A genotype-specific baseline score predicts post-treatment response to peginterferon alfa-2a in hepatitis B e antigen-negative chronic hepatitis B

Pietro Lampertico*a, Diethelm Messingerb, Markus Cornbergc, Maurizia Brunettof, Jörg Petersens, Patrick Kennedyt, Tarik Asselahg, Vivien Rotheb, Antonietta Caputoh, Georgios Bakalosi, Vedran Pavlovicj, George V. Papatheodoridisk

Fondazione IRCCS Ca’ Granda Ospedale Maggiore Policlinico, Università degli Studi di Milano, Milan, Italy; PROMETRIS GmbH, Mannheim, Germany; Hannover Medical School, Hannover, Germany; University Hospital of Pisa and Internal Medicine, Pisa, Italy; Asklepios Klinik St. Georg, University of Hamburg, Germany; Queen Mary University of London, London, UK; University of Paris Diderot, Hôpital Beaujon, Clichy, France; Roche S.p.A. Monza, Italy; F. Hoffmann-La Roche Ltd., Basel, Switzerland; Roche Products Ltd., Welwyn Garden City, UK; Medical School of National & Kapodistrian University of Athens, Laiko General Hospital, Athens, Greece

Background  Peginterferon alfa-2a induces durable responses in some hepatitis B e antigen-negative patients, but robust pretreatment predictors are not available to identify likely responders. In this study we aimed to develop genotype-specific baseline scoring systems to predict response.

Methods  Data from 323 hepatitis B e antigen-negative peginterferon alfa-2a recipients from three studies were analyzed. Scoring systems were developed using generalized additive models and multiple logistic regression analysis. Response was defined as hepatitis B virus DNA <2000 IU/mL alone (virological response) or in combination with alanine aminotransferase normalization (combined response) 48 weeks post-treatment.

Results  Points were assigned to genotype B/C patients for: age, alanine aminotransferase ratio, genotype B or C, and hepatitis B surface antigen level; and to genotype D patients for: age, hepatitis B surface antigen level and hepatitis B virus DNA level. Higher total scores (range 0-5 for B/C; 0-3 for D) indicated a higher likelihood of response. Among genotype B/C patients with scores of 0-1, 2 and ≥3, respectively, virological response rates were 16.7%, 25.8% and 70.2%, and combined response rates were 12.5%, 21.0% and 57.4%. Among genotype D patients with scores of 0-1, 2 and 3, respectively, virological response rates were 10.1%, 28.0% and 50.0%, and combined response rates were 7.8%, 28.0% and 33.3%.

Conclusion  Genotype-specific baseline scoring systems can identify hepatitis B e antigen-negative patients with low or high likelihood of achieving sustained responses to peginterferon alfa-2a.

Keywords  Predictors, treatment, HBsAg, HBeAg, virological response

Ann Gastroenterol 2018; 31 (6): 712-721

Abstract

Background  Peginterferon alfa-2a induces durable responses in some hepatitis B e antigen-negative patients, but robust pretreatment predictors are not available to identify likely responders. In this study we aimed to develop genotype-specific baseline scoring systems to predict response.

Methods  Data from 323 hepatitis B e antigen-negative peginterferon alfa-2a recipients from three studies were analyzed. Scoring systems were developed using generalized additive models and multiple logistic regression analysis. Response was defined as hepatitis B virus DNA <2000 IU/mL alone (virological response) or in combination with alanine aminotransferase normalization (combined response) 48 weeks post-treatment.

Results  Points were assigned to genotype B/C patients for: age, alanine aminotransferase ratio, genotype B or C, and hepatitis B surface antigen level; and to genotype D patients for: age, hepatitis B surface antigen level and hepatitis B virus DNA level. Higher total scores (range 0-5 for B/C; 0-3 for D) indicated a higher likelihood of response. Among genotype B/C patients with scores of 0-1, 2 and ≥3, respectively, virological response rates were 16.7%, 25.8% and 70.2%, and combined response rates were 12.5%, 21.0% and 57.4%. Among genotype D patients with scores of 0-1, 2 and 3, respectively, virological response rates were 10.1%, 28.0% and 50.0%, and combined response rates were 7.8%, 28.0% and 33.3%.

Conclusion  Genotype-specific baseline scoring systems can identify hepatitis B e antigen-negative patients with low or high likelihood of achieving sustained responses to peginterferon alfa-2a.

Keywords  Predictors, treatment, HBsAg, HBeAg, virological response

Ann Gastroenterol 2018; 31 (6): 712-721

Introduction

Chronic hepatitis B (CHB) is associated with serious complications, including liver failure, cirrhosis, and hepatocellular carcinoma, and is estimated to contribute to 650,000 deaths each year [1]. Hepatitis B e antigen (HBeAg)-negative CHB represents a late phase of infection characterized by persistent viral replication, progressive liver damage and cirrhosis [2,3]. The prevalence of HBeAg-negative CHB is increasing worldwide because of aging of populations [4,5]. Current guidelines recommend the use of nucleot(s)ide analogs (NAs) or conventional or pegylated interferon (PegIFN) alfa as first-line therapy for HBeAg-negative CHB [5,6]. NAs are highly effective at suppressing hepatitis B virus (HBV) DNA
replication and are well tolerated; however, relapse occurs frequently after withdrawal of therapy [5,6]. Hepatitis B surface antigen (HBsAg) levels decline very slowly during NA therapy [7]. Consequently, NA therapy must be continued for life to remain effective in the majority of patients [5], placing a burden both on the patients, who must adhere to the daily dosing regimen, and on healthcare systems, which must sustain potentially lifelong treatment programs.

In contrast to NAs, PegIFN alfa-2a induces virological responses that are durable after completion of treatment in approximately one-third of HBeAg-negative patients [8,9] and are sustained during long-term follow-up [10,11]. The decline in HBsAg is more rapid during PegIFN than NA therapy [12], and PegIFN appears to be the driver of HBsAg decline and clearance when used in combination with tenofovir [13].

However, treatment is often associated with side effects that may lead to discontinuation of therapy in some patients. The ability to identify patients most likely to respond to PegIFN alfa would be clinically useful, allowing clinicians to select patients most likely to respond to a finite course of such treatment, as well as those most likely to require long-term treatment with NAs.

The use of on-treatment factors, such as the decline in HBV DNA or HBsAg, to identify patients likely or unlikely to respond to PegIFN alfa is well established in clinical practice [14]. Monitoring HBsAg and HBV DNA during PegIFN alfa-2a therapy can identify patients likely to achieve a post-treatment response [9,15-19]. Although the ability to identify potential responders during treatment is useful, it would be better to identify such patients before initiating treatment. Current guidelines recommend consideration of baseline factors in treatment decisions regarding HBeAg-positive patients only [5]. Limited data suggest that baseline factors can identify HBeAg-negative patients likely to respond to PegIFN alfa-2a [8,10,20,21]; however, the use of baseline factors to predict response is not currently part of routine clinical practice, because individual factors have relatively low negative and positive predictive values for treatment response.

The objective of this analysis was to develop a baseline scoring system to estimate, prior to treatment, the likelihood of an HBeAg-negative patient achieving a durable off-treatment response after receiving PegIFN alfa-2a therapy.

Materials and methods

We conducted a pooled retrospective analysis of data from three studies of PegIFN alfa-2a (PEGASYS®, Roche, Basel, Switzerland) in HBeAg-negative patients: a large phase III randomized study (WV16241) [8]; a phase IV randomized study (PegBeLiver, ML18253, NCT01095835) [9]; and a nonrandomized study (PERSEAS, ML22016, NCT01283074) [22]. Data from 323 HBeAg-negative patients were included (192 from WV16241 [8], 48 from PegBeLiver [9] and 83 from PERSEAS [22]). All patients with HBV genotype B/C infection were derived from WV16241 [8], while patients with genotype D infection were derived from all three studies [8,9,22]. Differences in response rates were derived from the three studies were not statistically significant (P>0.05, Chi Square tests).

Selection criteria

Inclusion and exclusion criteria for these studies have been described previously [8,9,22]. Patients infected with HBV genotype B, C, or D, assigned to 48 weeks’ treatment with PegIFN alfa-2a 180 µg/week (without/with lamivudine) and who had baseline quantitative HBsAg, HBV DNA and alanine aminotransferase (ALT) data, were included. Patients from WV16241 were included only if they had participated in a long-term follow-up study (WV16866).
Outcome definitions

Two outcomes were used in the analyses, both determined at 48 weeks post-treatment: HBV DNA <2000 IU/mL (virological response), and the combination of HBV DNA <2000 IU/mL plus normal ALT (combined response). Patients with missing HBV DNA or ALT values at 48 weeks post-treatment were considered non-responders.

Development of baseline prediction score

HBV genotype influences the response to IFN [5]; hence, separate scores were developed for patients infected with genotype B or C (B/C) and those with genotype D. The number of patients with genotype A infection was insufficient to develop a scoring system.

The baseline prediction scoring systems were developed following several steps. First, generalized additive models (GAMs) with the logit link were used to identify appropriate cutoffs via visual inspection for continuous predictors. This included a search for the most statistically significant cutoff for each continuous baseline factor using logistic regression analysis, which could be used to stratify patients into two subgroups corresponding to high and low response rates (virological response and combined response data were considered when defining cutoffs). Baseline factors included in these analyses were age, HBV DNA, HBsAg, ALT level and sex. Next, separate multiple logistic regression (MLR) models were developed for patients infected with genotypes B/C and D, using a backward elimination process that considered only those factors (and cutoffs) associated with a response with a P-value of <0.2 in the univariate analysis. In the backward selection procedure, a P-value of 0.15 was used to select factors that would remain in the models, which allowed for the inclusion of factors with moderate predictive value. Internal validation methods were employed to assess the stability of the selected model, using bootstrap resampling methods [23]. A variable was considered to be a reliable predictor if it was selected in at least 50% of 500 bootstrapped samples with replacement. The discrimination of the model was quantified by calculating the area under the receiver operating characteristics curve (ROC-AUC: target ≥0.7); corresponding 95% confidence interval (CI) and the optimism-corrected ROC-AUCs were determined using the bootstrap samples. The Hosmer-Lemeshow test was applied to assess the goodness of fit of the final model (i.e., no P-value <0.05).

Factors that remained in the model of only one endpoint were added to the model for the other, to ensure that the models for both endpoints contained the same set of factors. To devise the scoring system for patients infected with HBV genotype B/C or D, predictive baseline characteristics retained in the MLR models were assigned points, taking into account the magnitude of the regression coefficients estimated in the models [24]: a value of 0 points was assigned to the reference category for each predictive factor; next, points were assigned according to the size of the regression coefficient, using 1 as the unit for 1 point with rounding to integer values.

Points assigned for individual characteristics were summed to arrive at a total score for each patient, with higher scores indicating a higher chance of achieving a response. Response rates by prediction scores were determined, including 95% CIs. Prediction characteristics (sensitivity, specificity, positive predictive value, negative predictive value [NPV]) and the positive likelihood ratio were determined for the two endpoints for both prediction scores.

An analysis of patients with genotype B/C infection enrolled in China, Hong Kong and Taiwan is provided in the Supplementary Data.

Applicability of stopping rules

The applicability of the stopping rule recommended by the European Association for the Study of the Liver (no decline in HBsAg and <2-log_{10} decline in HBV DNA at treatment Week 12) was evaluated after stratifying patients by baseline score.

Results

Patients

A total of 778 patients were assigned to treatment, of whom 323 were included in the analysis of baseline characteristics. Reasons for exclusion from the analysis are shown in Supplementary Fig. 1. The study population comprised 157 patients with HBV genotype B/C infection, all of Asian ethnicity, and 166 patients with HBV genotype D infection, all of Caucasian ethnicity (Table 1). Baseline characteristics of patients from each study are shown in Supplementary Table 1. Among genotype B/C patients, the mean age was 38.9 years and 83.4% were male. Among genotype D patients, the mean age was 43.7 years, and 70.5% were male. The prevalence of METAVIR Stage 4 fibrosis was 7.6% among patients with genotype B/C infection who had a pretreatment biopsy result and 2.6% among patients with genotype D infection.

Development of baseline prediction score

Graphic analysis and univariate logistic regression analysis to identify baseline predictive factors and appropriate cutoffs

The GAM plots show the relationships between continuous variables (age, ALT ratio, HBsAg level, and HBV DNA level) and treatment response in patients infected with HBV genotype B/C (Supplementary Fig. 2A-H) and D (Supplementary Fig. 3A-H). Based on these analyses and the search for optimal cutoffs using univariate logistic regression (ULR) analysis, the following cutoffs were considered for further analysis in patients infected with HBV genotype B/C: age ≤30, >30-45 and >45 years; ALT ratio <5 and ≥5 × upper limit of normal (ULN); HBsAg level <1500 and ≥1500 IU/mL; and in patients infected with HBV genotype D: age ≤45 and >45 years; HBsAg <2500 and ≥2500 IU/mL; HBV DNA <35,000 and ≥35,000 IU/mL. Inspection of the GAM plots...
showed that the relationship between the response variables and HBV DNA level in genotype B/C patients (Supplementary Fig. 2A-H), and ALT ratio in genotype D patients was not monotonic (Supplementary Fig. 3A-H); therefore, no cutoffs were selected for further analyses.

Among the 157 patients with HBV genotype B/C infection, 57 individuals (36.3%) had a virological response and 46 individuals (29.3%) achieved a combined response at 48 weeks post-treatment. Among the 166 patients with HBV genotype D infection, 26 individuals (15.7%) achieved a virological response and 21 individuals (12.7%) achieved a combined response at 48 weeks post-treatment.

Response rates by baseline characteristics included in the ULR analyses are shown by HBV genotype (B/C and D) and type of response in Fig. 1A-D. The ULR analyses of response rates according to baseline characteristics are shown in Table 2 for patients infected with HBV genotype B/C and in Table 3 for patients infected with HBV genotype D.

**Multivariate logistic regression**

For patients infected with HBV genotype B/C, four baseline factors (age, ALT ratio, HBV genotype and HBsAg level) identified by ULR analysis as being predictive for responses at 48 weeks post-treatment (P<0.2) were included in the MLR model selection procedure (backward elimination with P<0.15). For virological response, age, HBV genotype and HBsAg remained significant after the backward selection process (all P<0.01, Table 2), while age, ALT ratio and genotype were retained after the selection process for a combined response (P<0.1, Table 2). The covariates of the final MLR models were selected in ≥60% of the bootstrap samples showing the model's stability, and the ROC-AUCs and Hosmer-Lemeshow test indicate sufficient discrimination and goodness of fit (Table 2, Supplementary Figs. 4,5). When all four factors were included in the MLR models for both endpoints, the regression coefficients and odds ratios of the significant factors were very similar (data not shown).

For patients infected with HBV genotype D, three baseline factors were retained in the final model for virological response (Table 3): age, HBsAg level and HBV DNA level (all P<0.1); and two factors were retained in the final model for a combined response (Table 3): age and HBsAg level (all P<0.01). Again, the covariates of the final MLR models were selected in ≥60% of the bootstrap samples showing the model's stability, and the ROC-AUCs and Hosmer-Lemeshow test indicate sufficient discrimination and goodness of fit (Table 3, Supplementary Figs. 6,7). When all three factors were included in the MLR model for a combined response, the regression coefficients and odds ratios of the significant factors were very similar (data not shown).
To formulate the scoring systems, baseline factors were ranked according to the regression coefficients, and points were assigned according to the size of the regression coefficient (Tables 2 and 3) using 1 as the unit for 1 point. For example, the regression coefficient for genotype (C vs. B), for virological response (1.1725), and for combined response (0.9366) divided by 1 and rounded to an integer value results in an assigned value of 1 point. In genotype B/C patients, 1 point was assigned for HBsAg <1250 IU/mL and for ALT $\geq 5 \times$ ULN, although these factors were only significant for one of the two post-treatment response variables. In genotype D patients, the final MLR model for virological response was used to devise the scoring system. The resulting scoring systems used to determine an individual patient’s probability of achieving a post-treatment response with PegIFN alfa-2a (Table 4) have a maximum score of 5 for genotype C, 4 for genotype B, and 3 for genotype D.

Among patients infected with HBV genotype B or C, the distribution of prediction scores in the patient population (N=157) was: 0 points: 5.1% (n=8); 1 point: 25.5% (n=40); 2 points: 39.5% (n=62); 3 points: 21.7% (n=34); 4 points: 7.6% (n=12); and 5 points: 0.6% (n=1). Predictive scores were grouped as follows: 0-1 (n=48), 2 (n=62) and ≥3 (n=47) points to form groups of patients with a low, moderate and high chance of response.

Among patients infected with HBV genotype D, the distribution of prediction scores in the patient population (N=166) was: 0 points: 27.1% (n=45); 1 point: 50.6% (n=84); 2 points: 15.1% (n=25); and 3 points: 7.2% (n=12). Predictive scores were grouped as follows: 0-1 (n=129), 2 (n=25) and 3 (n=12) points to form groups of patients with a low, moderate and high chance of response.

**Response rates according to predictive score**

Response rates at 48 weeks post-treatment increased with increasing baseline score in patients infected with HBV genotypes B/C or D. Among 47 patients infected with HBV...
Predictors of PegIFN response in HBeAg-negative CHB

Genotype B/C and scores ≥3, 33 individuals (70.2%) achieved a virological response and 27 individuals (57.4%) achieved a combined response and among 48 patients with scores of 0-1, eight individuals (16.7%) achieved a virological response and six individuals (12.5%) achieved a combined response (Fig. 2A). The performance characteristics of the two scores are provided in Supplementary Table 2 and the response rates by prediction score are shown in Supplementary Table 3 for genotypes B and C separately. The results are consistent across genotypes B and C, for example, for patients with genotype B or C the combined response rate is ≥50% in those with prediction scores ≥3 and ≤25% for those with scores ≤2. Among genotype B patients with a baseline score of 0-1 the NPV for a combined response was 90.9% (30 of 33 patients did not achieve a response); however, for patients with a score of 0, the NPV was 100% (45/45), and for those patients with a score of 1, the NPV was 88.1% (74/84).

HBsAg clearance rates also increased with increasing baseline score (Fig. 2C).

**Applicability of on-treatment stopping rule by predictive score**

When patients were categorized by baseline score, there were no statistically significant differences in the number of individuals (genotype B/C or D) who met the criterion for discontinuing therapy at Week 12 (Supplementary Table 4).

**Discussion**

This analysis shows that HBeAg-negative patients with a high or low chance of achieving a post-treatment response to PegIFN alfa-2a can be identified with genotype-specific baseline scoring systems. The systems employ readily available demographic and laboratory data and could easily be incorporated into routine patient visits to assist
With the prediction scores developed here, points are assigned for age, ALT ratio, HBV genotype and HBsAg level for patients with genotype B/C infection, and for age, HBsAg and HBV DNA levels for patients with genotype D infection. Individual patient scores range from 0 to 5 for genotype B/C patients and from 0 to 3 for genotype D patients. Higher scores indicate a greater likelihood of achieving a virological response (HBV DNA <2000 IU/mL) and a combined response (HBV DNA <2000 IU/mL plus normal ALT) 48 weeks after completing a standard 48-week course of treatment with PegIFN alfa-2a. Among genotype B/C patients, patients with scores ≥3 had a virological response of 97%, compared to 42% for patients with scores 0-2 (unadjusted odds ratio 3.493, 95% CI 1.571-7.766, p = 0.0022). Among genotype D patients, patients with scores ≥3 had a virological response of 77%, compared to 27% for patients with scores 0-2 (unadjusted odds ratio 2.607, 95% CI 0.985-6.901, p = 0.0536).

### Table 2

| Baseline characteristic | Univariate logistic regression | Final multiple logistic regression model | Selected in bootstrap models |
|-------------------------|-------------------------------|----------------------------------------|-----------------------------|
|                         | Regression coefficient | Odds ratio (95% CI) | P-value | Regression coefficient | Odds ratio (95% CI) | P-value |
| Virological response (HBV DNA <2000 IU/mL) at 48 weeks post-treatment | | | | | | |
| Sex, male vs. female | 0.2960 | 1.344 (0.544, 3.323) | 0.5214 | 3.230 (1.460, 7.149) | 0.0038 | 94% |
| Genotype C vs. B | 0.8609 | 2.365 (1.177, 4.755) | 0.0157 | 1.1725 | <0.0001 | |
| Age, years | | | | | | |
| ≤30 vs. >45 | 1.9135 | 6.777 (2.447, 18.771) | 0.0002 | 2.4469 | 11.552 (3.680, 36.267) | <0.0001 | |
| 31–45 vs. >45 | 0.7049 | 2.024 (0.820, 4.996) | 0.1264 | 0.9584 | 2.607 (0.985, 6.901) | 0.0536 | |
| ALT ratio ≥5 vs. <5 × ULN | 0.7689 | 2.157 (0.883, 5.271) | 0.0916 | | | |
| HBsAg <1250 vs. ≥1250 IU/mL | 0.8517 | 2.344 (1.173, 4.681) | 0.0158 | 1.2506 | 3.493 (1.571, 7.766) | 0.0022 | 97% |

**Optimism-corrected AUC = 0.761, 95% CI 0.682-0.840**

**Hosmer-Lemeshow P-value = 0.91**

| Combined response (HBV DNA <2000 IU/mL and normal ALT) at 48 weeks post-treatment | | | | | | |
| Sex, male vs. female | 0.1403 | 1.151 (0.448, 2.957) | 0.7708 | 2.551 (1.131, 5.755) | 0.0240 | 84% |
| Genotype C vs. B | 0.7690 | 2.158 (1.027, 4.534) | 0.0423 | 0.9366 | 2.607 (1.082, 6.527) | 0.0536 | |
| Age, years | | | | | | |
| ≤30–45 vs. >45 | 2.3871 | 10.882 (3.221, 36.767) | 0.0001 | 2.4158 | 11.199 (3.169, 39.573) | 0.0002 | |
| 31–45 vs. >45 | 1.2726 | 3.570 (1.139, 11.195) | 0.0291 | 1.2392 | 3.453 (1.082, 11.016) | 0.0363 | |
| ALT ratio ≥5 vs. <5 × ULN | 1.1658 | 3.209 (1.297, 7.939) | 0.0117 | 0.8729 | 2.394 (0.878, 6.527) | 0.0880 | 60% |
| HBsAg <1250 vs. ≥1250 IU/mL | 0.0496 | 1.051 (0.504, 2.192) | 0.8947 | | | |

**Optimism-corrected AUC = 0.758, 95% CI: 0.676-0.840**

**Hosmer-Lemeshow P-value = 0.75**

*Wald chi-square test for each regression coefficient; **P>0.25 in the univariate analysis and thus not considered in the multiple logistic regression analysis and the bootstrap samples*

**ALT, alanine aminotransferase; AUC, area under the curve; CI, confidence interval; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; ULN, upper limit of normal**
A high chance (>50%), with a score of 2, a moderate chance (~25%), and with scores of 0 or 1, a low chance (~15%) of achieving a post-treatment response.

A smaller proportion of patients with the more-challenging-to-treat genotype D infection were identified as having a high chance of achieving a post-treatment response. Moreover, the response rates were lower in these individuals than in those with genotype B/C infection. Although genotype D patients are less likely to be suitable candidates for PegIFN alfa-2a, the tool may be most useful in this difficult-to-treat group, because it identifies those few patients who may benefit from PegIFN alfa-2a. In the PegBeLiver study, a total of 12% of patients (6/51) randomized to a standard 48-week course of PegIFN alfa-2a therapy achieved a combined response (which was defined as HBV DNA <3400 IU/mL and normal ALT levels 48 weeks post-treatment); none of these cleared HBsAg [9]. In the present analysis, few patients were identified as having a good chance of a response (baseline score of 3); however, one-third of patients with a baseline score of 3 achieved a combined response at 48 weeks post-treatment. Thus, the results imply that, although overall response rates to PegIFN alfa-2a are quite low in patients with genotype D infection, it is possible to identify the few patients who are most likely to respond on the basis of their baseline characteristics.

The utility of the scoring systems may vary according to genotype. For patients with genotype B infection, who usually have the highest response rates, the score may be most useful to confirm the suitability of patients for peginterferon alfa-2a, whereas, for genotype C and particularly D infection, response rates are generally lower and the score may be most appropriate for excluding patients who may have been considered for peginterferon alfa-2a treatment.

No decline in HBsAg and <2-log₁₀ decline in HBV DNA by Week 12 of treatment with PegIFN alfa (PARC rule) is recommended as a stopping rule because few patients that meet

### Table 3 Univariate and multiple logistic regression analysis of factors associated with virological response (HBV DNA <2000 IU/mL) and combined response (HBV DNA <2000 IU/mL and normal ALT) in patients infected with HBV genotype D

| Baseline characteristic | Virological response (HBV DNA <2000 IU/mL) at 48 weeks post-treatment | Combined response (HBV DNA <2000 IU/mL and normal ALT) at 48 weeks post-treatment |
|-------------------------|------------------------------------------------------------------------|----------------------------------------------------------------------------------|
|                         | Regression coefficient | Odds ratio (95% CI) | P-value⁠ | Regression coefficient | Odds ratio (95% CI) | P-value⁠ |
| Sex, female vs. male    | 0.0706                  | 1.073               | 0.8790   | 0.4453                  | 1.561               | 0.3592   |
| Age ≤45 vs. >45 years   | 1.0608                  | 2.889               | 0.0322   | 1.6675                  | 5.299               | 0.0098   |
| HBsAg <2500 vs. ≥2500 IU/mL | 1.1214              | 3.069               | 0.0129   | 1.0987                  | 3.000               | 0.0242   |
| HBV DNA <35,000 vs. ≥35,000 IU/mL | 1.5756           | 4.833               | 0.0005   | 1.1423                  | 3.134               | 0.0194   |

AUC = 0.763, 95% CI: 0.672–0.854, Optimism-corrected AUC = 0.721, Hosmer-Lemeshow P-value = 0.32

AUC = 0.730, 95% CI: 0.631–0.829, Optimism-corrected AUC = 0.691, Hosmer-Lemeshow P-value = 0.92

⁠<sup>a</sup>Wald chi-square test for each regression coefficient; ⁠<sup>b</sup>P>0.25 in the univariate analysis and thus not considered in the multiple logistic regression analysis and the bootstrap samples

ALT, alanine aminotransferase; AUC, area under the curve; CI, confidence interval; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus
Table 4 Scoring system for predictive baseline characteristics in patients infected with HBV genotypes B or C, or D

| Characteristic                  | Score |
|--------------------------------|-------|
| **HBV genotype B or C**        |       |
| Age, years                     |       |
| >45                            | 0     |
| >30-45                         | 1     |
| ≤30                            | 2     |
| ALT ratio, × ULN               |       |
| <5                             | 0     |
| ≥5                             | 1     |
| **HBV genotype**               |       |
| B                              | 0     |
| C                              | 1     |
| **HBsAg, IU/mL**               |       |
| ≥1250                          | 0     |
| <1250                          | 1     |
| **HBV genotype D**             |       |
| Age, years                     |       |
| >45                            | 0     |
| ≤45                            | 1     |
| **HBsAg, IU/mL**               |       |
| ≥2500                          | 0     |
| <2500                          | 1     |
| **HBV DNA, IU/mL**             |       |
| ≥35,000                        | 0     |
| <35,000                        | 1     |

ALT, alanine aminotransferase; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; ULN, upper limit of normal

these criteria have achieved a response in clinical trials [25,26]. In this analysis we were not able to correlate the baseline prediction scoring system with the PARC rule.

Strengths of the present analysis include the development of separate scoring systems for patients with genotype B/C and D infection and the use of 48-week post-treatment data in the outcome definition. The ability to identify patients likely to develop off-treatment responses with PegIFN alfa-2a is desirable because it would allow clinicians to target treatment to those patients most likely to respond, while minimizing the number of likely non responders who are exposed to the potential adverse events of treatment.

Limitations include the retrospective nature of the study and the pooled analysis, the relatively small number of patients (especially genotype D patients), the lack of a score for genotype A patients and the lack of an external validation cohort. The lack of ethnic diversity within each genotype group (i.e., all genotype B/C patients were Asian and all genotype D patients were Caucasian) and the lack of comprehensive information on liver disease severity are also limitations. HBV genotyping may not be reimbursed in some countries, so the need for HBV genotype may be seen as a limitation; however, the subanalysis in Asian patients shows that the score can still be used when this factor is not available. The scoring system must be prospectively validated before it can be recommended for use in routine clinical practice, and it would be reassuring to confirm whether patients identified by these scoring systems also have a high probability of experiencing HBsAg clearance or seroconversion. Future analyses should also attempt to evaluate the baseline scoring system prospectively, while applying on-treatment stopping rules. Application of these prediction scores requires that clinicians have knowledge of the HBV genotype and quantitative HBsAg levels. These tests may not be available in all practice settings, which would limit the utility of this tool.

In conclusion, the proposed baseline scoring systems for HBeAg-negative patients infected with HBV genotype B/C and D use readily available baseline characteristics and enable physicians to identify patients with a low, moderate, or high chance of achieving a post-treatment response to PegIFN alfa-2a. The benefit/risk ratio should be carefully considered before initiating treatment in patients with scores of 0-1, given that these indicate a low chance of success.

**Summary Box**

**What is already known:**

- Peginterferon alfa-2a induces durable responses in approximately one-third of hepatitis B e antigen-negative patients with chronic hepatitis B
- Baseline disease factors, such as higher hepatitis B virus DNA and hepatitis B surface antigen levels, have been associated with a reduced likelihood of response
- Criteria to identify specific patients likely or unlikely to respond to peginterferon alfa-2a would permit targeting of resources and sparing of treatment for patients unlikely to respond

**What the new findings are:**

- For patients with genotype B or C disease, a baseline scoring system from 0 to 5 was developed based on age, genotype, alanine aminotransferase ratio and HBsAg level
- Only 16.7% of genotype B or C patients with scores of 0 or 1 achieved a virological response, while 70.2% of those with scores of 3-5 responded
- For patients with genotype D disease, a scoring system from 0 to 3 based on age, hepatitis B surface antigen level and hepatitis B virus DNA level was predictive of response


Acknowledgments

The authors thank the study investigators, the patients and their families. The authors would like to thank the investigators of the studies included in the analysis for their contribution to the generation of data. Support for third-party writing assistance for this manuscript, furnished by Blair Jarvis, BSc and Leigh Church, PhD, of Health Interactions, was provided by F. Hoffmann-La Roche Ltd, Basel, Switzerland.

References

1. Naghavi M, Wang H, Lozano R, et al; GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015;385:117-171.

2. Hadziyannis SJ, Vassilopoulos D. Hepatitis B e antigen-negative chronic hepatitis B. Hepatology 2001;34:617-624.

3. Brunetto MR, Oliveri F, Coco B, et al. Outcome of anti-HBe positive chronic hepatitis B in alpha-interferon treated and untreated patients: a long term cohort study. J Hepatol 2002;36:263-270.

4. Funk ML, Rosenberg DM, Lok AS. World-wide epidemiology of HBeAg-negative chronic hepatitis B and associated precore and core promoter variants. J Viral Hepat 2002;9:52-61.

5. European Association for the Study of the Liver. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. J Hepatol 2017;67:370-398.

6. Sarin SK, Kumar M, Lau GK, et al. Asian-Pacific clinical practice guidelines on the management of hepatitis B: a 2015 update. Hepatol Int 2016;10:1-98.

7. Chevaliez S, Hézode C, Bahrami S, Grare M, Pawlotsky JM. Long-term hepatitis B surface antigen (HBsAg) kinetics during nucleoside/nucleotide analogue therapy: finite treatment duration unlikely. J Hepatol 2013;58:676-683.

8. Marcellin P, Lau GK, Bonino F, et al; Peginterferon Alfa-2a-Negative Chronic Hepatitis B Study Group. Peginterferon alfa-2a alone, lamivudine alone, and the two in combination in patients with HBeAg-negative chronic hepatitis B. N Engl J Med 2004;351:1206-1217.

9. Lampertico P, Viganò M, Di Costanzo GG, et al; PegBeLiver Study Group. Randomised study comparing 48 and 96 weeks peginterferon α-2a therapy in genotype D HBeAg-negative chronic hepatitis B. Gut 2013;62:290-298.

10. Marcellin P, Bonino F, Lau GK, et al; Peginterferon alfa-2a in HBeAg-negative Chronic Hepatitis B Study Group. Sustained response of hepatitis B e antigen-negative patients 3 years after treatment with peginterferon alpha-2a. Gastroenterology 2009;136:2169-2179.

11. Marcellin P, Bonino F, Yurdyydín C, et al. Hepatitis B surface antigen levels: association with 5-year response to peginterferon alfa-2a in hepatitis B e-antigen-negative patients. Hepatol Int 2013;7:88-97.

12. Reijnders JG, Rijckborst V, Sonneveld MJ, et al. Kinetics of hepatitis B surface antigen differ between treatment with peginterferon and entecavir. J Hepatol 2011;54:449-454.

13. Marcellin P, Ahn SH, Ma X, et al. Combination of tenofovir disoproxil fumarate and peginterferon α-2a increases loss of hepatitis B surface antigen in patients with chronic hepatitis B. Gastroenterology 2016;150:134-144.

14. Martinot-Peignoux M, Asselah T, Marcellin P. HBsAg quantification to optimise treatment monitoring in chronic hepatitis B patients. Liver Int 2015;35(Supppl 1):82-90.

15. Brunetto MR, Moriconi F, Bonino F, et al. Hepatitis B virus surface antigen levels: a guide to sustained response to peginterferon alfa-2a in HBeAg-negative chronic hepatitis B. Hepatology 2009;49:1141-1150.

16. Moucari R, Korevaar A, Lada O, et al. High rates of HBsAg seroconversion in HBeAg-positive chronic hepatitis B patients responding to interferon: a long-term follow-up study. J Hepatol 2009;50:1084-1092.

17. Brunetto MR. A new role for an old marker, HBsAg. J Hepatol 2010;52:475-477.

18. Janssen HL, Sonneveld MJ, Brunetto MR. Quantification of serum hepatitis B surface antigen: Is it useful for the management of chronic hepatitis B? Gut 2012;61:641-645.

19. Brunetto MR, Marcellin P, Cherubini B, et al. Response to peginterferon alfa-2a (40KD) in HBeAg-negative CHB: on-treatment kinetics of HBsAg serum levels vary by HBV genotype. J Hepatol 2013;59:1153-1159.

20. Bonino F, Marcellin P, Lau GK, et al; Peginterferon Alfa-2a HBeAg-Negative Chronic Hepatitis B Study Group. Predicting response to peginterferon alpha-2a, lamivudine and the two combined for HBeAg-negative chronic hepatitis B. Gut 2007;56:699-705.

21. Takkenberg RB, Jansen L, de Niet A, et al. Baseline hepatitis B surface antigen (HBsAg) as predictor of sustained HBsAg loss in chronic hepatitis B patients treated with pegylated interferon-α2a and adefovir. Antivir Ther 2013;18:895-904.

22. Gouili S, Karatapanis S, Akriviadis E, et al. On-treatment prediction of sustained response to peginterferon alfa-2a for HBeAg-negative chronic hepatitis B patients. Liver Int 2015;35:1540-1548.

23. Steyerberg E. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Springer-Verlag: New York; 2009.

24. Sullivan LM, Massaro JM, D’Agostino RB Sr. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. Stat Med 2004;23:1631-1660.

25. Rijckborst V, Hansen BE, Cakaloglu Y, et al. Early on-treatment kinetics of HBsAg serum levels vary by HBV genotype. J Hepatol 2009;50:1084-1092.

26. Rijckborst V, Hansen BE, Ferkic P, et al. Validation of a stopping rule at week 12 using HBsAg and HBV DNA levels. Hepatology 2010;52:454-461.

27. Shiga H, Zalata K. Chapter 10. Ishak versus METAVIR: Terminology, convertibility and correlation with laboratory changes in chronic hepatitis. In: Takahashi H (editor): Liver Biopsy. InTech, DOI: 10.5772/20110.
Supplementary Figure 1 Reasons for exclusion from the analysis by study

*HBV, hepatitis B virus*
Supplementary Figure 2 (Continued)
Supplementary Figure 2 GAM analysis in patients infected with HBV genotypes B or C: Virological response (HBV DNA <2000 IU/mL) 48 weeks post-treatment by (A) age, (B) ALT ratio, (C) HBsAg and (D) HBV DNA; Combined response (HBV DNA <2000 IU/mL and normal ALT) 48 weeks post-treatment by (E) age, (F) ALT ratio, (G) HBsAg, and (H) HBV DNA

ALT, alanine aminotransferase; BL, baseline; DF, degree of freedom; GAM, generalized additive model; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; trt, treatment; ULN, upper limit of normal
Supplementary Figure 3 (Continued)
Supplementary Figure 3 (Continued)
Application of the scoring system in Asian patients with HBV genotype B or C infection

Among the 157 patients with HBV genotype B or C infection, 135 individuals (86.0%) were enrolled in centers in China, Hong Kong, or Taiwan. The prediction score for genotype B/C was applied after removing the point assigned for genotype C. This was done because HBV genotype is rarely determined in patients with chronic hepatitis B in this region. The distribution of scores changed markedly after the removal of one point assigned for genotype in all patients with genotype C infection. Thus, among the 135 Chinese patients, prediction scores were distributed as follows: 0 points: 12.6% (n=17); 1 point: 39.3% (n=53); 2 points: 34.8% (n=47); 3 points: 12.6% (n=17); and 4 points: 0.7% (n=1). Predictive scores were grouped as follows: 0-1 (n=70), 2 (n=47), and ≥3 (n=18) points to form groups of patients with a low, moderate, and high chance of response.

Among Chinese patients infected with HBV genotype B/C, the percentage of patients with a virological response at 48 weeks post-treatment increased with increased score from 22.9% (scores 0-1), to 44.7% (score 2), to 83.3% (scores ≥3), while the corresponding combined response rates were 15.7%, 40.4%, and 66.7% (Supplementary Fig. 8). This indicates that the modified score might be applicable in Chinese patients with missing genotype determination, where almost all patients have genotype B or C.
**Supplementary Figure 4** ROC curves for HBV DNA <2000 IU/mL at 48 weeks post-treatment in genotype B/C patients for final model and for each individual factor in the model.

*HBV*, hepatitis B virus; *HBsAg*, hepatitis B surface antigen; *ROC*, receiver operating characteristics

**Supplementary Figure 5** ROC curves for combined response at 48 weeks post-treatment in genotype B/C patients for final model and for each individual factor in the model.

*ALT*, alanine aminotransferase; *HBV*, hepatitis B virus; *ROC*, receiver operating characteristics

**Supplementary Figure 6** ROC curves for HBV DNA <2000 IU/mL at 48 weeks post-treatment in genotype D patients for final model and for each individual factor in the model.

*HBV*, hepatitis B virus; *HBsAg*, hepatitis B surface antigen; *ROC*, receiver operating characteristics
**Supplementary Figure 7** ROC curves for combined response at 48 weeks post-treatment in genotype D patients for final model and for each individual factor in the model.

*HBsAg, hepatitis B surface antigen; ROC, receiver operating characteristics*

**Supplementary Figure 8** Response rates at 48 weeks post-treatment by baseline prediction score in patients enrolled in centers in China, Hong Kong, or Taiwan (N=135) infected with HBV genotype B or C (Error bars are 95% confidence intervals)
**Supplementary Table 1** Baseline characteristics of patients in the analysis according to original study

| Characteristic | WV16241 (n=182) | ML18253 (PegBeLiver) (n=48) | ML22016 (PERSEAS) (n=83) |
|---------------|----------------|-----------------------------|--------------------------|
| Age, years, mean ±SD |                  |                            |                          |
| Sex, male, n (%) | 158 (82.3)      | 31 (64.6)                   | 59 (71.18)               |
| Race, n (%)      |                  |                            |                          |
| Caucasian        | 35 (18.2)        | 48 (100)                    | 83 (100)                 |
| Asian            | 157 (81.8)       | 0                           | 0                        |
| ALT ratio, x ULN, mean±SD | 3.00±2.60       | 2.91±2.6                    | 2.42±1.62                |
| HBsAg, log\(_{10}\) IU/mL, mean±SD | 3.35±0.57       | 3.77±0.42                   | 3.53±0.59                |
| HBV DNA, log\(_{10}\) IU/mL, mean±SD | 6.54±1.97       | 6.22±1.22                   | 5.36±1.31                |

**Supplementary Table 2** Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and positive likelihood ratio for the two prediction scores for genotype B/C and D

| Prediction score | True | False | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | Positive likelihood ratio |
|------------------|------|-------|----------------|-----------------|---------|---------|--------------------------|
| Positive Negativ  |       |       |                |                 |         |         |                          |
| Pol € 2 or C patients |     |       |                |                 |         |         |                          |
| Combined response (HBV DNA<2000 IU/mL and normal ALT) at 48 weeks post-treatment | | | | | | | |
| ≥1               | 45   | 7     | 104            | 1               | 97.8    | 6.3     | 30.2                     | 87.5      | 1.04                   |
| ≥2               | 40   | 42    | 69             | 6               | 87.0    | 37.8    | 36.7                     | 87.5      | 1.40                   |
| ≥3               | 27   | 91    | 20             | 19              | 58.7    | 82.0    | 57.4                     | 82.7      | 3.26                   |
| ≥4               | 9    | 107   | 4              | 37              | 19.6    | 96.4    | 69.2                     | 74.3      | 5.43                   |
| ≥5               | 1    | 111   | 0              | 45              | 2.2     | 100     | 100                      | 71.2      | NA                     |
| Virological response (HBV DNA<2000 IU/mL) at 48 weeks post-treatment | | | | | | | |
| ≥1               | 56   | 7     | 93             | 1               | 98.2    | 7.0     | 37.6                     | 87.5      | 1.06                   |
| ≥2               | 49   | 40    | 60             | 8               | 86.0    | 40.0    | 45.0                     | 83.3      | 1.43                   |
| ≥3               | 33   | 86    | 14             | 24              | 57.9    | 86.0    | 70.2                     | 78.2      | 4.14                   |
| ≥4               | 11   | 98    | 2              | 46              | 19.3    | 98.0    | 84.6                     | 68.1      | 9.65                   |
| ≥5               | 1    | 100   | 0              | 56              | 1.8     | 100     | 100                      | 64.1      | NA                     |
| Genotype D patients |     |       |                |                 |         |         |                          |
| Combined response (HBV DNA<2000 IU/mL and normal ALT) at 48 weeks post-treatment | | | | | | | |
| ≥1               | 21   | 45    | 100            | 0               | 31.0    | 17.4    | 100                      | 1.45      |                       |
| ≥2               | 11   | 119   | 26             | 10              | 52.4    | 82.1    | 29.7                     | 92.2      | 2.92                   |
| ≥3               | 4    | 137   | 8              | 17              | 19.0    | 94.5    | 33.3                     | 89.0      | 3.45                   |
| Virological response (HBV DNA<2000 IU/mL) at 48 weeks post-treatment | | | | | | | |
| ≥1               | 25   | 44    | 96             | 1               | 96.2    | 31.4    | 20.7                     | 97.8      | 1.40                   |
| ≥2               | 13   | 116   | 24             | 13              | 50.0    | 82.9    | 35.1                     | 89.9      | 2.92                   |
| ≥3               | 6    | 134   | 6              | 20              | 23.1    | 95.7    | 50.0                     | 87.0      | 5.39                   |

Sensitivity=100 x true positive/(true positive+false negative); Specificity=100 x true negative/(true negative+false positive); PPV=100 x true positive/(true positive+false positive); NPV=100 x true negative/(true negative+false negative); Positive likelihood ratio=Sensitivity/(100 – Specificity)

ALT, alanine aminotransferase; HBV, hepatitis B virus
### Supplementary Table 3 Response rates by prediction score in genotype B or C patients

| Prediction score | N  | n  | %  |
|------------------|----|----|----|
| **Genotype B patients** |    |    |    |
| Combined response (HBV DNA<2000 IU/mL and normal ALT) at 48 weeks post-treatment |    |    |    |
| 0-1              | 33 | 3  | 9.1 |
| 2                | 24 | 6  | 25.0|
| 3                | 7  | 4  | 57.1|
| Total            | 64 | 13 | 20.3|
| Virological response (HBV DNA<2000 IU/mL) at 48 weeks post-treatment |    |    |    |
| 0-1              | 33 | 5  | 15.2|
| 2                | 24 | 6  | 25.0|
| 3                | 7  | 5  | 71.4|
| Total            | 64 | 16 | 25.0|
| **Genotype C patients** |    |    |    |
| Combined response (HBV DNA<2000 IU/mL and normal ALT) at 48 weeks post-treatment |    |    |    |
| 0-1              | 15 | 3  | 20.0|
| 2                | 38 | 7  | 18.4|
| 3                | 27 | 14 | 51.9|
| ≥4               | 13 | 9  | 69.2|
| Total            | 93 | 33 | 35.5|
| Virological response (HBV DNA<2000 IU/mL) at 48 weeks post-treatment |    |    |    |
| 0-1              | 15 | 3  | 20.0|
| 2                | 38 | 10 | 26.3|
| 3                | 27 | 17 | 63.0|
| ≥4               | 13 | 11 | 84.6|
| Total            | 93 | 41 | 44.1|

ALT, alanine aminotransferase; HBV, hepatitis B virus

### Supplementary Table 4 Baseline prediction score by PARC rule status*

| PARC rule status | Baseline prediction score | Total |
|------------------|---------------------------|-------|
|                  | 0 | 1 | 2 | 3 | 4 | 5 |
| **Genotype B or C (n=157)** |    |    |    |    |    |    |
| Fulfilled        | 0/5 | 4/19 | 3/39 | 0/17 | 1/6 | 0/1 | 8/87 |
| Not fulfilled    | 5/5 | 15/19 | 36/39 | 17/17 | 5/6 | 1/1 | 79/87 |
| Missing          | 3 | 21 | 23 | 17 | 6 | 0 | 70 |
| **Genotype D (n=166)** |    |    |    |    |    |    |
| Fulfilled        | 8/38 | 9/66 | 2/20 | 2/7 | NA | NA | 21/131 |
| Not fulfilled    | 30/38 | 57/66 | 18/20 | 5/7 | NA | NA | 110/131 |
| Missing          | 7 | 18 | 5 | 5 | NA | NA | 35 |

*PARC rule fulfilled=no decline in HBsAg and $<$2-log$_{10}$ decline in HBV DNA from baseline to Week 12 of treatment

Pearson Chi-Square Test for association between baseline prediction score and PARC rule status: P=0.4886 for HBV genotype B or C and P=0.4121 for HBV genotype D

HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; NA, not applicable