adapt the randomisation allocations correctly. For example, the ribavirin groups may perform better with one shortening strategy but much worse with another strategy; in this case it is unclear how the ribavirin randomisation allocation should be changed. Secondly, if a group is performing particularly badly and there is no prospect of that strategy being adopted, then continuing with that group would be a burden on resources, including on the number of participants enrolled. Thirdly, in an open-label trial such as

VIETNARMS, response-adaptive randomisation risks unblinding investigators to the relative performance of open groups, information which is usually privy only to a Data Monitoring Committee.

Table 5.8: Probabilities of stopping a group in VIETNARMS using either a posterior probability rule or a predictive probability rule

	Stop if >95% posterior probability of true cure <90%			**Stop if >95% predictive probability of >95% posterior probability of true cure rate <90% in fully recruited arm (i.e. 13 failures)**		
Total at EOT+12	**Failures needed to stop (G)**	**P(observed failures ≥G | true cure=89%)**	**P(observed failures ≥G | true cure=90%)**	**Failures needed to stop (F)**	**P(observed failures ≥F | 12 failures in fully recruited arm)**	**P(observed failures ≥F | true cure=90%)**
3	3	0.001	0.001	3	0.003	0.001
4	3	0.005	0.004	3	0.012	0.004
5	3	0.011	0.009	4	0.002	0.0005
6	3	0.021	0.016	4	0.005	0.001
7	3	0.033	0.026	4	0.011	0.003
8	4	0.007	0.005	4	0.020	0.005
9	4	0.012	0.008	4	0.033	0.008
10	4	0.018	0.013	5	0.008	0.002
11	4	0.026	0.019	5	0.013	0.003
12	4	0.035	0.026	5	0.021	0.004
13	4	0.046	0.034	5	0.030	0.006
14	5	0.014	0.009	5	0.043	0.009
15	5	0.019	0.013	6	0.012	0.002
16	5	0.025	0.017	6	0.017	0.003
17	5	0.032	0.022	6	0.025	0.005
18	5	0.041	0.028	6	0.034	0.006
19	5	0.050	0.035	6	0.009	0.002
20	5	0.061	0.043	6	0.013	0.002
21	6	0.022	0.014	7	0.018	0.003
22	6	0.028	0.018	7	0.025	0.004
23	6	0.034	0.023	7	0.033	0.005
24	6	0.042	0.028	7	0.042	0.007

	Stop if >95% posterior probability of true cure <90%			Stop if >95% predictive probability of >95% posterior probability of true cure rate <90% in fully recruited arm (i.e. 13 failures)		
Total at EOT+12	**Failures needed to stop (G)**	**P(observed failures ≥G | true cure=89%)**	**P(observed failures ≥G | true cure=90%)**	**Failures needed to stop (F)**	**P(observed failures ≥F | 12 failures in fully recruited arm)**	**P(observed failures ≥F | true cure=90%)**
25	6	0.050	0.033	7	0.054	0.009
26	6	0.059	0.040	7	0.068	0.011
27	7	0.024	0.015	8	0.022	0.004
28	7	0.029	0.018	8	0.029	0.005
29	7	0.034	0.022	8	0.037	0.006
30	7	0.041	0.026	8	0.046	0.008
31	7	0.048	0.031	8	0.058	0.010
32	7	0.056	0.036	8	0.071	0.012
33	7	0.064	0.042	9	0.023	0.004
34	8	0.028	0.017	9	0.029	0.005
35	8	0.033	0.020	9	0.036	0.006
36	8	0.039	0.024	9	0.045	0.008
37	8	0.045	0.027	9	0.056	0.009
38	8	0.052	0.032	9	0.069	0.011
39	8	0.059	0.037	9	0.083	0.013
40	8	0.067	0.042	9	0.100	0.016
41	8	0.075	0.048	10	0.032	0.006
42	9	0.036	0.021	10	0.040	0.007
43	9	0.042	0.024	10	0.049	0.009
44	9	0.047	0.028	10	0.061	0.010
45	9	0.054	0.032	10	0.074	0.012
46	9	0.061	0.036	10	0.089	0.014
47	9	0.068	0.041	10	0.107	0.016
48	9	0.076	0.046	10	0.127	0.019
49	10	0.038	0.022	11	0.037	0.008
50	10	0.043	0.025	11	0.046	0.009
51	10	0.049	0.028	11	0.057	0.011
52	10	0.055	0.032	11	0.070	0.013

Total at EOT+12	Stop if >95% posterior probability of true cure <90%			Stop if >95% predictive probability of >95% posterior probability of true cure rate <90% in fully recruited arm (i.e. 13 failures)		
	Failures needed to stop (G)	P(observed failures ≥G | true cure=89%)	P(observed failures ≥G | true cure=90%)	Failures needed to stop (F)	P(observed failures ≥F | 12 failures in fully recruited arm)	P(observed failures ≥F | true cure=90%)
53	10	0.061	0.035	11	0.085	0.014
54	10	0.068	0.040	11	0.103	0.017
55	10	0.075	0.044	11	0.124	0.019
56	11	0.040	0.021	11	0.148	0.021
57	11	0.044	0.024	11	0.175	0.024
58	11	0.050	0.027	12	0.041	0.011
59	11	0.055	0.031	12	0.051	0.013
60	11	0.061	0.034	12	0.064	0.015
61	11	0.068	0.038	12	0.080	0.017
62	11	0.074	0.042	12	0.098	0.019
63	11	0.081	0.047	12	0.122	0.021
64	12	0.045	0.024	12	0.150	0.024
65	12	0.050	0.026	12	0.184	0.026
66	12	0.055	0.029	12	0.224	0.029
67	12	0.060	0.033	12	0.273	0.033
68	12	0.066	0.036	12	0.332	0.036
69	12	0.073	0.040	12	0.402	0.040
70	12	0.079	0.044	12	0.485	0.044
71	12	0.086	0.048	12	0.583	0.048
72	13	0.049	0.025	12	0.700	0.053
73	13	0.054	0.028	12	0.838	0.058
74	13	0.059	0.031	12	-	0.063

5.5 Key findings

- With the given trial design and stopping guideline, the probability of incorrectly stopping a group is maintained <0.05.
- The probability of correctly stopping a group is very high when the true cure rate is 60% and there is still a reasonable chance of stopping a group with higher true cure rates (such as 80%).
- For this design, four interim analyses were chosen in month 7, 10, 13, and 18 after starting recruitment.
- Power is high for all comparisons in the final analysis.
- Overall, the design is statistically adequate for the aims of the trial.

6 Discussion

6.1 Summary of key results

In this thesis I have investigated which patients and which treatment strategies would be most suitable for short-course HCV treatment by exploring associations between patient demographics and treatment options and several outcomes associated with treatment failure, including QoL.

Baseline VL was the key determinant in deciding which patient might be most suitable for receiving short-course treatment as it was associated with all three outcomes investigated: those with higher baseline VLs were more likely to have an early treatment failure and a higher peak rebound VL once they had a treatment failure. Other groups of patients that were more likely to fail, and have an earlier failure when they did so, were those with HCV genotype 1a (compared to 1b) and those who received shorter durations of treatment. Having resistance to prescribed DAAs and current substance abuse greatly reduced the probability of a patient curing on short-course HCV treatment. Other factors that should be considered when deciding whether a patient might cure with short-course treatment are IL28B genotype, bilirubin, Fibroscan score and age. Ribavirin was associated with higher rates of SVR12 in those receiving the shorter 4-7 week length treatment and, despite being known to cause more adverse effects than DAAs alone, did not have a strong or lasting negative effect on QoL. Treatment length did not have a discernible effect on QoL, either.

I have also developed the design of a complex trial that aimed to test different strategies and treatment options for short-course HCV treatment. I showed that the operating characteristics of the stopping guideline for the trial were appropriate, with the probability of incorrectly stopping a group adequately maintained below the expected 0.05 and a reasonably high probability of correctly stopping a group with unacceptably low cure rates, that increased as the true cure rate in the strategies decreased. I developed a monitoring plan consisting of four interim analyses, timed to maximise the chance of stopping a poorly performing strategy. I also found that power at the end of the trial would be very high.

6.1.1 Chapter 2 – Factors associated with SVR12

Chapter 2 of the thesis investigated factors associated with SVR12 with data from the STOP-HCV-1 trial and estimated probabilities of cure for specific groups of patients based on these factors. For the 4-7 week variable duration treatment, a lower baseline VL, HCV genotype 1b (rather than 1a), IL28B genotype CC (rather than CT or TT), higher baseline bilirubin, not having baseline resistance to prescribed DAAs or no current substance abuse were all independently associated with higher rates of SVR12. No on-treatment or end of treatment factors added

important prognostic value compared with baseline factors alone. The probabilities of cure estimated from the baseline model indicate that patients infected with HCV genotype 1a who do not have resistance to prescribed DAAs or current substance abuse, or patients infected with genotype 1b who do not have current substance abuse have the highest probability of cure on short-course treatment. Due to the low probability of cure, patients with both resistance to prescribed DAAs and current substance abuse should not receive short-course treatment.

Alternative model building techniques were used and factor selection and prediction probabilities between the different models selected were compared. For all three model building techniques, the factors selected were similar, and even identical for the backwards elimination and best subset regression models using multiply imputed data. The predicted probabilities of SVR12 were very similar between the backwards elimination and best subsets regression models; the predictions from the lasso model were less likely to indicate very high or very low chance of cure but correlation between predictions from all models was very high (Spearman rho>0.94).

Factors associated with SVR12 in participants who received 8-week fixed duration treatment were also investigated, but no strong associations were identified due to the very high cure rate in this group. Albumin and creatinine were associated with SVR12, but their effects were not large and the evidence for their effects was not particularly strong.

6.1.2 Chapter 3 – Factors associated with time to treatment failure and viral kinetics

Chapter 3 of the thesis extended the work of Chapter 2 by estimating the kinetics of viral load rebound in those who actually failed using linear mixed models, allowing for random effects between participants, and explored factors associated with the kinetics and associated with time to treatment failure in the STOP-HCV-1 and SEARCH-1 trials. A longer length of DAA treatment, HCV genotype 1b (compared to 1a), lower baseline VL and being younger were all independently associated with a longer time until treatment failure. The trial a participant was enrolled into was also a factor (SEARCH-1 participants had a longer time until treatment failure, despite a more intense visit schedule after completing treatment), which likely acts as a proxy factor for a combination of other factors, including HCV genotype 6.

Without adjusting for factors associated with viral kinetics, the peak rebound VL was 6.02 $\log_{10}$ IU/ml (95% CI 5.77, 6.28) with an increase in the slope from treatment failure to peak VL of 0.60 $\log_{10}$ IU/ml (95% CI 0.40, 0.80) per week. Factors associated with a higher peak VL were being enrolled into SEARCH-1, receiving ribavirin, HCV VL being <LLOQ at end of treatment, a higher baseline VL, not having baseline resistance to prescribed DAAs and a higher Fibroscan

score. Having a shorter time between the treatment failure visit and the preceding VL measurement and genotype IL28B CC were associated with a greater increase in VL between treatment failure and peak rebound VL (first slope in the models); there were no factors associated with the change in VL after peak VL (second slope in the models). After adjusting for all the factors above, most of the variance in the peak VL had been explained indicating that the key factors associated with peak VL had mostly likely been identified. The predicted peak rebound VLs were numerically similar or slightly lower than the observed baseline VLs. A key limitation of the kinetics analysis is the differences in visit schedules between the two trials and the assumed common timing of peak rebound VL after the first VL observed to meet the prespecified failure definition.

6.1.3 Chapter 4 – Quality of life of short-course treatment

Chapter 4 of the thesis examined QoL of participants in the STOP-HCV-1 trial at the point of completing treatment (EOT) and at the point of the visit confirming cure or starting retreatment (EOT+12/RT D0). For the two comparisons of treatment length (4-6 week VUS1 vs 4-7 weeks VUS2; 4-7 weeks variable duration vs 8 week fixed duration) there was no clear indicator that length of treatment had an impact on QoL at either time point.

Participants who took ribavirin reported more problems in performing usual activities at EOT and had a trend towards worse QoL in the SF-12 summary scores for both physical and mental health at EOT. However, at EOT+12/RT D0 these effects were no longer apparent, which is consistent with previous studies that have shown a return to baseline QoL or higher after an initial decline while taking ribavirin (90, 92, 100, 104, 105). As anaemia is a common adverse event associated with taking ribavirin, the impact of anaemia on QoL was also investigated. After adjusting for anaemia, the trend towards worse physical QoL on the SF-12 for those taking ribavirin was no longer present suggesting that the negative effects on physical QoL may be due to ribavirin-associated anaemia, and not ribavirin itself. However, the trend towards negative mental QoL was strengthened after adjusting for anaemia, suggesting this association may have a different pathway.

At EOT, those who later cured reported numerically similar QoL to those who later experienced treatment failure. At EOT+12/RT D0, there were some areas on the EQ-5D-5L where those who failed had numerically worse QoL scores. Previous studies have shown a more positive benefit to QoL after achieving SVR12 (91-96, 101-103); however, the comparison in these previous studies was to a participant's baseline score and not to participants who failed on treatment, which was the analysis performed here.

The main limitation of this analysis is that a large number of statistical tests were performed and no adjustment was made for multiple testing, so despite the small number of associations found, any significant result may be erroneous.

6.1.4 Chapter 5 – Statistical aspects of the VIETNARMS trial

Chapter 5 of the thesis assessed the statistical aspects of the VIETNARMS trial, which planned to investigate several HCV drug shortening or sparing strategies using a complex factorial design and a Bayesian early stopping mechanism for strategies that were unlikely to perform well. The stopping guideline stated that recruitment into a group would stop if there was a >0.95 posterior probability of the true cure rate being <90%. The probability of incorrectly stopping a group using the guideline was always maintained below 0.05 if the true cure rate was 90%, with the probability decreasing as the true cure rate increased. The probability of correctly stopping a group with a low cure rate was very high when the true cure rate was low (>90% chance of stopping after analysing 21 or more outcomes if the true cure rate was 60%) and was still acceptable in higher cure rates (>50% chance of stopping after analysing 40 or more outcomes if the true cure rate was 80%).

A monitoring plan was developed with four interim analyses chosen to be performed in month 7, 10, 13 and 18 after starting recruitment. The average probability (assuming a true cure rate uniformly distributed over 60-90%) of stopping a genuinely inferior group at the first analysis was very low, but would be higher if the true cure rate was lower than this. The average probability of stopping a group with cure rate genuinely <90% increased to >0.43 for all strategies at the third analysis. A limitation of the monitoring plan is that it was based on one strategy reaching a threshold of being stopped; however due to the Bayesian design, there was no penalty from performing more interim analyses if the independent DMC felt it necessary to monitor a different group's outcomes at a separate time point.

As the original sample size calculation was determined by the monitoring analyses, power at the end of the trial was also assessed. I showed that the power to conclude non-inferiority using a 5% non-inferiority margin for the regimen comparison and a 10% non-inferiority margin for the strategy comparison and to conclude superiority using a 5% absolute difference for the ribavirin comparison, was always high (>91%) regardless of the comparator strategy group (non-ribavirin group, ribavirin group, pooled group) and the assumption of the effect of ribavirin. These power calculations assumed that no groups would be stopped during recruitment: power would decrease if recruitment into a group was stopped. However, power would likely remain at a sufficiently high level given the very large power estimated, and that any remaining participants would be recruited into other continuing groups.

6.2 Comparison of factors selected between models

When investigating factors associated with treatment failure, I considered three different types of outcomes: failure vs cure at SVR12, rebound kinetics in those who failed and time to failure after EOT. Each model included different groups and numbers of participants and so there were varying amounts of power to detect associations with the outcomes. For the outcome of failure vs cure, 87 participants in the 4-7 week variable duration group in STOP-HCV-1 trial were included. For the rebound kinetics outcome, 69 participants with treatment failure from SEARCH-1 and STOP-HCV-1 were included, regardless of duration randomisation. For the time to treatment failure outcome, 242 participants from both trials were included, regardless of SVR12 status and duration randomisation. Although there may be clinical reasons as to why one factor was associated with one outcome and not another, it may also be due to a lack of power in the models containing fewer participants.

While the outcome of rebound kinetics is clearly distinct from the other outcomes, the outcomes of SVR12 and time to treatment failure after EOT are similar as the outcome of time to treatment failure incorporates if a participant failed as well as the timing of their treatment failure if they did fail. I investigated time to treatment failure in addition to SVR12 as, due to incorporating information about the timing of failure, this could theoretically have more power to identify factors associated with treatment failure. Those taking 8-week fixed length treatment were included to also increase power; however the inclusion of these participants may also have diluted the effect of some factors if the effect was not present or was reversed in these participants. Of note no strong factors were associated with SVR12 in these participants when they were considered separately in Chapter 2.

Considering the three different types of outcomes, baseline HCV VL was the only factor selected in all three models, with a very strong association with the outcome being investigated in each case ($p<0.001$) (Table 6.1). This indicates that it is likely to be the most important factor in determining which patients are suitable for short-course treatment. Low baseline VL (most commonly defined as <2,000,000 IU/ml) has been widely shown to be associated with SVR12 (44-47, 50), and will likely be part of any future criteria for selection for short-course DAA therapy. However, it is not clear what length treatment a patient needs given a specific baseline VL as few studies have attempted to trial different lengths of treatment based on baseline VL: apart from STOP-HCV-1, which used a sliding scale to determine days of treatment based on baseline VL, only one study restricted eligibility based on a low baseline VL (<2,000,000 IU/ ml) (42).

Within this thesis, baseline VL was considered only as a continuous factor and not dichotomised into groups that would receive either short-course or standard duration treatment or categorised into several groups each with different treatment lengths. Categorisation of continuous factors can lead to decreased power, which was already limited due to the small sample size, and it can also introduce residual confounding (143). However, a clear threshold determining which patients can receive short-course treatment is simpler to use in a clinical setting and is therefore more useful to clinicians. Further work could attempt to identify this threshold by comparing the discrimination of various VL thresholds for identifying SVR12 with short-course treatment using area under the ROC curves, and then further assessing the suitability of the thresholds by seeing how they perform in the models examining time to treatment failure and VL kinetics. The majority of previous studies using DAAs have used <2,000,000 IU/ml as the criteria for low baseline VL so this could be used as an initial threshold for investigation; however other studies have also used 800,000 IU/ml and 4,000,000 IU/ml (47, 50). Higher thresholds would not be possible to test due to the low number of participants enrolled in STOP-HCV-1 with high baseline VLs: only 11 participants had a baseline VL >6,000,000 IU/ml.

Having HCV genotype 1b was associated both with SVR12 and a longer time to treatment failure, however it was not associated with the kinetics of VL rebound in those who failed. It is possible that those infected with genotype 1a simply needed more days of treatment to cure, potentially due to being more likely to develop DAA resistance (58) with no other impact on VL if they actually failed. However, there may also have been a lack of power to detect a difference in peak rebound VL due to the smaller number of participants included in the kinetics analysis compared to the other analyses.

HCV genotype has previously been associated with cure in the literature, predominantly a lower chance of curing with genotype 3 and DAAs (in contrast with PEG-IFN alone where lowest cure rates occurred in genotype 1). Specifically for subtype 1b, one previous study had identified genotype 1b as associated with higher SVR12 (45) and another study which limited eligibility to those infected with genotype 1b as well as a rapid viral response (defined as <500 IU/ml two days after starting treatment) saw 100% SVR12 with only 3 weeks' treatment (41).

Although there was strong evidence for the effect of genotype 1b on SVR12 ($p<0.001$), it was selected in only 9% of bootstrap models for SVR12. As it was also selected as a factor for time to treatment failure with reasonably strong evidence ($p=0.02$) and has been previously identified in the literature, instead of HCV genotype being an unlikely factor, it is possible that it was not selected more frequently due to the instability of the bootstrap models. Another possibility is that genotype is highly confounded or mediated by other factors that were

selected in the bootstrap models instead. To determine the reason for why HCV genotype was not selected in more bootstrap models, the factors that were selected in models including genotype could have been compared with the models not including genotype. However, due to the large number of factors and number of bootstraps this may not have been a simple process.

Treatment length was also independently associated with both SVR12 and time to treatment failure: for the SVR12 model, this was through the selection of VUS strategy (4-6 weeks vs 4-7 weeks) and in the time to treatment failure model this was simply days of DAA treatment. As for HCV genotype, longer treatment was not associated with VL rebound kinetics in those who failed. Longer lengths of treatment, even within short-course therapy, are widely known to lead to higher SVR12 rates (78), however longer durations of treatment are not suitable for all patients and the higher risk of treatment failure with shorter treatment durations may be a beneficial compromise if the alternative is for the patient to receive no treatment.

Although it is known that longer treatments have a higher chance of cure, it can be important to quantify the difference in cure rate between different lengths of treatment depending on the reason for why short-course therapy is necessary. If a patient can only afford a specific length of treatment, it is important for them and the clinician to know their chance of curing on that length of treatment so that they can make an informed choice about whether it would be sensible to proceed or not, particularly as the patient would be unlikely to afford retreatment which would be longer in duration. However, it is less helpful in this scenario as a patient who would be unable to afford the extra length in DAA treatment would also be unlikely to afford the VL testing, genotype and DAA resistance testing required for an accurate prediction of a suitable short-course treatment length. A similar situation would occur in patients who would be unlikely to adhere to treatment in the community but may be able to receive monitored therapy for a short time, such as for an inpatient hospital admission; if it is known at admission the likely length of stay, if the probability of cure was high enough it may be advantageous to initiate DAA treatment as soon as possible. Additionally, if a patient has already received a given length of treatment but is likely to require a lot of support to complete a week, or more, of extra treatment, the increase in cure rate from the extra treatment can be considered alongside any costs to determine if the extra treatment is ultimately beneficial to the patient.

Table 6.1: Factors selected in each multivariable model (complete cases only)

	Predictors of SVR12 model (STOP-HCV-1 N=87; Chapter 2)	Time to treatment failure after EOT model (STOP-HCV-1 & SEARCH-1 N=242; Chapter 3)	VL rebound model (STOP-HCV-1 & SEARCH-1 N=69; Chapter 3)	
			Peak rebound VL	Slope from treatment failure to peak VL
Higher baseline HCV VL	-	-	+	
Genotype: 1b vs 1a	+	+		
IL28B: CC vs CT/TT	+			+
Baseline resistance to DAAs	-		-	
Current substance abuse	-			
Higher bilirubin	+			
Received ribavirin	+		+	
VUS strategy (VUS2 vs VUS1)/days of DAA treatment*	+	+		
Trial (SEARCH-1 vs STOP-HCV-1)**	N/A	+	+	
Older age		-		
Higher Fibroscan score			+	
>LLOQ HCV VL at EOT**	N/A		-	
Time of preceding visit to treatment failure**	N/A	N/A		-

*VUS strategy was only available for selection in the SVR12 model; days of DAA treatment was available for selection in the time to treatment failure and kinetics model. Although not measuring exactly the same thing, they are combined in one row here as both are measuring longer treatment.
**Trial and >LLOQ HCV VL at EOT were only available for selection in the time to treatment failure and VL rebound models which included data from SEARCH-1 as well as STOP-HCV-1; time of preceding visit to treatment failure was only available for selection in the rebound model.
Note: +/- indicates factor as a positive/negative effect on the outcome: for SVR12 and time to failure this is increased risk of failure or shorter time to failure and for rebound model is change in VL.

Both baseline resistance to prescribed DAAs and ribavirin were associated with SVR12 and peak rebound VL, but not time to treatment failure. Those with baseline DAA resistance were more likely to have a treatment failure, but had a lower peak rebound VL when they failed, whereas participants who took ribavirin were less likely to fail, but had a higher peak rebound VL when they failed. A lower peak rebound VL indicates that there may well be a fitness cost to being less susceptible to DAAs that hinders viral replication. It is possible that although higher levels of virus or infected cells remain in those with DAA resistance, treatment failure is detected at a similar time to those without DAA resistance due to the slower replication. DAA resistance was not found to be associated with time to treatment failure; this may be explained by the inclusion of SEARCH-1 participants who had higher levels of DAA resistance than those in STOP-HCV-1 (92% vs 14%) and participants who received 8-weeks of treatment in STOP-HCV-1 diluting the effect of DAA resistance that was identified in those receiving 4-7 weeks variable length treatment in STOP-HCV-1 alone.

Resistance to DAAs is a factor that has previously been identified as associated with SVR12 in short-course treatment (45, 47); however I could not identify a reason in the literature why ribavirin why would lead to a higher peak VL. The higher peak VL may be related to the lower rates of emergent resistance to prescribed DAAs observed in those who took ribavirin in the STOP-HCV-1 trial, which would explain why its effect on peak VL in those who fail is opposite to the effect of baseline resistance. If the rebounding virus in those who did not take ribavirin had more mutations than in those who did due to higher levels of emergent resistance, if the mutations had a fitness cost that impacted viral replication, then a lower peak rebound VL would be expected in those who did not take ribavirin than those who did. Models may simply have identified ribavirin rather than resistance as the factor most strongly associated with outcome due to small numbers and instability or chance.

Having the CC variant of the IL28B gene was weakly associated with SVR12 ($p=0.09$) and was also associated with the rate of change in VL between treatment failure and peak VL, but was not associated with time to treatment failure. Previous research has shown an association between this variant and both SVR12 with treatment and spontaneous cure in some patients (47, 50, 59). As those with IL28B CC have a stronger immune response to HCV, it is unclear why they would have a faster rebound; it may be due to treatment failure being detected at a different time point in the rebound than in those with genotype CT or TT, and so instead of comparing the same point of the trajectory curve for both groups, it is instead comparing different parts of the trajectory curve. It is unlikely that viral rebound occurs at a constant rate for the entirety of its trajectory and it may be faster at times than others; thus it is possible that, for those with genotype CC, their treatment failure is being detected at a time point

when the trajectory curve is steeper than for those with CT or TT, but ultimately their overall rebound trajectories are similar.

Although IL28B genotype was not associated with time to treatment failure, this may be partly because different significance thresholds were used in for model selection in the SVR12 models and time to failure models: the SVR12 model used $p<0.157$ (AIC) and the time to failure model used $p<0.05$. If the same threshold of $p<0.157$ was used for both models, IL28B would be selected as a factor in the time to failure model ($p=0.11$). However, as power should be higher in the time to failure model, but the strength of the evidence decreased, it is possible that IL28B is not genuinely associated with SVR12 and was only selected in the SVR12 model due to the higher significance threshold, and so is a false-positive finding.

The trial a participant was enrolled into was associated with the time to treatment failure and viral rebound kinetics, both models where data was combined from the two trials. It is not clear what aspects of the trials this is acting for a proxy for, but it is likely to at least in part represent participants with HCV genotype 6, who were only eligible for SEARCH-1, as the effect of trial reduced when genotype 6 was forced in the kinetics model. As trial was selected in addition to HCV genotype in the time to failure model (although there was no evidence of a difference between genotype 6 and 1a, $p=0.39$), it is likely to represent additional factors such as other aspects of patient management and potentially also adherence to treatment. For example, 10% of SEARCH-1 participants reported missing doses, compared to 28% in STOP-HCV-1. As trial is a proxy for several factors, including potentially some that were not offered for selection in the model, there will be higher power to detect the association with trial overall compared to any individual factors offered to the model, and therefore trial may have been selected instead of these factors.

All other factors identified were associated with only one of the different outcomes I considered. It is possible that these factors were only related to this outcome alone, or that their effects were being mediated through other factors that were selected in the other models. It is also possible that they were not identified due to the different demographics of the two trials, reducing power to detect an effect in models combining both. For example, although current substance abuse was identified as associated with SVR12 in STOP-HCV-1 participants, no SEARCH-1 participant reported current substance abuse and therefore the power to detect the association in the larger data set would have been lower and the effect of current substance abuse diluted.

All the associations identified and discussed here are patient baseline demographics or aspects of treatment and trial design. Although on treatment and end of treatment factors were not

considered for the time to treatment failure and rebound kinetics models, apart from viral suppression at EOT and the timing of the VL measurement preceding the treatment failure visit, they were investigated as factors associated with SVR12 and none were independently associated. However, post-baseline factors were only considered after baseline factors had already been selected. It is possible that if these factors had been explored alongside the baseline factors, they may have been chosen instead of the baseline factors. However, due to the limited sample size, I chose not to test all potential factors at the same time and any future work that did consider baseline and post-baseline factors would likely need a larger sample size that also had a large proportion of treatment failures.

Other work within the SEARCH-1 trial has detected an association between change in ALT or AST and treatment failure, but this was the change from EOT to EOT+12 and no adjustment was made for any other baseline, on- or post-treatment factors (144). Additionally, while post-treatment factors may be helpful in providing a cheaper or simpler way of identifying treatment failure than monitoring VLs, it is less helpful in determining which patients need longer treatment as treatment will have stopped several weeks earlier and treatment failure may have already taken place.

6.3 An update to the VIETNARMS trial and design

VIETNARMS was designed as a factorial trial with two full factorial randomisations (sofosbuvir/velpatasvir vs sofosbuvir/daclatasvir; standard 12-week duration vs 4-week treatment with adjunctive PEG-IFN vs 4–12-week response guided therapy vs 12-week induction/maintenance) and one partial factorial randomisation (adjunctive ribavirin vs none, in those receiving one of the three drug shortening strategies only). Randomisation was also planned to be stratified by HCV genotype (6 vs not 6). As described in Chapter 5, there was a Bayesian stopping guideline specifying that recruitment into a group would stop if there was a >0.95 posterior probability of the true cure rate being <90%, with the primary final analysis performed only on those groups that had not closed to recruitment. It planned to recruit 1092 participants at one site in Ho Chi Minh City in Vietnam over a period of two years, with four interim analyses planned during the recruitment phase.

At the time of performing the analysis for and writing Chapter 5, the VIETNARMS trial was not yet open to recruitment. Since then, the trial opened to recruitment on 19th June 2020 and recruited its last participant on 10th May 2023 with a planned first-line last participant last visit in October 2023, although retreatment visits may take place after this time point. Recruitment was not stopped in any group due to meeting the stopping guidelines I developed.

Due to the COVID-19 pandemic and local lockdowns in Vietnam, recruitment became difficult and had to pause at times. To help improve recruitment, a second site opened in Hanoi, which, due to local variability in case rates and timing of lockdowns, could often recruit when the main site in Ho Chi Minh City was unable to, and similarly for the Ho Chi Minh City site when Hanoi was unable to recruit. Given the overall challenges in recruiting with COVID-19, the Trial Steering Committee agreed that recruitment into the ribavirin groups should stop as this would reduce the overall sample size considerably from 1092 to 624 and data from the STOP-HCV-1 trial suggested there was unlikely to be a benefit from adjunctive ribavirin, whereas there was no such data for the shortening strategies. This change in design had no impact on the operating characteristics of the stopping guidelines as there was no change in size to the remaining groups and there was no substantial change to the overall power at the final analysis for the remaining comparisons.

For the timing of interim analyses, it was no longer possible for the fourth review to take place as the planned total recruitment at that analysis was larger than the new total sample size. It was possible for the DMC to have chosen to have the fourth interim analysis at an earlier than planned time point, or any other additional interim analysis at any time of their choosing, but this did not happen and only three interim analyses took place.

Table 6.2: Comparison of planned and actual interim analyses timing

	First analysis		**Second analysis**		**Third analysis**	
	Planned	**Actual**	**Planned**	**Actual**	**Planned**	**Actual**
Months since recruitment started	7	13	10	19	13	27
Total recruited/included in analysis (out of maximum 624)	205	192/170	361	262/203	517	520/415
Total reached EOT+12 weeks	44	57	144	170	286	279
At EOT+12 weeks in:*						
PEG-IFN group	5	7	14	18	24	36
RGT group	3	5	11	16	21	34
Induction/maintenance group	2	7	8	19	17	31

*For actual interim analysis, this is the mean of the groups for each strategy.

A comparison of the timing of and outcomes contributing to each interim analysis as planned and as happened is in Table 6.2. Due to the slower than anticipated recruitment, interim analyses could not take place at the original time points based on the start date of recruitment and were instead based on the timing of the expected number of participants reaching 12 weeks after the end of first-line treatment, which could be estimated earlier from recruitment

numbers. There was a slight discrepancy in the planned and actual number of outcomes due to how recruitment numbers were monitored and the planning of the data entry deadlines and the DMC meeting.

Although the total number of outcomes was similar between the planned and actual interim analyses, the total in each group was much higher in the actual analyses. This is because the participants were no longer being distributed between ribavirin and non-ribavirin groups but were only allocated to non-ribavirin groups and so these increased faster than they were originally planned to. The monitoring plan did not account for recruitment into groups being stopped, especially into a large number of groups, and this would have been difficult to incorporate due to the uncertainties of which groups would be successful at the design stage. This difference meant that the probability of stopping recruitment into a failing group would have been higher than planned and, as no group met the stopping guideline, this gives extra support that the strategies have performed well. However, if recruitment into a group had stopped, it would have stopped later than planned and fewer participants would have been recruited into more successful groups.

The analysis of the first-line data will be performed shortly after the last first-line visit, currently expected October 2023. Analysis of retreatment data will be performed either at the same time or at a later time point, depending on when the last treatment failure occurs (the latest time point a retreatment visit could theoretically occur will be April 2024, if the last recruited participant fails at their very last first-line visit).

6.4 Strengths and limitations

The strength of this thesis is that very little previous work has examined data from short-course HCV treatment trials to this extent. Some previous trials investigating short-course treatment have identified some participant factors that were associated with SVR12, mostly baseline VL, but these trials were smaller in size and the range of factors tested was more limited (see Section 2.2 for more detail on these studies). While studies of standard duration DAA treatment were larger, the cure rate was so high it would be very difficult to identify many factors, unless they had implausibly large effect sizes. Although STOP-HCV-1 did not fully recruit, its low cure rate combined with its moderate size provided me with an opportunity to identify factors associated with failure that other studies could not. I could also find no previous research into investigating factors associated with time to treatment failure with DAAs or into the kinetics of viral load rebound. While many previous studies have examined QoL during and after DAA treatment, none were conducted using short-course HCV treatment.

As the VIETNARMS design was novel, there was no previous information about how the design would work and my assessment of the statistical aspects of the trial were important to ensure the trial functioned appropriately to collect valid results. Another strength of the analysis of the VIETNARMS design is that, alongside the design itself, it is highly generalisable to testing other interventions in many different disease areas.

The main limitation of the thesis is the small sample size in the STOP-HCV-1 and SEARCH-1 trials. Although including larger numbers than other studies of short-course treatment, it has limited the analysis I was able to perform and, ultimately, has limited the scope of the factors I was able to identify, my confidence in the associations I found and the potential effects of treatment on QoL. One analysis choice I made, partly due to the small sample size, was not to investigate interactions between the factors identified, which is another weakness of the thesis. Another limitation is that the two main factors representing length of treatment, VUS1/VUS2 in STOP-HCV-1 and days of DAA treatment in SEARCH-1, were not randomised comparisons and there may be unmeasured confounders that affect the estimates of their effect. The primary confounder not accounted for in STOP-HCV-1 is calendar time as VUS2 was implemented after VUS1, with no overlap in the strategies, meaning there is complete confounding. However, in the QoL analysis, the comparison between variable and fixed duration was randomised and results for this comparison were largely similar to that of the VUS1 vs VUS2 comparison, in that there were no clear effects on QoL regardless of treatment duration.

Another potential limitation of the thesis is that the results from Chapters 2-4 may not be generalisable to other strategies for short-course HCV treatment, particularly as the strategy used in STOP-HCV-1 was unique and used DAAs that are not one of the pangenotypic DAA regimens now recommended by the WHO and therefore are not now widely used. However, the inclusion of SEARCH-1 data in Chapter 3 increases the generalisability, as SEARCH-1 used a different strategy based on treatment response rather than baseline VL and a commonly used WHO-recommended drug regimen. Additionally, the factors identified in these Chapters have mostly been identified in the literature previously and none contradict those previously established.

The factors associated with SVR12 and VL rebound identified within this thesis, particularly the key factors, are likely to be generalisable to the wider population of HCV patients with mild liver disease, with the exception of HCV genotype, as these are generally consistent with previous research. The inclusion of only a limited number of genotypes (1a and 1b throughout, 6 in Chapter 3) affects global generalisability due to the global distribution of genotypes (see Chapter 1 for more information); it is known that SVR12 rates are lower in genotype 3 and

there is a limited amount of previous research into other genotypes. It is also unlikely that these results will be generalisable to those with more advanced liver disease as fibrosis and cirrhosis are key factors determining SVR12 and may reduce the effects of the factors identified here. Short course treatment will only be suitable for a small group of patients and primarily those who will not be able to adhere to the standard-length treatment, so generalisability towards these patients is most important. While clinically these factors may be generalisable to this group, they are more likely to have more complex needs, including social needs, that have not been fully accounted for within this analysis. However, it would be difficult to research what, if any, factors would be different within this group of patients as, for similar reasons to their need for short-course treatment, they would not be suitable for inclusion in a clinical trial and observational studies may also lack data.

6.5 Future work

In each chapter I have discussed further work that could extend the analysis contained in the chapter. Overall, the key piece of work that could take place in future would be to incorporate the data from the VIETNARMS trial into the analysis performed in this thesis. Once the trial has finished follow-up and been analysed, how the trial monitoring approach I developed performed in real-life could also be assessed fully and compared to what was planned in Chapter 5. If combined with data from the STOP-HCV-1 and SEARCH-1 trials, the power to detect important factors associated with the outcomes in Chapters 2 and 3 would increase greatly. The visit schedule in VIETNARMS is equivalent to that of STOP-HCV-1, although this may have been disrupted in some participants due to COVID-19 lockdowns, so it may be easier to combine the data from these two trials to investigate the kinetics of viral rebound and time to treatment failure than with SEARCH-1. Due to the larger sample size of VIETNARMS than SEARCH-1, it may also be easier to find differences between UK and Vietnamese HCV patients that was not possible in this thesis. Alternatively, the model for predicting SVR12 in Chapter 2 could be validated using the VIETNARMS data. However, as the stopping guideline was not met for any randomised group in VIETNARMS, the rate of treatment failure may be too low within the trial to contribute meaningful information about which patients are suitable for short-course treatment.

Since I started working on this thesis topic, the cost of DAAs has decreased and more countries have access to cheap generics using licensing agreements and the cost of treatment is no longer a barrier for many patients still requiring treatment. A cost-effectiveness analysis could be performed using the current cost of DAAs and any projected further reductions in price to determine whether short-course treatment is able to lower treatment costs either for an individual or for a public health system. This cost-effectiveness analysis could be performed for

both the targeted strategy of identifying suitable patients via the models built in Chapter 2, with the additional costs of HCV genotyping and DAA resistance testing for the more discriminative models, and the more general strategies being tested in VIETNARMS, which require extra costs for PEG-IFN or increased VL testing for the response guided therapy strategy.

Participants in the VIETNARMS trial are asked to complete the EQ-5D-**3L** QoL questionnaire at the same time points as the participants in the STOP-HCV-1 trial were asked to complete the EQ-5D-**5L** and, after reducing the five levels in the EQ-5D-5L to three, it would be possible to combine the responses from the two trials. However, responses to these questionnaires are subjective and there may be cultural differences that mean the results in each country are not directly comparable. Most of the studies investigating HCV have recruited participants from Western high-income countries (HICs) (Europe, North America and Oceania) with two studies in East Asia (China, South Korea and Taiwan) (96, 105), and one each in Brazil, Indonesia and Iran (92, 94, 98). Data on QoL from VIETNARMS could therefore contribute valuable insight into the QoL of HCV patients after treatment in South-East Asian and low- and middle-income countries (LMICs).

This thesis has focused on short-course HCV treatment as a method to increase rates of SVR12 in hard-to-reach populations; however, length of and adherence to treatment are not the only barriers to achieving the WHO's goals of elimination of HCV as a global threat by 2030. The WHO has identified that to meet these targets, both diagnostics and service delivery need to be simplified and decentralised (145). Cost can often be a barrier to decentralisation as it may require duplication of expensive services, such as viral load or genotype testing, and alternatives must be found to replace these. One suggestion to replace viral load testing to identify treatment failures is to monitor the change in patients' ALT or AST levels instead as an increase from EOT to EOT+12 has been shown to be highly sensitive in detecting treatment failures and could save US$18-$46,000 per thousand patients treated in Vietnam (144). This approach could be tested in a large scale trial.

Within this thesis only 3/69 (4%) participants had a treatment failure at EOT+12, with an additional 2 participants with a treatment failure at EOT+24 who were considered cured within the thesis as they achieved SVR12. Currently, SVR12 is the standard definition for cure, but as there were very few treatment failures after EOT+8, one potential way to simplify delivery of treatment is to consider SVR8 as a marker of cure instead of SVR12. This may be beneficial for participants or settings where the risk of loss to follow-up increases over time to increase detection of any treatment failures. However, this may only be suitable for short-course DAAs and would need to be verified as a proxy for cure over longer time periods using larger

datasets. It may also come with additional risks of patients who have a treatment failure after EOT+8 who believe that they are cured that would need to be considered alongside the benefits of detecting more treatment failures in those who would otherwise be unaware of their status.

However, before those with HCV can be treated, they need to be identified. The WHO estimated in 2019 that 79% of those living with HCV, around 45.8 million people, were unaware of their status and therefore unable to initiate treatment (8). Large screening programmes will be needed to identify these people, who are likely to be in the hardest-to-reach groups or in low-income countries, and financial and societal support given to help them achieve cure. If the strategies tested in VIETNARMS have high cure rates, these could help in improving access to treatment for these populations. Although VIETNARMS has defined a successful strategy using a non-inferiority margin of 10%, that is the lower confidence bound of a cure rate in a shortening strategy group should be no larger than 10% below the cure rate of the control group, a strategy that performs slightly worse than this may be also of interest to clinicians if they could be used for particularly difficult-to-reach groups. The more strategies are successful, the more options clinicians have for treating different groups of patients that may have different needs, making it easier to treat a larger number of people.

6.6 Conclusion

In this thesis I have aimed to identify patients which a clinician could consider for short-course treatment, which treatment options are suitable (considering differing treatment lengths and adjunctive ribavirin) and to develop a trial design for testing strategies for short-course treatment which could work on a population level, regardless of patient demographics. However, which strategy can be used, whether the personalisation of treatment based on a patient's characteristics or a broad approach as used in VIETNARMS, will also depend on the setting in which it is to be used as well as the patient group it is to be used in: more options will be available to those in HICs rather than LMICs.

I have shown that baseline HCV VL is the key determinant of which patients can successfully cure using short-course treatment; a higher baseline VL was associated with worse outcomes in all three areas of time to treatment failure, rebound VL kinetics and treatment failure overall. Baseline VL was not part of the strategies used in VIETNARMS, but on-treatment response, which may be related to baseline VL, was. I have also identified that having HCV genotype 1b and IL28B genotype CC are also more likely to cure on short-course treatment, while having baseline resistance to prescribed DAAs are strong indicators that a patient may not cure with short durations of treatment.

As baseline VL is so important, it should be included as part of any criteria for determining which patients are suitable for short-course treatment, particularly in HICs where VL testing is widely available and affordable. In LMICs, where there may be limited access to VL testing due to either costs or access, this may not be as feasible and therefore selection of patients for short-course treatment may not be as successful. However, performing one VL at baseline may save more costs in terms of DAA treatment or from the expense of not curing a patient, from either treatment of the sequalae of chronic HCV or from further spread of the virus to other individuals, so it may still be cost effective to test baseline VL. HCV genotype testing, IL28B testing and testing for DAA resistance are even more difficult and expensive in LMICs, furthering the importance of baseline VL testing if possible. There are point of care tests for IL28B genotype that can increase access to testing as it does not require a laboratory (146), but these are still expensive and may only be suitable for HICs. In LMICS, it may not be possible to identify patients for short-course treatment based on their characteristics, and one of the broader approaches tested in VIETNARMS would be more appropriate.

After treatment failure, viral load rebounds to a similar level or slightly lower to that at baseline, but this may be affected by HCV genotype, baseline DAA resistance, the use of ribavirin, baseline Fibroscan score and whether the patient was suppressed at end of treatment. The higher peak rebound VL in those with HCV genotype 1a and a higher Fibroscan score could indicate that these patients should be given longer treatment than those with genotype 1b and lower Fibroscan scores. However, those with resistance to prescribed DAAs also had a lower peak rebound VL and they should not be given short-course treatment in most patients, as seen in the probabilities of SVR12 I generated. However, this may explain why retreatment is so successful in these patients. There are limitations to this analysis that assumes that treatment failure is identified at a similar time point in all participants in the trials and the same length of time from treatment failure being identified to peak rebound VL. Further work should be done to separate out the effect of demographic factors from those caused by the timing of the detection of treatment failure before clinicians decide on treatment options based on VL rebound.

Patients who take longer treatment are more likely to cure and the longer length does not impact their quality of life, so longer treatments can be considered if poor adherence can be mitigated in other ways, such as the induction/maintenance approach in VIETNARMS. Although ribavirin is often poorly tolerated, there is no strong evidence this impacts quality of life and it has no lasting effects, so it could be considered for patients where it may increase the probability of cure. As any negative effects on physical health are mediated through

ribavirin associated anaemia, patients could also be monitored closely and any anaemia promptly treated to minimise a reduction in quality of life.

The knowledge gained from the results generated in this thesis could be used to determine criteria to select the patients most suitable for short-course treatment and this could in theory be tested in a trial design similar to VIETNARMS, which was shown to be appropriately designed with high power and an effective stopping mechanism. However, it is unlikely that further large scale trials into HCV treatment will take place due to the wider availability of DAAs given reductions in their costs that previously may have incentivised patients to join trials to access treatment. Although there are still hard-to-reach populations who need treatment, they are unlikely to be suitable participants for a clinical trial for the same reasons they need short-course treatment. As the majority of patients can afford standard duration treatment and are able to adhere to treatment and achieve cure, drug shortening strategies will likely only be used with patients who would otherwise not cure and, in this small proportion of cases, there may be sufficient evidence to provide clinicians confidence in prescribing short-course treatment to those I have identified could benefit.

7 References

1. Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science (New York, NY). 1989;244(4902):359-62.
2. Polaris Observatory HCV Collaborators. Global change in hepatitis C virus prevalence and cascade of care between 2015 and 2020: a modelling study. Lancet Gastroenterol Hepatol. 2022;7(5):396-415.
3. Stanaway JD, Flaxman AD, Naghavi M, Fitzmaurice C, Vos T, Abubakar I, et al. The global burden of viral hepatitis from 1990 to 2013: findings from the Global Burden of Disease Study 2013. Lancet (London, England). 2016;388(10049):1081-8.
4. Polaris Observatory H. C. V. Collaborators. Global prevalence and genotype distribution of hepatitis C virus infection in 2015: a modelling study. The Lancet Gastroenterology & Hepatology. 2017;2(3):161-76.
5. World Health Organization. Global health sector strategy on viral hepatitis 2016-2021. http://apps.who.int/iris/bitstream/handle/10665/246177/WHO-HIV-2016.06-eng.pdf 2016.
6. Ayoub HH, Chemaitelly H, Omori R, Abu-Raddad LJ. Hepatitis C virus infection spontaneous clearance: Has it been underestimated? International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases. 2018;75:60-6.
7. Aisyah DN, Shallcross L, Hully AJ, O'Brien A, Hayward A. Assessing hepatitis C spontaneous clearance and understanding associated factors-A systematic review and meta-analysis. Journal of viral hepatitis. 2018;25(6):680-98.
8. World Health Organization. Global progress report on HIV, viral hepatitis and sexually transmitted infections, 2021. https://www.who.int/publications/i/item/9789240027077: World Health Organization; 2021.
9. Backus LI, Belperio PS, Shahoumian TA, Mole LA. Direct-acting antiviral sustained virologic response: Impact on mortality in patients without advanced liver disease. Hepatology (Baltimore, Md). 2018;68(3):827-38.
10. Chekuri S, Nickerson J, Bichoupan K, Sefcik R, Doobay K, Chang S, et al. Liver Stiffness Decreases Rapidly in Response to Successful Hepatitis C Treatment and Then Plateaus. PLoS One. 2016;11(7):e0159413-e.
11. McHutchison JG, Poynard T. Combination therapy with interferon plus ribavirin for the initial treatment of chronic hepatitis C. Seminars in liver disease. 1999;19 Suppl 1:57-65.
12. Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, et al. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. Lancet. 2001;358(9286):958-65.
13. Carter W, Connelly S, Struble K. Reinventing HCV Treatment: Past and Future Perspectives. J Clin Pharmacol. 2017;57(3):287-96.
14. Zeuzem S, Feinman SV, Rasenack J, Heathcote EJ, Lai MY, Gane E, et al. Peginterferon alfa-2a in patients with chronic hepatitis C. The New England journal of medicine. 2000;343(23):1666-72.
15. McHutchison JG, Gordon SC, Schiff ER, Shiffman ML, Lee WM, Rustgi VK, et al. Interferon alfa-2b alone or in combination with ribavirin as initial treatment for chronic hepatitis C. Hepatitis Interventional Therapy Group. The New England journal of medicine. 1998;339(21):1485-92.
16. Bodenheimer HC, Jr., Lindsay KL, Davis GL, Lewis JH, Thung SN, Seeff LB. Tolerance and efficacy of oral ribavirin treatment of chronic hepatitis C: a multicenter trial. Hepatology (Baltimore, Md). 1997;26(2):473-7.

17. Yu M-L, Dai C-Y, Huang J-F, Chiu C-F, Yang Y-HC, Hou N-J, et al. Rapid virological response and treatment duration for chronic hepatitis C genotype 1 patients: a randomized trial. Hepatology (Baltimore, Md). 2008;47(6):1884-93.
18. Ferenci P, Laferl H, Scherzer T-M, Gschwantler M, Maieron A, Brunner H, et al. Peginterferon alfa-2a and ribavirin for 24 weeks in hepatitis C type 1 and 4 patients with rapid virological response. Gastroenterology. 2008;135(2):451-8.
19. World Health Organization. Guidelines for the Care and Treatment of Persons Diagnosed with Chronic Hepatitis C Virus Infection. https://apps.who.int/iris/bitstream/handle/10665/273174/9789241550345-eng.pdf: World Health Organization; 2018.
20. Falade-Nwulia O, Suarez-Cuervo C, Nelson DR, Fried MW, Segal JB, Sulkowski MS. Oral Direct-Acting Agent Therapy for Hepatitis C Virus Infection: A Systematic Review. Annals of internal medicine. 2017;166(9):637-48.
21. Spengler U. Direct antiviral agents (DAAs) - A new age in the treatment of hepatitis C virus infection. Pharmacol Ther. 2018;183:118-26.
22. Guedj J, Dahari H, Rong L, Sansone ND, Nettles RE, Cotler SJ, et al. Modeling shows that the NS5A inhibitor daclatasvir has two modes of action and yields a shorter estimate of the hepatitis C virus half-life. Proc Natl Acad Sci U S A. 2013;110(10):3991-6.
23. Bagaglio S, Uberti-Foppa C, Morsica G. Resistance Mechanisms in Hepatitis C Virus: implications for Direct-Acting Antiviral Use. Drugs. 2017;77(10):1043-55.
24. Médecins Sans Frontières. Strategies to secure access to generic hepatitis C medicines: overcoming patent and regulatory barriers to secure access to generic hepatitis C medicines. https://msfaccess.org/strategies-secure-access-generic-hepatitis-c-medicines2015.
25. Douglass CH, Pedrana A, Lazarus JV, t Hoen EFM, Hammad R, Leite RB, et al. Pathways to ensure universal and affordable access to hepatitis C treatment. BMC Med. 2018;16(1):175-.
26. Barber MJ, Gotham D, Khwairakpam G, Hill A. Price of a hepatitis C cure: Cost of production and current prices for direct-acting antivirals in 50 countries. J Virus Erad. 2020;6(3):100001.
27. Cooke GS, Pett S, McCabe L, Jones C, Gilson R, Verma S, et al. Strategic treatment optimization for HCV (STOPHCV1): a randomised controlled trial of ultrashort duration therapy for chronic hepatitis C. Wellcome Open Res. 2021;6:93.
28. Flower B, Hung LM, McCabe L, Ansari MA, Le Ngoc C, Vo Thi T, et al. Efficacy of ultra-short, response-guided sofosbuvir and daclatasvir therapy for hepatitis C in a single-arm mechanistic pilot study. Elife. 2023;12.
29. McHutchison JG, Lawitz EJ, Shiffman ML, Muir AJ, Galler GW, McCone J, et al. Peginterferon alfa-2b or alfa-2a with ribavirin for treatment of hepatitis C infection. The New England journal of medicine. 2009;361(6):580-93.
30. Bakr I, Rekacewicz C, El Hosseiny M, Ismail S, El Daly M, El-Kafrawy S, et al. Higher clearance of hepatitis C virus infection in females compared with males. Gut. 2006;55(8):1183-7.
31. Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Gonçales FL, Jr., et al. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. The New England journal of medicine. 2002;347(13):975-82.
32. Ge D, Fellay J, Thompson AJ, Simon JS, Shianna KV, Urban TJ, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature. 2009;461(7262):399-401.
33. Arias A, Aguilera A, Soriano V, Benítez-Gutiérrez L, Lledó G, Navarro D, et al. Rate and predictors of treatment failure to all-oral HCV regimens outside clinical trials. Antiviral therapy. 2017;22(4):307-12.
34. Buggisch P, Vermehren J, Mauss S, Günther R, Schott E, Pathil A, et al. Real-world effectiveness of 8-week treatment with ledipasvir/sofosbuvir in chronic hepatitis C. Journal of hepatology. 2018;68(4):663-71.

35. Belperio PS, Shahoumian TA, Loomis TP, Mole LA, Backus LI. Real-world effectiveness of daclatasvir plus sofosbuvir and velpatasvir/sofosbuvir in hepatitis C genotype 2 and 3. Journal of hepatology. 2019;70(1):15-23.
36. Backus LI, Belperio PS, Shahoumian TA, Loomis TP, Mole LA. Real-world effectiveness of ledipasvir/sofosbuvir in 4,365 treatment-naive, genotype 1 hepatitis C-infected patients. Hepatology (Baltimore, Md). 2016;64(2):405-14.
37. Benhammou JN, Dong TS, May FP, Kawamoto J, Dixit R, Jackson S, et al. Race affects SVR12 in a large and ethnically diverse hepatitis C-infected patient population following treatment with direct-acting antivirals: Analysis of a single-center Department of Veterans Affairs cohort. Pharmacology research & perspectives. 2018;6(2):e00379-e.
38. Terrault NA, Zeuzem S, Di Bisceglie AM, Lim JK, Pockros PJ, Frazier LM, et al. Effectiveness of Ledipasvir-Sofosbuvir Combination in Patients With Hepatitis C Virus Infection and Factors Associated With Sustained Virologic Response. Gastroenterology. 2016;151(6):1131-40.e5.
39. Jansen JW, Linneman TW, Powderly GM, Moenster RP, Nayak L. Association Between Baseline Creatinine Clearance and Treatment Failure in Patients With Hepatitis C Virus Treated With Ledipasvir and Sofosbuvir. Open Forum Infect Dis. 2019;6(3):ofz087-ofz.
40. Persico M, Aglitti A, Milella M, Coppola C, Messina V, Claar E, et al. Real-life glecaprevir/pibrentasvir in a large cohort of patients with hepatitis C virus infection: The MISTRAL study. Liver international : official journal of the International Association for the Study of the Liver. 2019;39(10):1852-9.
41. Lau G, Benhamou Y, Chen G, Li J, Shao Q, Ji D, et al. Efficacy and safety of 3-week response-guided triple direct-acting antiviral therapy for chronic hepatitis C infection: a phase 2, open-label, proof-of-concept study. Lancet Gastroenterol Hepatol. 2016;1(2):97-104.
42. Ovrehus ALH, Krarup H, Birkemose I, Holm DK, Mossner B, Ernst A, et al. Four weeks of ledipasvir/sofosbuvir and ribavirin with or without pegylated interferon for chronic hepatitis C in non-cirrhotic people who inject drugs. A randomized trial. J Hepatol. 2017.
43. Madsen LW, Christensen PB, Fahnøe U, Pedersen MS, Bukh J, Øvrehus A. Inferior cure rate in pilot study of 4-week glecaprevir/pibrentasvir treatment with or without ribavirin of chronic hepatitis C. Liver Int. 2021;41(11):2601-10.
44. Madsen LW, Christensen PB, Hansen JF, Røge BT, Holm DK, Dröse S, et al. Four Weeks Treatment with Glecaprevir/Pibrentasvir + Ribavirin—A Randomized Controlled Clinical Trial. Viruses. 2022;14(3):614.
45. Kohli A, Kattakuzhy S, Sidharthan S, Nelson A, McLaughlin M, Seamon C, et al. Four-Week Direct-Acting Antiviral Regimens in Noncirrhotic Patients With Hepatitis C Virus Genotype 1 Infection: An Open-Label, Nonrandomized Trial. Annals of internal medicine. 2015;163(12):899-907.
46. Sulkowski MS, Flamm S, Kayali Z, Lawitz EJ, Kwo P, McPhee F, et al. Short-duration treatment for chronic hepatitis C virus with daclatasvir, asunaprevir, beclabuvir and sofosbuvir (FOURward study). Liver Int. 2017;37(6):836-42.
47. Lawitz E, Poordad F, Gutierrez JA, Wells JT, Landaverde CE, Evans B, et al. Short-duration treatment with elbasvir/grazoprevir and sofosbuvir for hepatitis C: A randomized trial. Hepatology. 2017;65(2):439-50.
48. Gane EJ, Schwabe C, Hyland RH, Yang Y, Svarovskaia E, Stamm LM, et al. Efficacy of the Combination of Sofosbuvir, Velpatasvir, and the NS3/4A Protease Inhibitor GS-9857 in Treatment-Naïve or Previously Treated Patients With Hepatitis C Virus Genotype 1 or 3 Infections. Gastroenterology. 2016;151(3):448-56.e1.
49. Kattakuzhy S, Wilson E, Sidharthan S, Sims Z, McLaughlin M, Price A, et al. Moderate Sustained Virologic Response Rates With 6-Week Combination Directly Acting Anti-Hepatitis C Virus Therapy in Patients With Advanced Liver Disease. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2016;62(4):440-7.

50. Sulkowski MS, Feld JJ, Lawitz E, Felizarta F, Corregidor AM, Khalid O, et al. Efficacy and safety of 6 or 8 weeks of simeprevir, daclatasvir, sofosbuvir for HCV genotype 1 infection. Journal of viral hepatitis. 2018;25(6):631-9.
51. Zeuzem S, Bourgeois S, Greenbloom S, Buti M, Aghemo A, Lampertico P, et al. JNJ-4178 (AL-335, Odalasvir, and Simeprevir) for 6 or 8 Weeks in Hepatitis C Virus-Infected Patients Without Cirrhosis: OMEGA-1. Hepatology (Baltimore, Md). 2019;69(6):2349-63.
52. Chua JV, Ntem-Mensah A, Abutaleb A, Husson J, Mutumbi L, Lam KW, et al. Short-duration treatment with the novel non-nucleoside inhibitor CDI-31244 plus sofosbuvir/velpatasvir for chronic hepatitis C: An open-label study. J Med Virol. 2021;93(6):3752-60.
53. Morris TP, White IR, Carpenter JR, Stanworth SJ, Royston P. Combining fractional polynomial model building with multiple imputation. Stat Med. 2015;34(25):3298-317.
54. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological). 1996;58(1):267-88.
55. Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, et al. Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal. 2018;21(1):C1-C68.
56. Vallet-Pichard A, Mallet V, Nalpas B, Verkarre V, Nalpas A, Dhalluin-Venier V, et al. FIB-4: an inexpensive and accurate marker of fibrosis in HCV infection. comparison with liver biopsy and fibrotest. Hepatology (Baltimore, Md). 2007;46(1):32-6.
57. Wyles DL, Luetkemeyer AF. Understanding Hepatitis C Virus Drug Resistance: Clinical Implications for Current and Future Regimens. Top Antivir Med. 2017;25(3):103-9.
58. Wyles DL, Gutierrez JA. Importance of HCV genotype 1 subtypes for drug resistance and response to therapy. J Viral Hepat. 2014;21(4):229-40.
59. Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, O'Huigin C, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature. 2009;461(7265):798-801.
60. Zamor PJ, Brown A, Dylla DE, Dillon JF, Luetkemeyer AF, Feld JJ, et al. High Sustained Virologic Response Rates of Glecaprevir/Pibrentasvir in Patients With Dosing Interruption or Suboptimal Adherence. Am J Gastroenterol. 2021;116(9):1896-904.
61. Cunningham EB, Hajarizadeh B, Amin J, Litwin AH, Gane E, Cooper C, et al. Adherence to Once-daily and Twice-daily Direct-acting Antiviral Therapy for Hepatitis C Infection Among People With Recent Injection Drug Use or Current Opioid Agonist Therapy. Clin Infect Dis. 2020;71(7):e115-e24.
62. Cengiz M, Yılmaz G, Ozenirler S. The association between indirect bilirubin levels and liver fibrosis due to chronic hepatitis C virus infection. Pathology - Research and Practice. 2014;210(8):488-93.
63. Zhu Z, Wilson AT, Mathahs MM, Wen F, Brown KE, Luxon BA, et al. Heme oxygenase-1 suppresses hepatitis C virus replication and increases resistance of hepatocytes to oxidant injury. Hepatology. 2008;48(5):1430-9.
64. Zhu Z, Wilson AT, Luxon BA, Brown KE, Mathahs MM, Bandyopadhyay S, et al. Biliverdin inhibits hepatitis C virus nonstructural 3/4A protease activity: mechanism for the antiviral effects of heme oxygenase? Hepatology. 2010;52(6):1897-905.
65. Subhanova I, Muchova L, Lenicek M, Vreman HJ, Luksan O, Kubickova K, et al. Expression of Biliverdin Reductase A in Peripheral Blood Leukocytes Is Associated with Treatment Response in HCV-Infected Patients. PLOS ONE. 2013;8(3):e57555.
66. Neumann AU, Lam NP, Dahari H, Gretch DR, Wiley TE, Layden TJ, et al. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. Science. 1998;282(5386):103-7.
67. Perelson AS, Guedj J. Modelling hepatitis C therapy--predicting effects of treatment. Nat Rev Gastroenterol Hepatol. 2015;12(8):437-45.
68. Dahari H, Ribeiro RM, Perelson AS. Triphasic decline of hepatitis C virus RNA during antiviral therapy. Hepatology. 2007;46(1):16-21.

69. Huang CF, Yeh ML, Huang CI, Liang PC, Lin YH, Hsieh MY, et al. Ribavirin facilitates early viral kinetics in chronic hepatitis C patients receiving daclatasvir/asunaprevir. J Gastroenterol Hepatol. 2020;35(1):151-6.
70. Etzion O, Dahari H, Yardeni D, Issachar A, Nevo-Shor A, Cohen-Naftaly M, et al. Response guided therapy for reducing duration of direct acting antivirals in chronic hepatitis C infected patients: a Pilot study. Sci Rep. 2020;10(1):17820.
71. Fourati S, Guedj J, Chevaliez S, Nguyen THT, Roudot-Thoraval F, Ruiz I, et al. Viral kinetics analysis and virological characterization of treatment failures in patients with chronic hepatitis C treated with sofosbuvir and an NS5A inhibitor. Aliment Pharmacol Ther. 2018;47(5):665-73.
72. McKnight AH, Townsend ML, Hashem MG, Naggie S, Park LP, Britt RB. Standard Versus Extended Duration Direct-Acting Antiviral Therapy in Hepatitis C Patients With Slow Response to Treatment. Annals of Pharmacotherapy. 2020;54(11):1057-64.
73. Gorstein E, Martinello M, Churkin A, Dasgupta S, Walsh K, Applegate TL, et al. Modeling based response guided therapy in subjects with recent hepatitis C infection. Antiviral Res. 2020;180:104862.
74. Akkarathamrongsin S, Payungporn S, Thong VD, Poovorawan K, Prapunwattana P, Poovorawan Y, et al. Early viral kinetics during hepatitis C virus genotype 6 treatment according to IL28B polymorphisms. World J Gastroenterol. 2014;20(30):10599-605.
75. Zeuzem S, Herrmann E, Lee JH, Fricke J, Neumann AU, Modi M, et al. Viral kinetics in patients with chronic hepatitis C treated with standard or peginterferon alpha2a. Gastroenterology. 2001;120(6):1438-47.
76. Herrmann E, Lee JH, Marinos G, Modi M, Zeuzem S. Effect of ribavirin on hepatitis C viral kinetics in patients treated with pegylated interferon. Hepatology. 2003;37(6):1351-8.
77. Dixit NM, Layden-Almer JE, Layden TJ, Perelson AS. Modelling how ribavirin improves interferon response rates in hepatitis C virus infection. Nature. 2004;432(7019):922-4.
78. Jones CR, Flower BF, Barber E, Simmons B, Cooke GS. Treatment optimisation for hepatitis C in the era of combination direct-acting antiviral therapy: a systematic review and meta-analysis. Wellcome Open Res. 2019;4:132.
79. Spaan M, van Oord G, Kreefft K, Hou J, Hansen BE, Janssen HL, et al. Immunological Analysis During Interferon-Free Therapy for Chronic Hepatitis C Virus Infection Reveals Modulation of the Natural Killer Cell Compartment. J Infect Dis. 2016;213(2):216-23.
80. EuroQol Group. EQ-5D instrument [17/08/2022]. Available from: https://euroqol.org/eq-5d-instruments/.
81. Hays RD, Sherbourne CD, Mazel R. User's Manual for the Medical Outcomes Study (MOS) Core Measures of Health-Related Quality of Life. Santa Monica, CA: RAND Corporation; 1995.
82. QualityMetric. The SF-12v2 Health Survey [18/08/2022]. Available from: https://www.qualitymetric.com/health-surveys-old/the-sf-12v2-health-survey/.
83. Ware J, Kosinski M, Keller S. SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales. 1998.
84. Spiegel BM, Younossi ZM, Hays RD, Revicki D, Robbins S, Kanwal F. Impact of hepatitis C on health related quality of life: a systematic review and quantitative assessment. Hepatology. 2005;41(4):790-800.
85. Björnsson E, Verbaan H, Oksanen A, Frydén A, Johansson J, Friberg S, et al. Health-related quality of life in patients with different stages of liver disease induced by hepatitis C. Scand J Gastroenterol. 2009;44(7):878-87.
86. Yeung MW, Young J, Moodie E, Rollet-Kurhajec KC, Schwartzman K, Greenaway C, et al. Changes in quality of life, healthcare use, and substance use in HIV/hepatitis C coinfected patients after hepatitis C therapy: a prospective cohort study. HIV Clin Trials. 2015;16(3):100-10.

87. Ragusa R, Bertino G, Bruno A, Frazzetto E, Cicciu F, Giorgianni G, et al. Evaluation of health status in patients with hepatitis c treated with and without interferon. Health Qual Life Outcomes. 2018;16(1):17.
88. Vera-Llonch M, Martin M, Aggarwal J, Donepudi M, Bayliss M, Goss T, et al. Health-related quality of life in genotype 1 treatment-naïve chronic hepatitis C patients receiving telaprevir combination treatment in the ADVANCE study. Aliment Pharmacol Ther. 2013;38(2):124-33.
89. Stepanova M, Nader F, Cure S, Bourhis F, Hunt S, Younossi ZM. Patients' preferences and health utility assessment with SF-6D and EQ-5D in patients with chronic hepatitis C treated with sofosbuvir regimens. Aliment Pharmacol Ther. 2014;40(6):676-85.
90. Kracht PAM, Lieveld FI, Amelung LM, Verstraete CJR, Mauser-Bunschoten EP, de Bruijne J, et al. The Impact of Hepatitis C Virus Direct-Acting Antivirals on Patient-Reported Outcomes: A Dutch Prospective Cohort Study. Infectious Diseases and Therapy. 2018;7(3):373-85.
91. Younossi ZM, Stepanova M, Sulkowski M, Wyles D, Kottilil S, Hunt S. Patient-reported outcomes in patients co-infected with hepatitis C virus and human immunodeficiency virus treated with sofosbuvir and velpatasvir: The ASTRAL-5 study. Liver International. 2017;37(12):1796-804.
92. Fagundes RN, Ferreira L, Pace FHL. Health-related quality of life and fatigue in patients with chronic hepatitis C with therapy with direct-acting antivirals agents interferon-free. PLoS One. 2020;15(8):e0237005.
93. Santonicola A, Bilancio G, Zingone F, Donnarumma L, Caputo C, Ciacci C. The Psychosocial Burden of HCV Infection and the Impact of Antiviral Therapy on the Quality of Life in Liver and Kidney Transplant Recipients: A Pilot Study. Gastroenterology Research and Practice. 2020;2020.
94. Karimi-Sari H, Hosseini MA, Nikjoo N, Bagheri Baghdasht MS, Alavian SM. Patient-reported outcomes of sleep, mood and quality of life after treatment of chronic hepatitis C infection using direct-acting antiviral agents. Clinical Microbiology and Infection. 2020;26(8):1093.e5-.e8.
95. Nardelli S, Riggio O, Rosati D, Gioia S, Farcomeni A, Ridola L. Hepatitis C virus eradication with directly acting antivirals improves health-related quality of life and psychological symptoms. World J Gastroenterol. 2019;25(48):6928-38.
96. Yi CH, Bair MJ, Wang JH, Wong MW, Liu TT, Lei WY, et al. Improvement of patient-reported outcomes in patients achieving sustained virologic response with direct-acting antivirals for hepatitis C virus infection. J Microbiol Immunol Infect. 2022;55(4):643-50.
97. Younossi ZM, Stepanova M, Younossi I, Papatheodoridis G, Janssen HLA, Agarwal K, et al. Patient-reported outcomes in patients chronic viral hepatitis without cirrhosis: The impact of hepatitis B and C viral replication. Liver International. 2019;39(10):1837-44.
98. Yunihastuti E, Amelia F, Hapsari AI, Wicaksana B, Natali V, Widhani A, et al. Impact of sofosbuvir and daclastavir on health-related quality of life in patients co-infected with hepatitis C and human immunodeficiency virus. Health Qual Life Outcomes. 2021;19(1):154.
99. Cornberg M, Stoehr A, Naumann U, Teuber G, Klinker H, Lutz T, et al. Real-World Safety, Effectiveness, and Patient-Reported Outcomes in Patients with Chronic Hepatitis C Virus Infection Treated with Glecaprevir/Pibrentasvir: Updated Data from the German Hepatitis C-Registry (DHC-R). Viruses. 2022;14(7).
100. Ohlendorf V, Schäfer A, Christensen S, Heyne R, Naumann U, Link R, et al. Only partial improvement in health-related quality of life after treatment of chronic hepatitis C virus infection with direct acting antivirals in a real-world setting-results from the German Hepatitis C-Registry (DHC-R). J Viral Hepat. 2021;28(8):1206-18.
101. Goñi Esarte S, Juanbeltz R, Martínez-Baz I, Castilla J, San Miguel R, Herrero JI, et al. Long-term changes on health-related quality of life in patients with chronic hepatitis C after viral clearance with direct-acting antiviral agents. Rev Esp Enferm Dig. 2019;111(6):445-52.

102. Gormley MA, Akiyama MJ, Rennert L, Howard KA, Norton BL, Pericot-Valverde I, et al. Changes in Health-related Quality of Life for Hepatitis C Virus-Infected People Who Inject Drugs While on Opioid Agonist Treatment Following Sustained Virologic Response. Clin Infect Dis. 2022;74(9):1586-93.
103. Juanbeltz R, Martínez-Baz I, San Miguel R, Goñi-Esarte S, Cabasés JM, Castilla J. Impact of successful treatment with direct-acting antiviral agents on health-related quality of life in chronic hepatitis C patients. PLoS One. 2018;13(10):e0205277.
104. Younossi ZM, Stepanova M, Charlton M, Curry MP, O'Leary JG, Brown RS, et al. Patient-reported outcomes with sofosbuvir and velpatasvir with or without ribavirin for hepatitis C virus-related decompensated cirrhosis: an exploratory analysis from the randomised, open-label ASTRAL-4 phase 3 trial. The Lancet Gastroenterology & Hepatology. 2016;1(2):122-32.
105. Younossi ZM, Stepanova M, Henry L, Han K-H, Ahn SH, Lim Y-S, et al. Sofosbuvir and ledipasvir are associated with high sustained virologic response and improvement of health-related quality of life in East Asian patients with hepatitis C virus infection. Journal of Viral Hepatitis. 2018;25(12):1429-37.
106. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5D-5L value set for England. Health Econ. 2018;27(1):7-22.
107. de Knegt RJ, Bezemer G, Van Gool AR, Drenth JP, Hansen BE, Droogleever Fortuyn HA, et al. Randomised clinical trial: escitalopram for the prevention of psychiatric adverse events during treatment with peginterferon-alfa-2a and ribavirin for chronic hepatitis C. Aliment Pharmacol Ther. 2011;34(11-12):1306-17.
108. Mahajan S, Avasthi A, Grover S, Chawla YK. Role of baseline depressive symptoms in the development of depressive episode in patients receiving antiviral therapy for hepatitis C infection. J Psychosom Res. 2014;77(2):109-15.
109. Cattie JE, Letendre SL, Woods SP, Barakat F, Perry W, Cherner M, et al. Persistent neurocognitive decline in a clinic sample of hepatitis C virus-infected persons receiving interferon and ribavirin treatment. Journal of NeuroVirology. 2014;20(6):561-70.
110. Younossi ZM, Guyatt G, Kiwi M, Boparai N, King D. Development of a disease specific questionnaire to measure health related quality of life in patients with chronic liver disease. Gut. 1999;45(2):295-300.
111. Reilly MC, Zbrozek AS, Dukes EM. The Validity and Reproducibility of a Work Productivity and Activity Impairment Instrument. PharmacoEconomics. 1993;4(5):353-65.
112. McCabe L, White IR, Cooke GS, Walker AS. Bayesian trial monitoring and power estimation in a complex Hepatitis C treatment trial (VIETNARMS). Trials. 2019;20(Suppl 1):PS2B.
113. McCabe L, White IR, Chau NVV, Barnes E, Pett SL, Cooke GS, et al. The design and statistical aspects of VIETNARMS: a strategic post-licensing trial of multiple oral direct-acting antiviral hepatitis C treatment strategies in Vietnam. Trials. 2020;21(1):413.
114. Montgomery AA, Peters TJ, Little P. Design, analysis and presentation of factorial randomised controlled trials. BMC Medical Research Methodology. 2003;3(1):26.
115. Parmar MK, Barthel FM, Sydes M, Langley R, Kaplan R, Eisenhauer E, et al. Speeding up the evaluation of new agents in cancer. J Natl Cancer Inst. 2008;100(17):1204-14.
116. Couper DJ, Hosking JD, Cisler RA, Gastfriend DR, Kivlahan DR. Factorial designs in clinical trials: options for combination treatment studies. J Stud Alcohol Suppl. 2005(15):24-32; discussion 6-7.
117. Freidlin B, Korn EL, Gray R, Martin A. Multi-arm clinical trials of new agents: some design considerations. Clin Cancer Res. 2008;14(14):4368-71.
118. Jaki T, Vasileiou D. Factorial versus multi-arm multi-stage designs for clinical trials with multiple treatments. Stat Med. 2017;36(4):563-80.
119. Royston P, Barthel FM, Parmar MK, Choodari-Oskooei B, Isham V. Designs for clinical trials with time-to-event outcomes based on stopping guidelines for lack of benefit. Trials. 2011;12:81.

120. Ghosh P, Liu L, Senchaudhuri P, Gao P, Mehta C. Design and monitoring of multi-arm multi-stage clinical trials. Biometrics. 2017;73(4):1289-99.
121. Berry DA. Bayesian clinical trials. Nat Rev Drug Discov. 2006;5(1):27-36.
122. Food and Drug Administration. Guidance for the use of Bayesian statistics in medical device clinical trials. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-use-bayesian-statistics-medical-device-clinical-trials 2010.
123. Wason JMS, Stecher L, Mander AP. Correcting for multiple-testing in multi-arm trials: is it necessary and is it done? Trials. 2014;15(1):364.
124. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Methods in health service research. An introduction to bayesian methods in health technology assessment. BMJ. 1999;319(7208):508-12.
125. Zohar S, Teramukai S, Zhou Y. Bayesian design and conduct of phase II single-arm clinical trials with binary outcomes: a tutorial. Contemp Clin Trials. 2008;29(4):608-16.
126. Mukherjee SD, Goffin JR, Taylor V, Anderson KK, Pond GR. Early stopping rules in oncology: considerations for clinicians. Eur J Cancer. 2011;47(16):2381-6.
127. Zhu H, Yu Q, Mercante DE. A Bayesian sequential design with binary outcome. Pharmaceutical statistics. 2017;16(3):192-200.
128. Pedroza C, Tyson JE, Das A, Laptook A, Bell EF, Shankaran S, et al. Advantages of Bayesian monitoring methods in deciding whether and when to stop a clinical trial: an example of a neonatal cooling trial. Trials. 2016;17(1):335-.
129. Ryan EG, Bruce J, Metcalfe AJ, Stallard N, Lamb SE, Viele K, et al. Using Bayesian adaptive designs to improve phase III trials: a respiratory care example. BMC medical research methodology. 2019;19(1):99-.
130. Ryan EG, Lamb SE, Williamson E, Gates S. Bayesian adaptive designs for multi-arm trials: an orthopaedic case study. Trials. 2020;21(1):83-.
131. Weiss HL, Urban DA, Grizzle WE, Cronin KA, Freedman LS, Kelloff GJ, et al. Bayesian monitoring of a phase 2 chemoprevention trial in high-risk cohorts for prostate cancer. Urology. 2001;57(4 Suppl 1):220-3.
132. Bartlett RH, Roloff DW, Cornell RG, Andrews AF, Dillon PW, Zwischenberger JB. Extracorporeal circulation in neonatal respiratory failure: a prospective randomized study. Pediatrics. 1985;76(4):479-87.
133. Murphy P, Moline M, Mayleben D, Rosenberg R, Zammit G, Pinner K, et al. Lemborexant, A Dual Orexin Receptor Antagonist (DORA) for the Treatment of Insomnia Disorder: Results From a Bayesian, Adaptive, Randomized, Double-Blind, Placebo-Controlled Study. Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine. 2017;13(11):1289-99.
134. Satlin A, Wang J, Logovinsky V, Berry S, Swanson C, Dhadda S, et al. Design of a Bayesian adaptive phase 2 proof-of-concept trial for BAN2401, a putative disease-modifying monoclonal antibody for the treatment of Alzheimer's disease. Alzheimer's & dementia (New York, N Y). 2016;2(1):1-12.
135. Berry SM, Petzold EA, Dull P, Thielman NM, Cunningham CK, Corey GR, et al. A response adaptive randomization platform trial for efficient evaluation of Ebola virus treatments: A model for pandemic response. Clin Trials. 2016;13(1):22-30.
136. Cellamare M, Ventz S, Baudin E, Mitnick CD, Trippa L. A Bayesian response-adaptive trial in tuberculosis: The endTB trial. Clin Trials. 2017;14(1):17-28.
137. Ahmed H, Abushouk AI, Attia A, Gadelkarim M, Gabr M, Negida A, et al. Safety and efficacy of sofosbuvir plus velpatasvir with or without ribavirin for chronic hepatitis C virus infection: A systematic review and meta-analysis. J Infect Public Health. 2018;11(2):156-64.
138. Sulkowski MS, Gardiner DF, Rodriguez-Torres M, Reddy KR, Hassanein T, Jacobson I, et al. Daclatasvir plus sofosbuvir for previously treated or untreated chronic HCV infection. N Engl J Med. 2014;370(3):211-21.

139. Food and Drug Administration. Non-Inferiority Clinical trials to Establish Effectiveness: Guidance for Industry. https://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf 2016.
140. Gamalo MA, Tiwari RC, LaVange LM. Bayesian approach to the design and analysis of non-inferiority trials for anti-infective products. Pharm Stat. 2014;13(1):25-40.
141. Lee JJ, Liu DD. A predictive probability design for phase II cancer clinical trials. Clin Trials. 2008;5(2):93-106.
142. Saville BR, Connor JT, Ayers GD, Alvarez J. The utility of Bayesian predictive probabilities for interim monitoring of clinical trials. Clin Trials. 2014;11(4):485-93.
143. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med. 2006;25(1):127-41.
144. Flower B, Nguyen Thi Ngoc P, McCabe L, Le Ngoc C, Vo Thi T, Thi Kim HV, et al. Rise in alanine aminotransferase after HCV treatment is a highly sensitive screen for treatment failure. Clin Liver Dis (Hoboken). 2023;21(5):138-42.
145. World Health Organization. Updated recommendations on HCV simplified service delivery and HCV diagnostics: policy brief. https://www.who.int/publications/i/item/9789240052697; 2022 2022.
146. Sy A, McCabe L, Hudson E, Ansari AM, Pedergnana V, Lin SK, et al. Utility of a buccal swab point-of-care test for the IFNL4 genotype in the era of direct acting antivirals for hepatitis C virus. PLoS One. 2023;18(1):e0280551.

8 Appendix

Table 8.1: Raw EQ-5D-5L scores at each time point

	Enrolment			End of treatment			EOT+12/Retreatment D0		
	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=97	Fixed duration N=98	p-value
Problems with mobility									
None	89 (92%)	86 (88%)	0.78	81 (86%)	82 (84%)	0.96	81 (85%)	79 (84%)	0.89
Slight	5 (5%)	9 (9%)		8 (9%)	10 (10%)		6 (6%)	8 (9%)	
Moderate	2 (2%)	2 (2%)		3 (3%)	3 (3%)		6 (6%)	4 (4%)	
Severe	1 (1%)	1 (1%)		2 (2%)	3 (3%)		2 (2%)	2 (2%)	
Unable/extreme	0	0		0	0		0	1 (1%)	
Problems with self-care									
None	91 (94%)	91 (93%)	1.00	88 (93%)	91 (94%)	0.42	87 (92%)	85 (90%)	0.96
Slight	4 (4%)	4 (4%)		5 (5%)	2 (2%)		5 (5%)	5 (5%)	
Moderate	2 (2%)	2 (2%)		2 (2%)	4 (4%)		3 (3%)	3 (3%)	
Severe	0	1 (1%)		0	0		0	1 (1%)	
Unable/extreme	0	0		0	0		0	0	

	Enrolment			End of treatment			EOT+12/Retreatment D0		
	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=97	Fixed duration N=98	p-value
Problems with usual activities									
None	84 (87%)	78 (80%)	0.24	72 (76%)	70 (72%)	0.97	73 (77%)	73 (78%)	0.52
Slight	7 (7%)	12 (12%)		14 (15%)	16 (16%)		14 (15%)	14 (15%)	
Moderate	4 (4%)	7 (7%)		6 (6%)	8 (8%)		4 (4%)	4 (4%)	
Severe	2 (2%)	0		2 (2%)	2 (2%)		4 (4%)	1 (1%)	
Unable/extreme	0	1 (1%)		1 (1%)	1 (1%)		0	2 (2%)	
Problems with pain or discomfort									
None	66 (68%)	69 (70%)	0.86	59 (62%)	71 (72%)	0.50	58 (61%)	66 (70%)	0.15
Slight	22 (23%)	19 (19%)		22 (23%)	16 (16%)		29 (31%)	17 (18%)	
Moderate	9 (9%)	8 (8%)		11 (12%)	9 (9%)		7 (7%)	7 (7%)	
Severe	0	1 (1%)		2 (2%)	2 (2%)		1 (1%)	4 (4%)	
Unable/extreme	0	1 (1%)		1 (1%)	0		0	0	

	Enrolment			End of treatment			EOT+12/Retreatment D0		
	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=97	Fixed duration N=98	p-value
Problems with anxiety or depression									
None	57 (59%)	59 (60%)	0.78	55 (58%)	61 (62%)	0.96	56 (59%)	58 (62%)	0.31
Slight	28 (29%)	27 (28%)		25 (26%)	21 (21%)		23 (24%)	19 (20%)	
Moderate	9 (9%)	10 (10%)		10 (11%)	11 (11%)		12 (13%)	8 (9%)	
Severe	1 (1%)	2 (2%)		3 (3%)	3 (3%)		1 (1%)	6 (6%)	
Unable/extreme	2 (2%)	0		2 (2%)	2 (2%)		3 (3%)	3 (3%)	
EQ-index value	0.94 (0.90, 1.00)	0.94 (0.89, 1.00)	0.17	0.92 (0.86, 1.00)	0.95 (0.86, 1.00)	0.66	0.94 (0.85, 1.00)	0.94 (0.86, 1.00)	0.63
EQ-visual analogue scale	85 (75, 95)	80 (75, 90)	0.98	80 (75, 90)	80 (75, 90)	0.37	85 (75, 90)	85 (75, 95)	0.99

Note: Data presented as n (%) or median (IQR).

Table 8.2: Raw MOS-Cog scores at each time point

	Enrolment			End of treatment			EOT+12/Retreatment D0		
	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=97	Fixed duration N=98	p-value
Difficulty reasoning or solving problems									
None of the time	56 (58%)	51 (52%)	0.09	50 (53%)	45 (46%)	0.30	56 (59%)	56 (61%)	0.76
A little bit of the time	25 (26%)	27 (28%)		19 (20%)	31 (32%)		19 (20%)	20 (22%)	
Some of the time	4 (4%)	11 (11%)		14 (15%)	11 (11%)		12 (13%)	6 (7%)	
A good bit of the time	8 (8%)	2 (2%)		7 (7%)	4 (4%)		2 (2%)	4 (4%)	
Most of the time	3 (3%)	6 (6%)		5 (5%)	5 (5%)		5 (5%)	5 (5%)	
All of the time	0	1 (1%)		0	2 (2%)		1 (1%)	1 (1%)	
Forgetting things									
None of the time	41 (42%)	35 (36%)	0.14	36 (38%)	25 (26%)	0.12	40 (43%)	32 (35%)	0.59
A little bit of the time	27 (28%)	36 (37%)		30 (32%)	43 (44%)		31 (33%)	31 (34%)	
Some of the time	16 (16%)	14 (14%)		17 (18%)	19 (19%)		15 (16%)	18 (20%)	
A good bit of the time	12 (12%)	6 (6%)		9 (9%)	4 (4%)		4 (4%)	7 (8%)	
Most of the time	1 (1%)	5 (5%)		2 (2%)	6 (6%)		3 (3%)	1 (1%)	
All of the time	0	2 (2%)		1 (1%)	1 (1%)		1 (1%)	3 (3%)	

	Enrolment			End of treatment			EOT+12/Retreatment D0		
	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=97	Fixed duration N=98	p-value
Trouble keeping attention on an activity									
None of the time	42 (44%)	37 (38%)	0.43	42 (44%)	36 (37%)	0.30	48 (52%)	42 (46%)	0.94
A little bit of the time	21 (22%)	30 (31%)		23 (24%)	39 (40%)		19 (20%)	22 (24%)	
Some of the time	22 (23%)	17 (17%)		20 (21%)	16 (16%)		14 (15%)	17 (19%)	
A good bit of the time	8 (8%)	6 (6%)		4 (4%)	3 (3%)		8 (9%)	7 (8%)	
Most of the time	2 (2%)	5 (5%)		5 (5%)	3 (3%)		2 (2%)	2 (2%)	
All of the time	1 (1%)	3 (3%)		1 (1%)	1 (1%)		2 (2%)	1 (1%)	
Difficulty doing things involving concentration or thinking									
None of the time	41 (43%)	44 (45%)	0.18	39 (41%)	37 (38%)	0.77	50 (53%)	41 (45%)	0.87
A little bit of the time	29 (30%)	28 (29%)		25 (27%)	35 (36%)		23 (24%)	26 (28%)	
Some of the time	17 (18%)	10 (10%)		18 (19%)	17 (17%)		11 (12%)	11 (12%)	
A good bit of the time	5 (5%)	11 (11%)		7 (7%)	4 (4%)		5 (5%)	7 (8%)	
Most of the time	4 (4%)	2 (2%)		3 (3%)	3 (3%)		4 (4%)	5 (5%)	
All of the time	0	3 (3%)		2 (2%)	2 (2%)		1 (1%)	2 (2%)	

	Enrolment			End of treatment			EOT+12/Retreatment D0		
	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=97	Fixed duration N=98	p-value
Confused and start several actions at a time									
None of the time	63 (66%)	57 (58%)	0.17	54 (57%)	48 (49%)	0.62	60 (63%)	53 (58%)	0.36
A little bit of the time	14 (15%)	17 (17%)		24 (25%)	31 (32%)		14 (15%)	22 (24%)	
Some of the time	11 (11%)	16 (16%)		6 (6%)	10 (10%)		13 (14%)	8 (9%)	
A good bit of the time	7 (7%)	2 (2%)		5 (5%)	4 (4%)		5 (5%)	4 (4%)	
Most of the time	1 (1%)	4 (4%)		6 (6%)	4 (4%)		1 (1%)	4 (4%)	
All of the time	0	2 (2%)		0	1 (1%)		2 (2%)	1 (1%)	
React slowly to things									
None of the time	64 (66%)	59 (60%)	0.45	53 (56%)	55 (56%)	0.87	58 (61%)	57 (62%)	0.99
A little bit of the time	18 (19%)	19 (19%)		25 (26%)	24 (24%)		23 (24%)	23 (25%)	
Some of the time	12 (12%)	13 (13%)		10 (11%)	10 (10%)		10 (11%)	8 (9%)	
A good bit of the time	3 (3%)	2 (2%)		5 (5%)	6 (6%)		2 (2%)	2 (2%)	
Most of the time	0	3 (3%)		2 (2%)	1 (1%)		2 (2%)	2 (2%)	
All of the time	0	2 (2%)		0	2 (2%)		0	0	
Summary score	90 (73, 100)	87 (70, 97)	0.48	87 (70, 100)	85 (73, 93)	0.49	90 (73, 100)	88 (73, 100)	0.54

Table 8.3: Raw SF-12 scores at each time point

	Enrolment			End of treatment			EOT+12/Retreatment D0		
	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=94	Fixed duration N=98	p-value	Variable duration N=95	Fixed duration N=94	p-value
General health									
Excellent	16 (16%)	8 (8%)	0.34	12 (13%)	15 (15%)	0.88	10 (11%)	15 (16%)	0.52
Very good	39 (40%)	44 (45%)		43 (46%)	38 (39%)		42 (44%)	41 (44%)	
Good	34 (35%)	32 (33%)		24 (26%)	28 (29%)		33 (35%)	25 (27%)	
Fair	6 (6%)	8 (8%)		10 (11%)	11 (11%)		6 (6%)	10 (11%)	
Poor	2 (2%)	5 (5%)		4 (4%)	6 (6%)		4 (4%)	3 (3%)	
Moderate activities									
Not limited	3 (3%)	4 (4%)	1.00	3 (3%)	4 (4%)	0.84	5 (5%)	5 (5%)	0.96
Limited a little	14 (14%)	15 (15%)		18 (19%)	16 (16%)		12 (13%)	13 (14%)	
Limited a lot	3 (3%)	4 (4%)		3 (3%)	4 (4%)		5 (5%)	5 (5%)	
Climbing several flights of stairs									
Not limited	4 (4%)	6 (6%)	0.56	8 (9%)	10 (10%)	0.43	6 (7%)	8 (9%)	0.30

	Enrolment			End of treatment			EOT+12/Retreatment D0		
	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=94	Fixed duration N=98	p-value	Variable duration N=95	Fixed duration N=94	p-value
Limited a little	14 (15%)	19 (20%)		19 (21%)	27 (28%)		15 (16%)	23 (25%)	
Limited a lot	4 (4%)	6 (6%)		8 (9%)	10 (10%)		6 (7%)	8 (9%)	
Accomplished less due to physical health									
All of the time	0	8 (8%)	0.02	3 (3%)	5 (5%)	0.50	1 (1%)	2 (2%)	0.62
Most of the time	8 (8%)	4 (4%)		10 (11%)	10 (10%)		7 (7%)	4 (4%)	
Some of the time	12 (13%)	11 (11%)		11 (12%)	18 (18%)		8 (8%)	13 (14%)	
A little of the time	19 (20%)	26 (27%)		23 (24%)	27 (28%)		28 (29%)	23 (24%)	
None of the time	57 (59%)	49 (50%)		47 (50%)	38 (39%)		51 (54%)	52 (55%)	
Limited in the kind of work or other activities due to physical health									
All of the time	3 (3%)	3 (3%)	0.37	3 (3%)	1 (1%)	0.22	1 (1%)	1 (1%)	0.41
Most of the time	3 (3%)	6 (6%)		4 (4%)	7 (7%)		4 (4%)	5 (5%)	
Some of the time	8 (8%)	16 (16%)		10 (11%)	21 (22%)		7 (8%)	13 (14%)	
A little of the time	19 (20%)	18 (18%)		24 (26%)	21 (22%)		22 (24%)	14 (15%)	
None of the time	63 (66%)	55 (56%)		50 (55%)	47 (48%)		59 (63%)	61 (65%)	

	Enrolment			End of treatment			EOT+12/Retreatment D0		
	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=94	Fixed duration N=98	p-value	Variable duration N=95	Fixed duration N=94	p-value
Accomplished less due to emotional problems									
All of the time	2 (2%)	4 (4%)	0.66	3 (3%)	4 (4%)	0.15	3 (3%)	3 (3%)	0.97
Most of the time	6 (6%)	8 (8%)		11 (12%)	8 (8%)		4 (4%)	6 (6%)	
Some of the time	13 (14%)	17 (17%)		17 (18%)	12 (12%)		15 (16%)	13 (14%)	
A little of the time	28 (29%)	21 (21%)		15 (16%)	30 (31%)		24 (25%)	22 (24%)	
None of the time	47 (49%)	48 (49%)		48 (51%)	44 (45%)		49 (52%)	49 (53%)	
Did work or other activities less due to emotional problems									
All of the time	2 (2%)	2 (2%)	0.61	3 (3%)	1 (1%)	0.13	1 (1%)	3 (3%)	0.30
Most of the time	4 (4%)	6 (6%)		7 (8%)	3 (3%)		2 (2%)	3 (3%)	
Some of the time	10 (11%)	14 (14%)		11 (12%)	14 (15%)		10 (11%)	14 (15%)	
A little of the time	21 (23%)	14 (14%)		12 (13%)	24 (25%)		30 (32%)	18 (20%)	
None of the time	56 (60%)	62 (63%)		59 (64%)	54 (56%)		51 (54%)	54 (59%)	
Pain interferes with work									
Not at all	63 (66%)	64 (65%)	0.32	52 (55%)	58 (59%)	0.61	59 (62%)	64 (69%)	0.47

	Enrolment			End of treatment			EOT+12/Retreatment D0		
	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=94	Fixed duration N=98	p-value	Variable duration N=95	Fixed duration N=94	p-value
A little bit	21 (22%)	15 (15%)		23 (24%)	23 (23%)		16 (17%)	13 (14%)	
Moderately	7 (7%)	6 (6%)		11 (12%)	6 (6%)		10 (11%)	5 (5%)	
Quite a bit	4 (4%)	11 (11%)		7 (7%)	8 (8%)		7 (7%)	10 (11%)	
Extremely	1 (1%)	2 (2%)		1 (1%)	3 (3%)		3 (3%)	1 (1%)	
How often felt calm and peaceful									
All of the time	9 (9%)	5 (5%)	0.21	10 (11%)	7 (7%)	0.93	10 (11%)	6 (6%)	0.80
Most of the time	42 (44%)	48 (50%)		39 (41%)	39 (41%)		49 (52%)	46 (49%)	
Some of the time	30 (31%)	23 (24%)		28 (30%)	31 (32%)		20 (21%)	23 (24%)	
A little of the time	9 (9%)	17 (18%)		9 (10%)	11 (11%)		12 (13%)	13 (14%)	
None of the time	6 (6%)	3 (3%)		8 (9%)	8 (8%)		4 (4%)	6 (6%)	
How often have a lot of energy									
All of the time	8 (8%)	6 (6%)	0.92	10 (11%)	8 (8%)	0.35	7 (7%)	8 (9%)	0.09
Most of the time	38 (40%)	35 (36%)		31 (33%)	25 (26%)		41 (43%)	40 (43%)	
Some of the time	30 (31%)	36 (37%)		31 (33%)	27 (28%)		31 (33%)	17 (18%)	

	Enrolment			End of treatment			EOT+12/Retreatment D0		
	Variable duration N=97	Fixed duration N=98	p-value	Variable duration N=94	Fixed duration N=98	p-value	Variable duration N=95	Fixed duration N=94	p-value
A little of the time	12 (13%)	12 (12%)		14 (15%)	23 (24%)		11 (12%)	19 (20%)	
None of the time	8 (8%)	8 (8%)		8 (9%)	13 (14%)		5 (5%)	10 (11%)	
How often felt downhearted or low									
All of the time	3 (3%)	5 (5%)	0.95	2 (2%)	3 (3%)	0.48	4 (4%)	3 (3%)	0.95
Most of the time	4 (4%)	5 (5%)		14 (15%)	9 (9%)		6 (6%)	7 (7%)	
Some of the time	26 (27%)	27 (28%)		17 (18%)	26 (27%)		23 (24%)	23 (24%)	
A little of the time	38 (40%)	35 (36%)		32 (34%)	28 (29%)		37 (39%)	33 (35%)	
None of the time	25 (26%)	25 (26%)		29 (31%)	30 (31%)		24 (26%)	28 (30%)	
Physical health or emotional problems interfered with social activities									
All of the time	1 (1%)	4 (4%)	0.14	6 (6%)	4 (4%)	0.37	4 (4%)	4 (4%)	0.92
Most of the time	8 (8%)	11 (11%)		6 (6%)	9 (9%)		7 (7%)	4 (4%)	
Some of the time	11 (11%)	21 (22%)		14 (15%)	23 (24%)		18 (19%)	18 (19%)	
A little of the time	28 (29%)	22 (23%)		26 (28%)	28 (29%)		22 (23%)	21 (22%)	
None of the time	48 (50%)	39 (40%)		42 (45%)	33 (34%)		44 (46%)	47 (50%)	

	Enrolment			End of treatment			EOT+12/Retreatment D0		
	Variable duration N=97	**Fixed duration N=98**	**p-value**	**Variable duration N=94**	**Fixed duration N=98**	**p-value**	**Variable duration N=95**	**Fixed duration N=94**	**p-value**
Physical component summary score	54 (46, 57)	52 (41, 56)	0.15	50 (44, 56)	49 (42, 55)	0.34	53 (45, 55)	54 (45, 57)	0.48
Mental component summary score	46 (37, 53)	45 (35, 53)	0.58	46 (36, 54)	43 (35, 52)	0.24	45 (36, 54)	46 (36, 53)	0.76

Note: Data presented as n (%) or median (IQR).