Supplementary Material

Fibrosis regression in HIV-Hepatitis C virus (HCV) co-infection before and after sustained virologic response: Is transient elastography the best non-invasive marker for long-term monitoring?

Nadine Kronfli[1], Jim Young[1,2,3], Shouao Wang[1], Joseph Cox[1,3], Sharon Walmsley[4,5], Mark Hull[6], Curtis Cooper[7], Valerie Martel-Laferriere[8], Alexander Wong[9], Neora Pick[10] and Marina B. Klein[1,5] for the Canadian Co-infection Cohort Study Investigators

Authors Kronfli and Young contributed equally to the manuscript

[1]Department of Medicine, Division of Infectious Diseases and Chronic Viral Illness Service, McGill University, Montreal, Quebec, Canada

[2]Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, Basel, Switzerland

[3]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Canada

[4]University Health Network, University of Toronto, Toronto, Canada

[5]CIHR Canadian HIV Trials Network, Vancouver, Canada

[6]BC Centre of Excellence, St. Paul's Hospital, Vancouver, Canada

[7]Ottawa Hospital Research Institute, Ottawa, Canada

[8]Departement de Microbiologie et Infectiologie, Centre Hospitalier de l'Université de Montréal, Montreal, Canada

[9]Regina Qu'Appelle Health Region, Regina, Canada

[10]Department of Medicine, Division of Infectious Diseases, University of British Columbia, Vancouver, Canada

Corresponding author:

Marina B. Klein

Department of Medicine, Division of Infectious Diseases

McGill University Health Centre

1001 Decarie Boulevard, D02.4110

Montreal, Quebec H4A 3J1, Canada

Email: marina.klein@mcgill.ca

Telephone: 1-514-843-2090

Fax: 1-514-843-2092

Introduction

We analyzed measurement sequences from each patient with a generalized additive mixed model [1]. For this model, we assumed measurements were distributed gamma, with the mean estimated from model parameters through a natural log link function; we included fixed effects for pre-specified covariates and two smoothing terms. The first smoothing term was a random intercept to accommodate repeated measurements from the same patient; the second smoothing term was a spline function representing the change in the mean measurement over time, both before, during and after treatment. The date treatment began was designated time zero.

When fitting generalized additive models, subjective decisions need to be made about how to construct a smooth response curve from the underlying data. These decisions may hide or exaggerate elements of a response curve. Here we describe the process we followed to arrive at a final response curve.

This supplementary material also includes additional results. Figure S1 shows the process of patient selection for three main analyses: aspartate aminotransferase to platelet ratio index (APRI) in the interferon era, APRI in the direct acting antiviral (DAA) era and transient elastography (TE) in the DAA era. Patient selection for the analysis of Fibrosis-4 index (FIB-4) in the DAA era is not shown because the process was almost identical to that shown for APRI. Table S1 has the covariate estimates for all four main analyses. Table S2 has slope estimates for sensitivity analysis where alternative sets of covariates were used when modeling APRI in the DAA era.

Generalized additive modeling strategy

Generalized additive mixed models were fit in R (version 3.5.2) using the mgcv package (version 1.8-28) [2]. A gamma distributed generalized linear model with a log link function is appropriate for data where the variance increases with the mean and where the systematic part of the model acts multiplicatively on the response [3]. Another option is a normal model of a log transformed response, but the gamma model avoids the need to transform the response so that prediction is on the original scale of measurement.

We first smoothed the response over time using a natural cubic spline. This form of smoothing

is relatively well-known and easily understood. It consists of combining sections of a set of simple polynomial functions; where these sections join are known as knots [4]. The default when using a cubic spline in R package mgcv is to construct the response curve from cubic sections using ten knots evenly spread through the data – that is, there are an equal number of data points between each knot.

Our modeling strategy was to (1) fit the default response curve and (2) fit a response curve where we chose where knots were placed, concentrating knots in regions of the response where we expected more rapid change – just prior to, during and immediately after treatment. When placing knots, we used the same approach with each main analysis: a knot for each year with appreciable data before treatment started or after treatment finished over the range of the data; and the same intensive pattern of knots throughout the treatment period.

Having fit these two curves, we then considered whether any further modification was necessary to create a plausible response curve. We found that most curves looked more plausible with adaptive smoothing. With conventional spline smoothing, a single (constant) smoothing parameter controls the tradeoff between fit to the data and smoothness. With adaptive smoothing, this smoothing parameter is allowed to vary with the response. We think that the concentration of data collected around the treatment period can have an undue influence on the smooth outside that period. We found that an adaptive p-spline – the default for adaptive smoothing in mgcv – gave a smoother response curves than an adaptive cubic spline [5]. We constructed the p-spline [6] from our usual set of knots, plus knots at the extremes of the data and three knots beyond each extreme; the additional knots needed to create a second order p-spline with a second order difference penalty, the default p-spline in mgcv [7].

Our strategy reflects a Bayesian smoothing perspective: that smooth mean responses are considered more likely than variable mean responses. Indeed the credible intervals shown in plots and in tables reflect a prior than makes this assumption explicit and gives equal probability to all models of equal smoothness [8]. Here we show response curves for each of the three steps in our modeling strategy over the full range of the available data. The response curves in the main paper show the final curve for each outcome over just the four years before and after treatment began.

APRI in the interferon era

With cubic spline smoothing, both the default response curve and the curve based on our knot placements were similar (Figure S2) – reference lines show where the knots were placed for each curve. By comparison, the response curve based on an adaptive p-spline had its peak closer to when treatment started and was more plausible for this reason.

APRI in the DAA era

With cubic spline smoothing, the default response curve did not capture the rapid change when treatment first began (Figure S3). In this situation, knots concentrated in regions where we expected rapid change better captured the likely response to treatment. With adaptive p-spline smoothing, the response curve showed slightly less variation in the period leading up to treatment and we preferred the third curve for that reason.

FIB-4 in the DAA era

As with APRI, adaptive p-spline smoothing gave the most plausible response curve for FIB-4 in the DAA era (Figure S4). This response curve had less variation prior to treatment and a peak response closer to the start of treatment than alternative curves based on cubic spline smoothing.

TE in the DAA era

Again with cubic spline smoothing, both the default response curve and the curve based on our knot placements were similar (Figure S5) – most likely due to the absence of rapid change during treatment. By comparison, the response curve based on an adaptive p-spline had its peak closer to when treatment started and was more plausible for this reason.

APRI in the DAA era in fibrosis subgroups

When modeling subgroups, our starting point was the model fitted to all APRI data in the DAA era by adaptive p-spline smoothing . With this model, however, the response curve seemed somewhat segmented between knots (Figure S6, A). We assumed this was due to too many knots given fewer data. We therefore fit a second model using cubic spline smoothing, with knots every two years before and

after treatment, although with the same intensive pattern of knots throughout the treatment period. But this second model offered no real improvement (Figure S6, B). We therefore used the adaptive p-spline model for both patients with and without fibrosis (Figure S6, C) to be consistent with other analyses.

FIB-4 in the DAA era in fibrosis subgroups

As with APRI, both an adaptive p-spline and a simpler cubic spline with fewer knots gave similar response curves for patients with fibrosis when starting treatment (Figure S7, A and B). We therefore used the adaptive p-spline model for both patients with and without fibrosis (Figure S7, C) to be consistent with other analyses.

TE in the DAA era in fibrosis subgroups

When modeling subgroups, our starting point was the model fitted to all TE data in the DAA era by adaptive p-spline smoothing. This model gave a plausible response curve for patients with fibrosis when starting treatment (Figure S8, A). An alternative model using cubic spline smoothing was less plausible, with an earlier peak relative to when treatment started (Figure S8, B). We therefore used the more complex adaptive p-spline model for both patients with and without fibrosis (Figure S8, C).

Additional APRI data in the DAA era

When modeling APRI with additional data collected during treatment monitoring (Figure S9), we following the same strategy as before for this outcome. Again with adaptive p-spline smoothing (Figure S9, C), the response curve shows slightly less variation in the period leading up to treatment and we prefer the third curve for that reason. Note that the two slope estimates (before and after treatment) are no more precise when found from treatment monitoring data and cohort follow-up data combined than the two estimates found from cohort follow-up data alone (Table 2).

References

1. Wood SM. Mixed methods and GAMMs (Chapter 6). Generalized additive models: An introduction with R. Boca Raton, Florida: Chapman & Hall/CRC Press; 2006 pp. 277-325.
2. Wood SM. GAMs in practice: mgvc (Chapter 5). Generalized additive models: An introduction with R. Boca Raton, Florida: Chapman & Hall/CRC Press; 2006 pp. 221-260.
3. McCullagh P, Nelder JA. Models for data with constant coefficient of variation (Chapter 8). Generalized linear models. London: Chapman and Hall; 1989 pp. 285-313.
4. Wood SM. Univariate smooth functions (Chapter 3.2). Generalized additive models: An introduction with R. Boca Raton, Florida: Chapman & Hall/CRC Press; 2006 pp.122-133.
5. Wood SM. Adaptive smooths in GAMs. R Documentation. Smooth.construct.ps.smooth.spec {mgcv}, Package *mgcv* version 1.8-28. [cited 2019 Aug. 30]. Available from: https://stat.ethz.ch/R-manual/R-patched/library/mgcv/html/smooth.construct.ad.smooth.spec.html
6. Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. Statistical Science, 1996; 11(2):89-121.
7. Wood SM. P-splines in GAMs. R Documentation. Smooth.construct.ps.smooth.spec {mgcv}, Package *mgcv* version 1.8-28. [cited 2019 Aug. 30]. Available from: https://stat.ethz.ch/R-manual/R-patched/library/mgcv/html/smooth.construct.ps.smooth.spec.html
8. Wood SM. Bayesian model and posterior distribution of the parameters, for an additive model (Chapter 4.8.1). Generalized additive models: An introduction with R. Boca Raton, Florida: Chapman & Hall/CRC Press; 2006 p.190.

Table S1. Covariate associations in generalized additive models: for aspartate aminotransferase to platelet ratio index (APRI) in the interferon era; and for APRI, Fibrosis-4 index (FIB-4) and transient elastography (TE) in the direct acting antiviral era.

Covariate risk ratio	Era:	Interferon	Direct acting antiviral		
(95% credible interval)	Marker:	APRI	APRI	FIB-4	TE
Female		0.83	1.12	1.16	1.16
		(0.58, 1.20)	(0.93, 1.35)	(1.01, 1.33)	(0.94, 1.43)
Body mass index, per 5 units					
< 25		0.97	1.01	0.98	0.99
		(0.82, 1.15)	(0.93, 1.09)	(0.93, 1.03)	(0.90, 1.09)
≥ 25		1.00	1.03	1.02	1.07
		(0.92, 1.09)	(0.98, 1.08)	(0.99, 1.05)	(0.99 1.15)
Age at HCV infection, per 10 years					
< 40		0.93	1.14	1.42	1.04
		(0.75, 1.17)	(0.99, 1.31)	(1.28, 1.57)	(0.89, 1.21)
≥ 40		1.23	0.95	1.27	0.93
		(0.64, 2.35)	(0.73, 1.25)	(1.04, 1.55)	(0.69, 1.26)
Duration of HCV infection, per 10 years		0.99	1.10	1.45	1.04
		(0.83, 1.18)	(0.99, 1.23)	(1.34, 1.57)	(0.92, 1.18)
Binge drinking [a]		1.10	1.12	1.06	1.05
		(0.98, 1.23)	(1.04, 1.19)	(1.02, 1.10)	(0.98, 1.13)
HIV RNA ≥ 50 copies/ml		1.00	1.10	1.09	0.96
		(0.90, 1.11)	(1.04, 1.15)	(1.06, 1.13)	(0.91, 1.02)

Abbreviations: HCV, hepatitis C virus.

[a] At least six or more drinks on at least one occasion each month in the last six months.

Table S2. Sensitivity analyses using models with different covariates: slope estimates for change in aspartate aminotransferase to platelet ratio index during the direct acting antiviral era.

Model	Slope before treatment (95% CrI) [a]	Slope after treatment (95% CrI) [b]
Main analysis	0.03 (-0.05, 0.12)	-0.03 (-0.06, 0.01)
Single additional covariate		
HCV Genotype 3	0.03 (-0.06, 0.12)	-0.03 (-0.06, 0.01)
CD4 < 200 cells/µl	0.05 (-0.08, 0.19)	-0.04 (-0.10, 0.01)
Type 2 diabetes	0.04 (-0.05, 0.12)	-0.02 (-0.06, 0.01)
Protease inhibitor-based ART	0.03 (-0.05, 0.12)	-0.03 (-0.06, 0.01)
Covariate substitutions		
Hazardous drinking	0.03 (-0.05, 0.12)	-0.02 (-0.06, 0.01)
Age when starting treatment	0.03 (-0.05, 0.13)	-0.03 (-0.07, 0.01)
Body mass index knot at 30	0.03 (-0.05, 0.12)	-0.03 (-0.06, 0.01)

Abbreviations: ART, antiretroviral therapy; CrI, credible interval; HCV, hepatitis C virus.

[a] Slope between 1.5 and 0.5 years before starting treatment.

[b] Slope between 0.5 and 1.5 years after treatment ends.

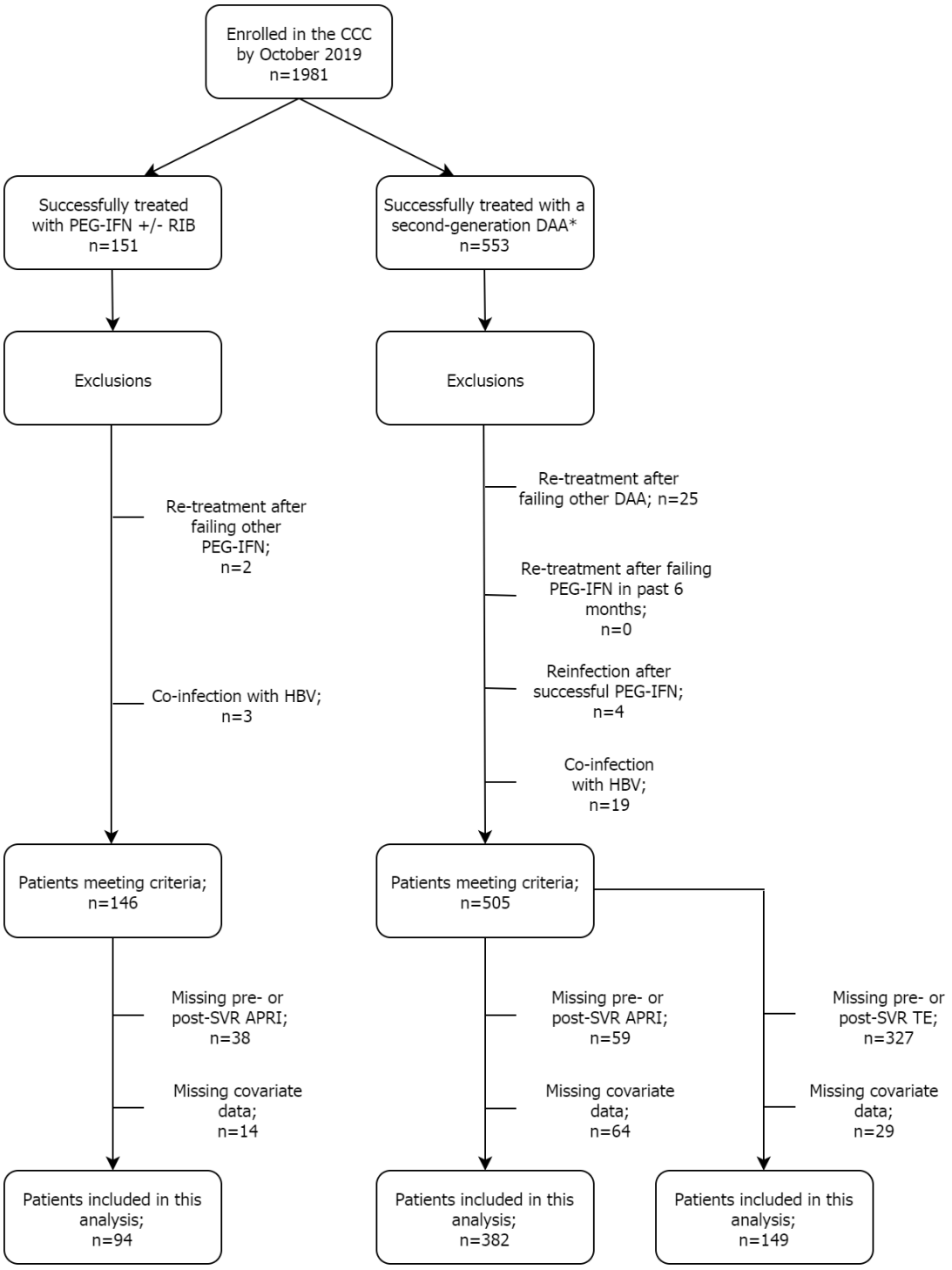

Figure S1. Patient selection for three main analyses.

Abbreviations: APRI, aspartate aminotransferase to platelet ratio index; CCC; Canadian Co-infection Cohort; DAA, direct acting antiviral; HBV, hepatitis B virus; PEG-IFN, pegylated interferon; RIB, ribavirin; SVR, sustained virologic response; TE, transient elastography.

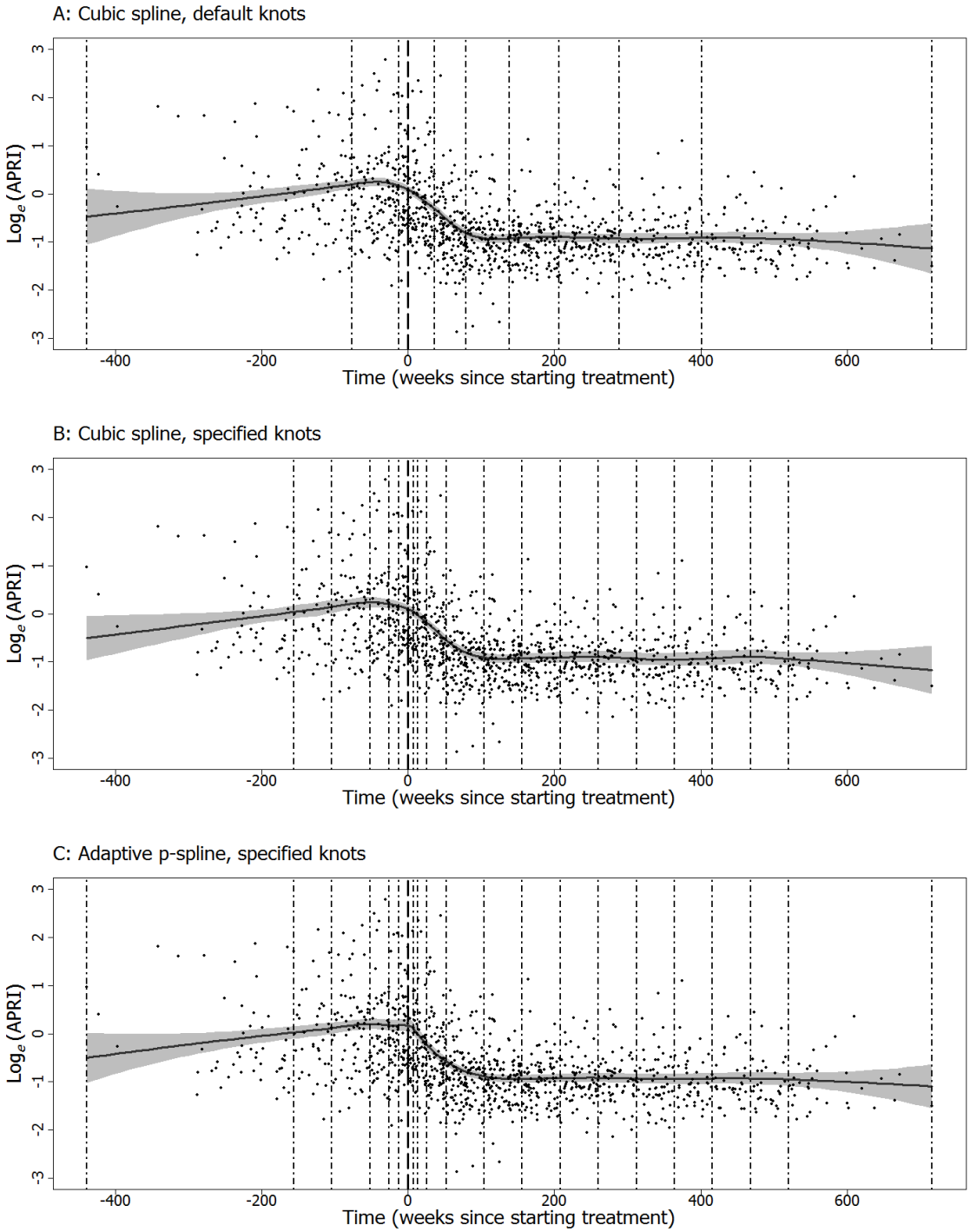

Figure S2. Response curves from generalized additive models for the aspartate aminotransferase to platelet ratio index (APRI) in the interferon era. Curves are shown for cubic spline smoothing with the default knots (A) and with a set of specified knots (B) and for adaptive p-spline smoothing with the same set of specified knots (C). In these plots, reference lines (• – •) show where knots were placed.

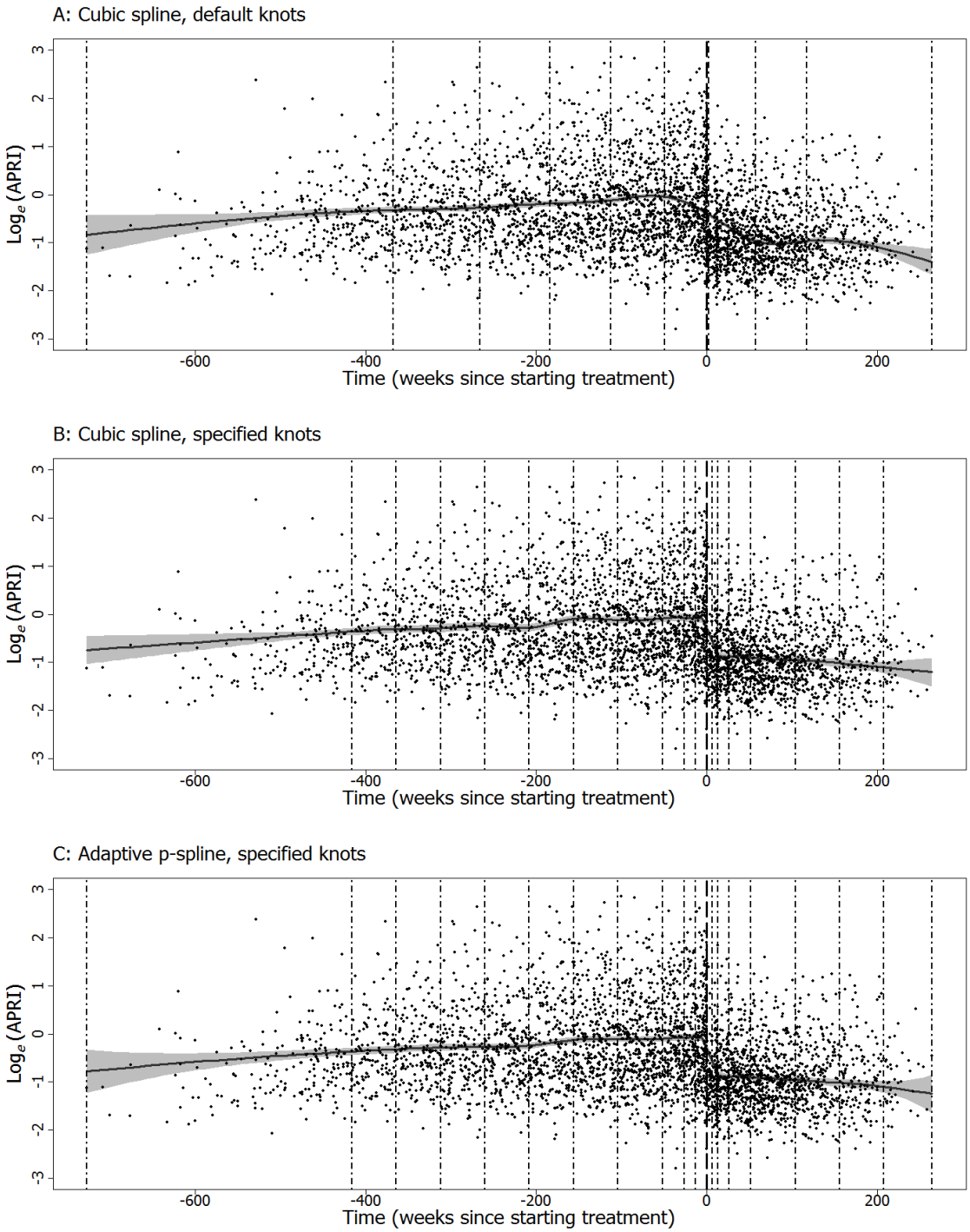

Figure S3. Response curves from generalized additive models for the aspartate aminotransferase to platelet ratio index (APRI) in the direct acting antiviral era. Curves are shown for cubic spline smoothing with the default knots (A) and with a set of specified knots (B) and for adaptive p-spline smoothing with the same set of specified knots (C). In these plots, reference lines (• – •) show where knots were placed.

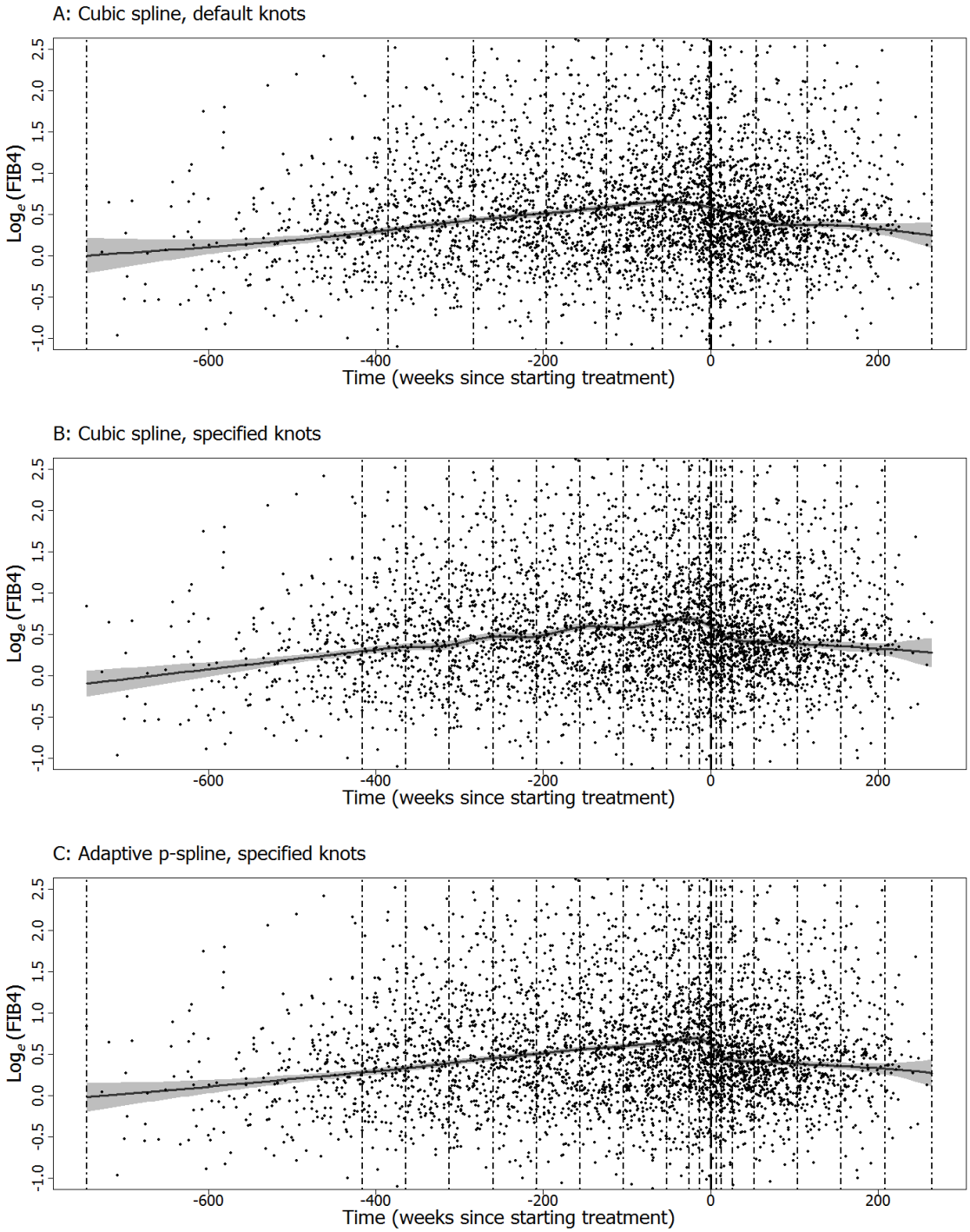

Figure S4. Response curves from generalized additive models for the Fibrosis-4 index (FIB-4) in the direct acting antiviral era. Curves are shown for cubic spline smoothing with the default knots (A) and with a set of specified knots (B) and for adaptive p-spline smoothing with the same set of specified knots (C). In these plots, reference lines (• – •) show where knots were placed.

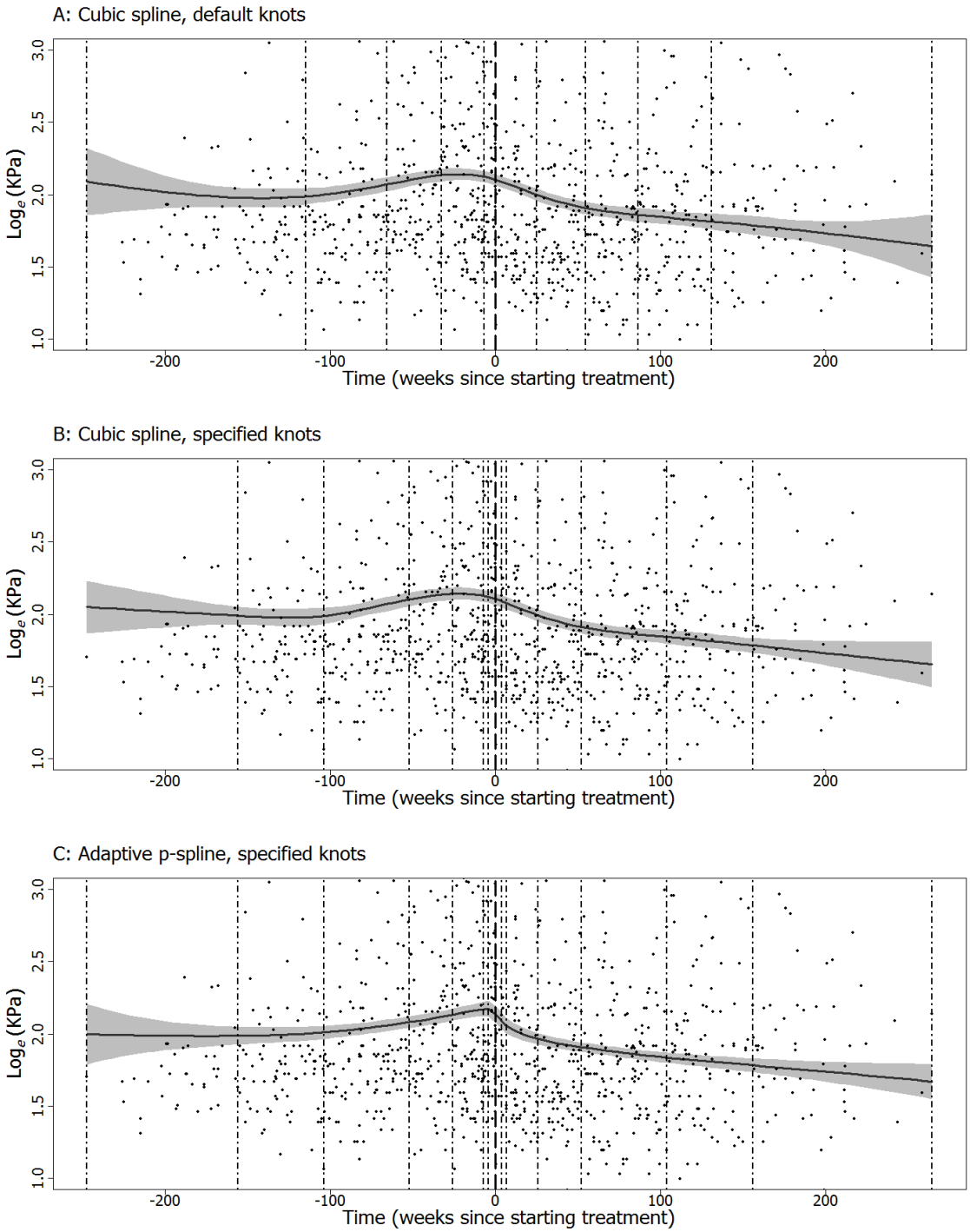

Figure S5. Response curves from generalized additive models for transient elastography in the direct acting antiviral era. Curves are shown for cubic spline smoothing with the default knots (A) and with a set of specified knots (B) and for adaptive p-spline smoothing with the same set of specified knots (C). In these plots, reference lines (• – •) show where knots were placed.

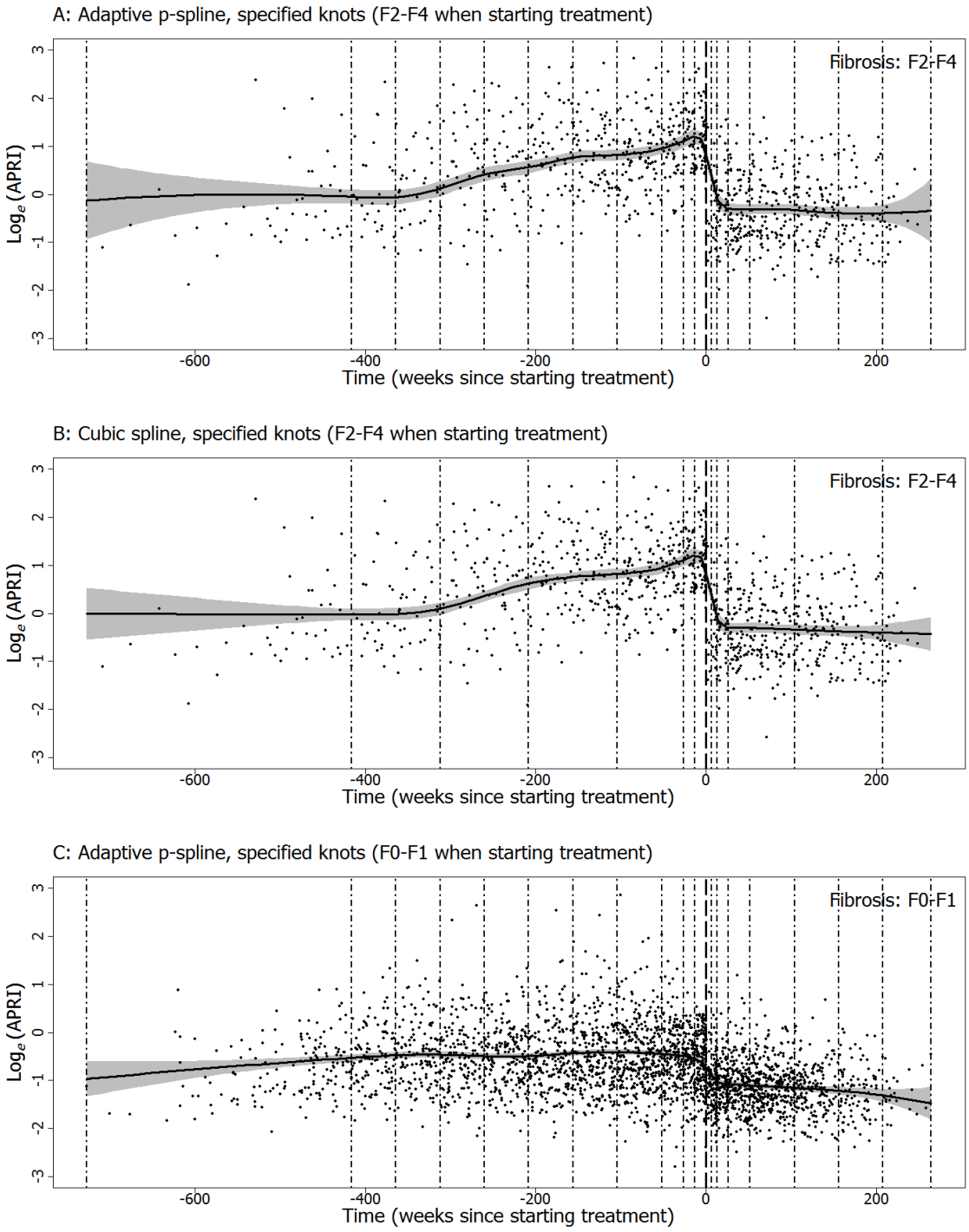

Figure S6. Response curves from generalized additive models for the aspartate aminotransferase to platelet ratio index (APRI) in the direct acting antiviral era. For patients with fibrosis (APRI ≥ 1.5) when starting treatment, curves are shown using adaptive p-spline smoothing with the usual set of specified knots (A) and then using cubic spline smoothing with a restricted set of knots (B). For patients without fibrosis (APRI < 1.5) when starting treatment, a curve is shown using adaptive p-spline smoothing with the usual set of specified knots (C). In these plots, reference lines (• – •) show where knots were placed.

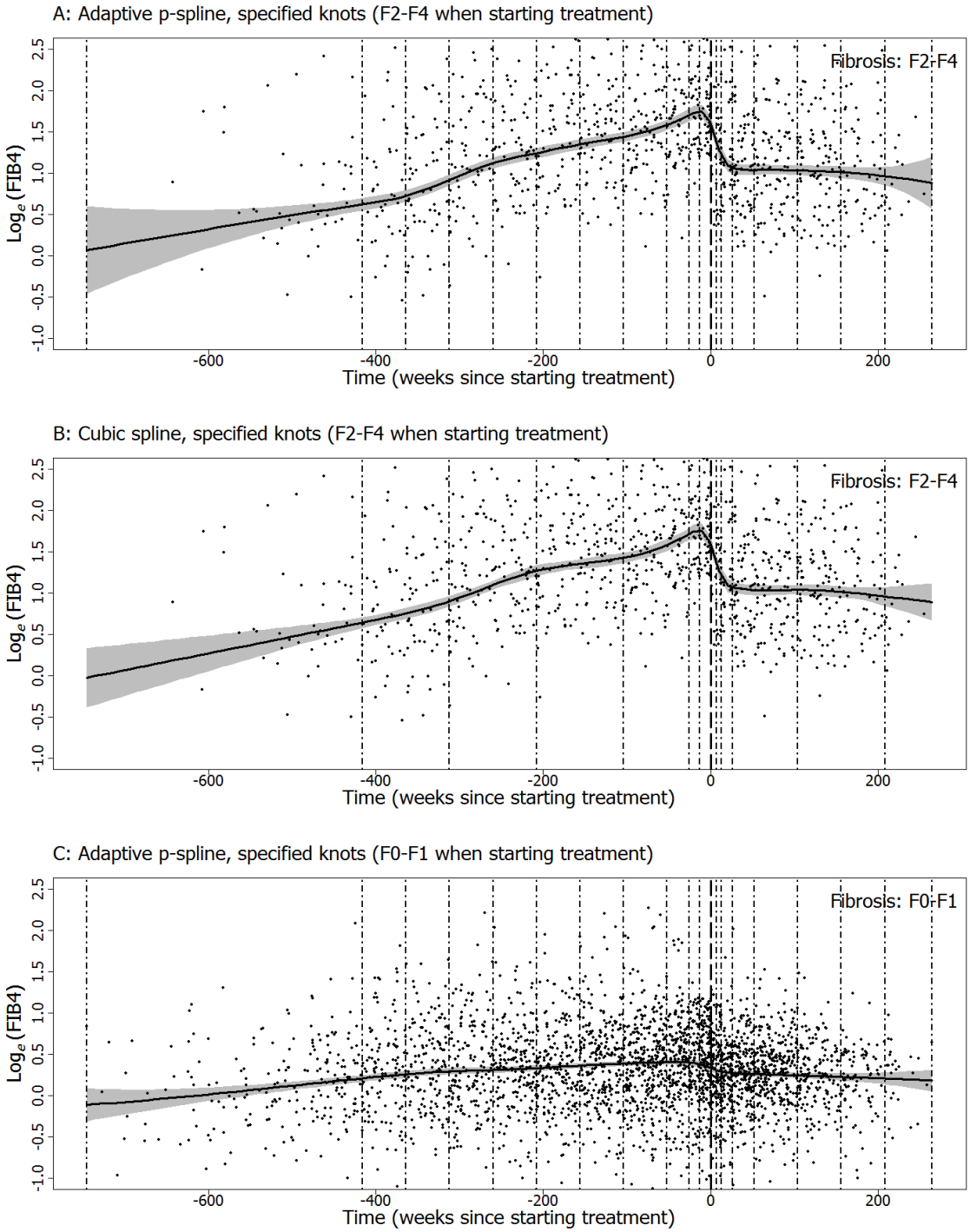

Figure S7. Response curves from generalized additive models for the Fibrosis-4 index (FIB-4) in the direct acting antiviral era. For patients with fibrosis (FIB-4 ≥ 3.25) when starting treatment, curves are shown using adaptive p-spline smoothing with the usual set of specified knots (A) and then using cubic spline smoothing with a restricted set of knots (B). For patients without fibrosis (FIB-4 < 3.25) when starting treatment, a curve is shown using adaptive p-spline smoothing with the usual set of specified knots (C). In these plots, reference lines (• – •) show where knots were placed.

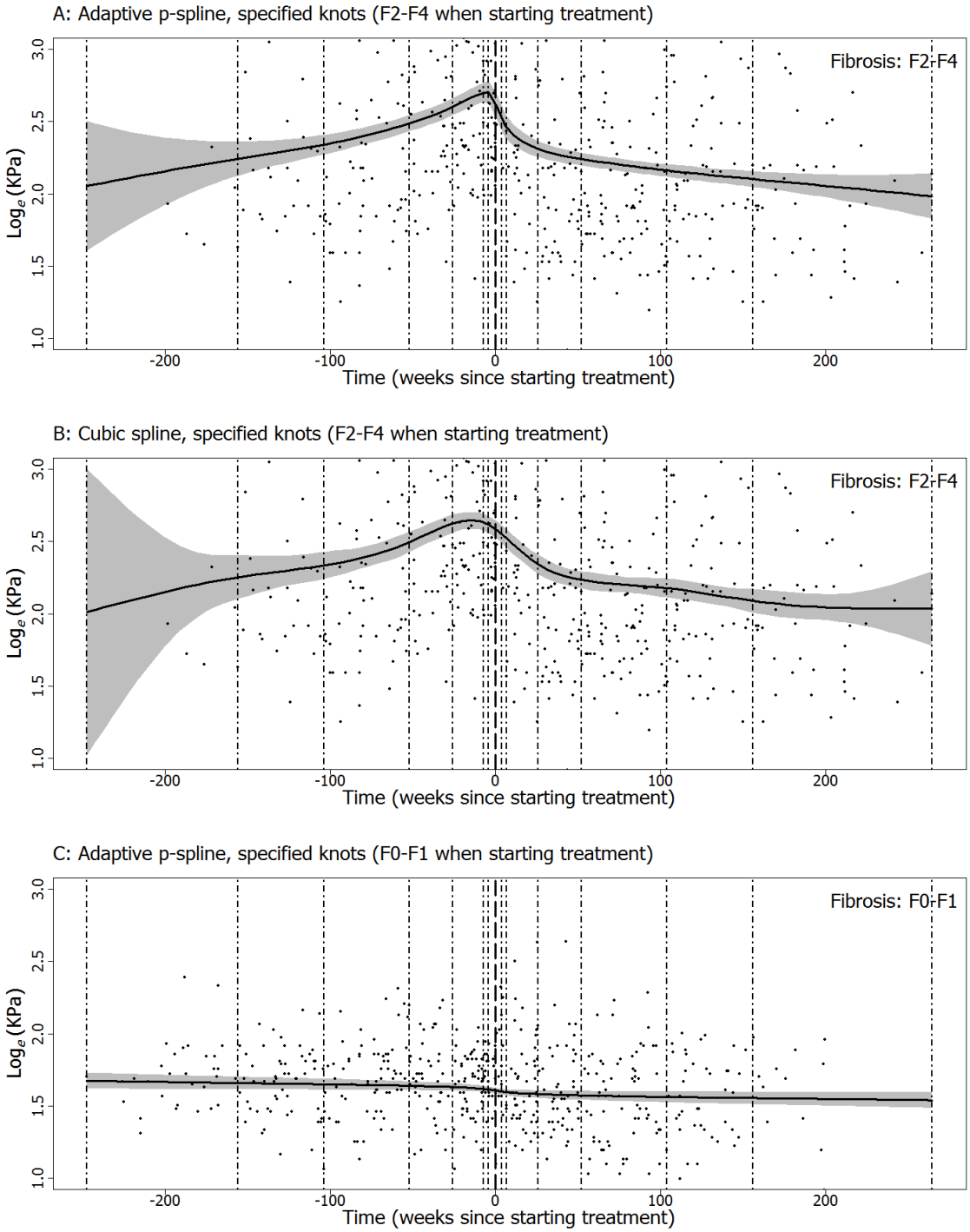

Figure S8. Response curves from generalized additive models for transient elastography (TE) in the direct acting antiviral era. For patients with fibrosis (TE ≥ 7.2) when starting treatment, curves are shown using adaptive p-spline smoothing with the usual set of specified knots (A) and then using cubic spline smoothing with the usual set of specified knots (B). For patients without fibrosis (TE < 7.2) when starting treatment, a curve is shown using adaptive p-spline smoothing with the usual set of specified knots (C). In these plots, reference lines (• – •) show where knots were placed.

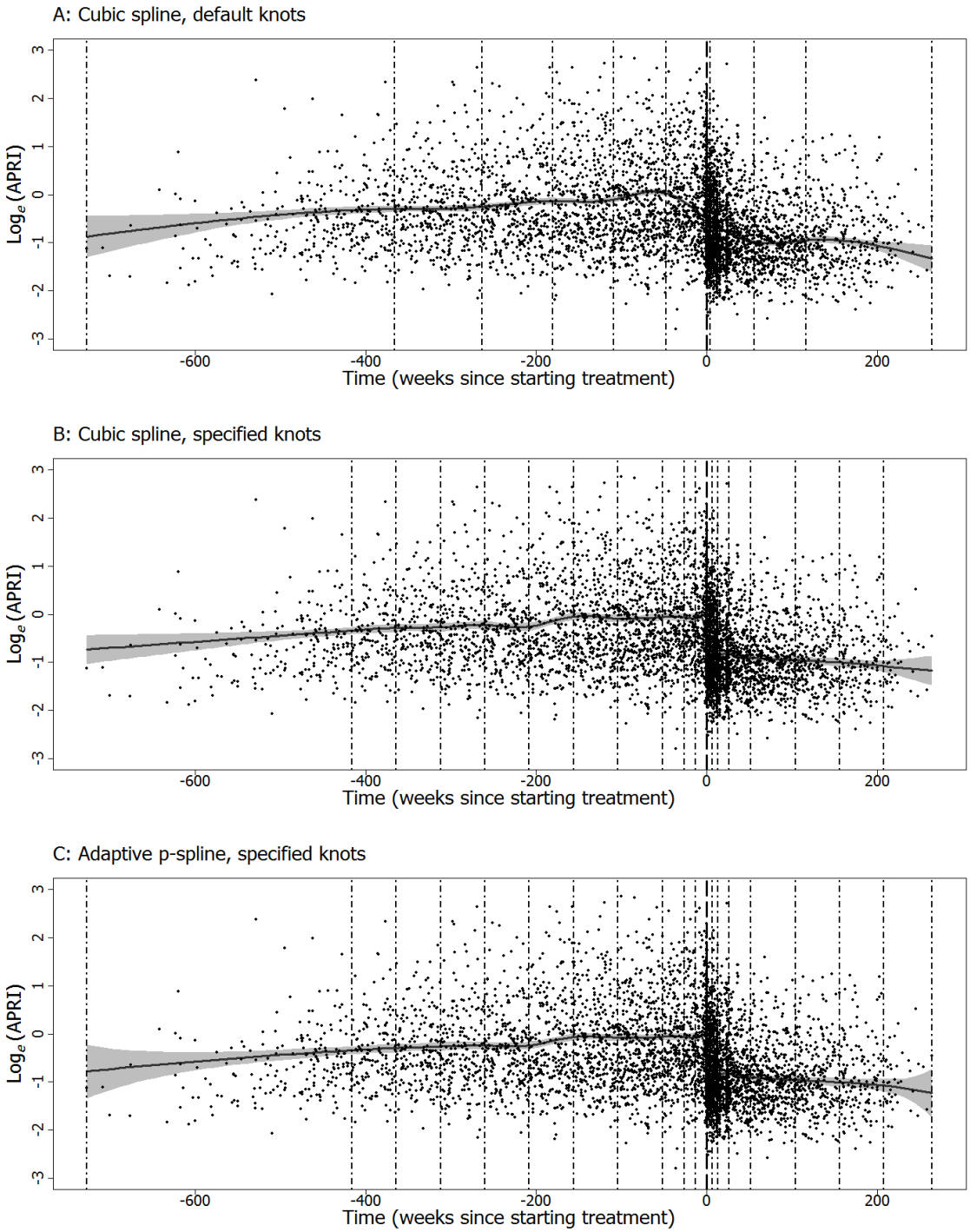

Figure S9. Response curves from generalized additive models for the aspartate aminotransferase to platelet ratio index (APRI) in the direct acting antiviral era. Additional data have been included collected during treatment monitoring. Curves are shown for cubic spline smoothing with the default knots (A) and with a set of specified knots (B) and for adaptive p-spline smoothing with the same set of specified knots (C). In these plots, reference lines (• – •) show where knots were placed.