2018-03

Application of prognostic scores in the STOPAH trial: Discriminant function is no longer the optimal scoring system in alcoholic hepatitis

Forrest, EH

10.1016/j.jhep.2017.11.017
Journal of Hepatology
Elsevier

PREDICTING OUTCOME IN ALCOHOLIC HEPATITIS: PERFORMANCE AND APPLICATION OF EXISTING PROGNOSTIC SCORES IN THE STOPAH TRIAL.

Ewan H Forrest[1], Stephen R Atkinson[2], Paul Richardson[3], Steven Masson[4], Stephen Ryder[6], Mark R Thursz[2], Michael Allison[5]
On behalf of the STOPAH Trial Management Group*

[1] Department of Gastroenterology, Glasgow Royal Infirmary, Glasgow
[2] Liver Unit, Imperial College, London
[3] Liver Unit, Royal Liverpool Hospital, Liverpool
[4] Liver Unit, Freeman Hospital, Newcastle
[5] Liver Unit, Addenbrooke's Hospital, Cambridge Biomedical Research Centre, Cambridge
[6]Liver Unit, Queen's Medical Centre, Nottingham

* Members of the STOPAH Trial Management Group other than the authors: Dr Anne McCune, Bristol; Dr Dermot Gleeson, Sheffield; Dr Andrew Austin, Derby; Dr David Patch, London, Dr Ashwin Dhanda, Plymouth; Dr Debbie Shawcross, London; Dr Mark Wright, Southampton.

Correspondence to:
Dr Ewan Forrest
Department of Gastroenterology
Glasgow Royal Infirmary
Castle Street
Glasgow G4 0SF
United Kingdom

Tel: 0141 232 0734
Fax: 0141 552 6126
E mail: Ewan.Forrest@ggc.scot.nhs.uk

KEYWORDS: Alcoholic Hepatitis; Prednisolone; Prognosis

Word Count: ; Figures: 2 Tables: 6; Supplementary Figures: 2; Supplementary Tables: 4
Conflicts of Interest: none
Financial Support: STOPAH Trial funded by the National Institute for Health Research Health Technology Assessment program.
Author Contributions: Study Concept and Design: EHF, SRA, MA; Analysis and Interpretation of Data: EHF, SRA, MA; Drafting of Manuscript: EHF; SRA, MA; Critical Revision of Manuscript for Important Intellectual Content: EHF, SRA, PR, SM, MRT, MA.

Acknowledgments:
Dr Alan H Forrest, Fellow of Royal Statistical Society, for statistical advice.

ABSTRACT

Background and Aims:

Prognostic models in alcoholic hepatitis (AH) may be 'static', using data from a single time point, or 'dynamic', incorporating data from subsequent time points, such as the bilirubin at 7 days. Commonly applied ‘static’ scores include the Discriminant Function (DF), Glasgow Alcoholic Hepatitis Score (GAHS), the Age, Bilirubin, INR and Creatinine (ABIC) score and the Model of End-Stage Liver Disease (MELD); common ‘dynamic’ scores are the percentage change in serum bilirubin and the Lille Score. The aim of this study was to assess the predictive performance of these scores in patients recruited to the STOPAH trial.

Methods:

The primary and secondary endpoints in this study were day 28 and 90 mortality, respectively. Predictive performance of scores was assessed by area under the Receiver Operating Curve (AUC). The effect of different therapeutic strategies upon survival was assessed by Kaplan-Meier analysis and tested using the log-rank test.

Results:

1068 patients were studied. The AUC for the DF was 0.670, significantly lower than for MELD, ABIC and GAHS at 0.704, 0.726 and 0.713 respectively. ‘Dynamic scores’ the evolution of ‘static’ scores by Day 7 revealed similar AUCs. Patients with consistently low ‘static’ scores had a good 28-day outcome not improved with prednisolone (MELD<25: 8.6%; ABIC<.71: 6.6%; GAHS<9: 5.9%). In patients not presenting with gastro-intestinal bleeding or sepsis, 28-day mortality was lower with prednisolone (MELD: 22.2% *v* 28.9%, p=0.13; ABIC 14.6% *v* 21% p=0.02 GAHS 21% *v* 29.3%, p=0.04). Compared with treating all patients with a DF ≥ 32 and subsequent Lille assessment (overall 90-day mortality 26.8%), strategies combining ‘static’ scores

with ‘dynamic’ scores led to fewer patients being exposed to prednisolone with a trend to an overall improvement in survival (overall 90-day mortality: MELD/Lille 21.8%; ABIC/Lille 23.7%; GAHS/Lille 20.6%).

Conclusion:

MELD, ABIC and GAHS perform better than the DF in the prognostic assessment of AH. Low scores are associated with a favourable outcome not improved with prednisolone. Combination of the baseline ‘static’ scores with Day 7 scores identified patients benefiting most from prednisolone treatment, reduced the number of patients exposed to corticosteroids and improved 90-day outcome.

Lay Summary:

Alcoholic hepatitis is a severe life-threatening condition. Several scores exist to determine the outcome of these patients as well as to identify those who may benefit from treatment. This study looked at the performance of existing scores in patients who had been recruited to the largest alcoholic hepatitis clinical trial: STOPAH.

‘Static’ scores are calculable at the start of assessment. The three newer static scores (ABIC, GAHS and MELD) were shown to be superior to the oldest score (DF). ABIC and GAHS could also identify patients who had a survival benefit 28 days after starting prednisolone treatment. ‘Dynamic’ scores relate to the change in disease over the first week of treatment. Combination of the ‘static’ scores ‘with the ‘dynamic’ scores or change in ‘static’ scores allowed identification of patients who could benefit from prednisolone up to 90 days.

INTRODUCTION

Alcoholic hepatitis is an acute and florid manifestation of alcoholic liver disease with high short and medium-term mortality[1]. Recognition of those at risk of a poor outcome is fundamental to structuring patient management. Several prognostic scores have been developed to predict the course of alcoholic hepatitis, however to be useful in a clinical context, scores should not only identify patients with a poor prognosis but also to direct patient care[2].

The Discriminant Function (DF) has become established in clinical practice with a threshold greater than or equal to 32 identifying those with severe disease[3,4]. However concerns have been raised regarding the reliability of the DF as it uses the absolute value of prothrombin time rather than a ratiometric value such as the International Normalised Ratio (INR)[5]. Combined analysis of five alcoholic hepatitis trials which used the DF to determine severity showed 28-day mortality of 20% for corticosteroid treated patients and 35% for untreated patients[6]. However, several of these studies were more than twenty years old[3,7,8]. More recent studies including the Steroids Or Pentoxifylline for Alcoholic Hepatitis (STOPAH) trial have shown improvements in 28-day outcome[9,10,11]. These improved outcomes with the modern management of alcoholic hepatitis will likely reduce further the specificity of the DF to identify those at greatest risk of death compared with historical trials.

Three alternative scores have been proposed to determine prognosis based upon variables obtained from a single time-point described as ‘static’ scores. These are the Glasgow Alcoholic Hepatitis Score (GAHS)[1], the Age, Bilirubin, INR and Creatinine (ABIC) score[12] and the Model of End-Stage Liver Disease (MELD) score. These scores seek to stratify patients into two (GAHS, MELD) or three (ABIC) groups characterised by significantly different prognoses. Well established cut-offs exist for

the GAHS (severe disease: >8) and ABIC (low, intermediate, high:) scores. There is no widely accepted optimal cut off for MELD; proposed values range from 18 to 30.5[13,14,15,16]. The ABIC score seeks to stratify patients into those at low, intermediate or high risk of death with those in the intermediate group perhaps most likely to benefit from corticosteroids. In a retrospective study the GAHS has been suggested as being able to identify those who are likely to benefit from corticosteroid treatment[17].

Other measures which include the change in bilirubin levels over the first week of treatment have been used to assess likely benefit from corticosteroid therapy: sometimes described as ‘dynamic' scores. These are the Early Change in Bilirubin Levels (ECBL)[18], percentage change in serum bilirubin (%ΔBili)[19] and the Lille Score[20]. These scores were originally described to be used in a dichotomous fashion with any fall in bilirubin (ECBL), a greater than or equal to 25%ΔBili or a Lille Score less than 0.45 being associated with a favourable outcome, or corticosteroid 'response'. Patients admitted with alcoholic hepatitis present at different stages of their illness. Whilst some will present at a point where their disease severity is at its worst and then improve; others will present at a point where their disease is on a continuing trajectory of deterioration. Given this variation it is not unexpected that baseline severity scores, although they can give an index of severity and mortality risk, may have relatively low predictive value. By the same logic, it would be anticipated that a means of assessment of the severity that incorporates the evolution of disease over time, with or without treatment, would be more accurate in predicting outcome. The combination of a ‘static’ score with a ‘dynamic’ score would seem to be a reasonable strategy to identify those with an initial poor prognosis whose condition does not show improvement after starting corticosteroid treatment[21].

The aim of this study was to assess the performance of these existing scores for alcoholic hepatitis to predict outcome and to assess their application to different treatment strategies in patients recruited to the STOPAH trial.

METHODS

Patients recruited to the STOPAH trial were studied. The characteristics of these patients have been described in detail previously[22]. Inclusion was based upon a clinical diagnosis of alcoholic hepatitis with recent onset of jaundice and heavy alcohol misuse and no other forms of liver disease. All patients had a baseline DF greater than or equal to 32. Patients were randomised by a factorial design to receive Prednisolone and Placebo, Pentoxifylline and Placebo, Prednisolone and Pentoxifylline or double Placebo. Analysis was performed on an intention to treat basis.

'Static' scores (GAHS, ABIC and MELD) were calculated on baseline data at the time of starting treatment: 'dynamic' scores (Lille, ECBL and %ΔBili) were calculated only in those patients with data available after 7 days of treatment. In addition the changes (Delta, Δ) of each of the 'static' scores between baseline and Day 7 were calculated. As the GAHS is a categorical score allocating a natural number to each variable to create a score, it was possible to categorise some patients as greater than or less than 9 even if some variables were missing if categorisation was arithmetically inevitable. The MELD was calculated using the UNOS variation. A high ABIC score was defined as a value greater than or equal to 6.71.

Assessment of Prognosis

For the purposes of assessing the overall discriminatory ability of static scores and comparison between them, the whole patient cohort was studied. Mortality at 28 and 90 days after randomisation were analysed. Survival beyond 90 days was not assessed in this context as other factors, such as continued alcohol use, were felt to influence outcome beyond that point and so not be reflective of a prognostic score's accuracy. Patients were consented for follow-up using the NHS data linkage service so that even if lost to follow-up their outcomes could be captured.

Application to Treatment Strategies

The application of these scores to different treatment strategies which integrated 'static' and 'dynamic' was assessed. The overall STOPAH results showed no therapeutic effect with pentoxifylline at any time point and so patients randomised to pentoxifylline were analysed as per untreated (placebo) patients. Thus the therapeutic comparison was for those treated with prednisolone or not. Whilst assessing the whole patient cohort, two pre-specified analyses were included. Initially all patients treated with prednisolone were analysed and then those who presented with either gastrointestinal bleeding (GIB) or sepsis were excluded for further analysis. Secondly patients with low MELD, GAHS or ABIC scores at baseline for whom no further score was available at seven days or whose Day 7 score remained favourable were analysed separately to represent those with consistently low scores.

The treatment strategies were based upon only treating patients with high initial baseline 'static' scores (MELD≥25; ABIC≥6.71 or GAHS>8) and assessing response at Day 7 by either change in 'static' score or by 'dynamic' scores.

Analysis was performed using MedCalc v17.4. Comparison of scores was performed by area under the Receiver Operating Curve (AUC) analysis. Optimal cut-offs were identified by calculating the Youden Index (*J*). Kaplan-Meier analysis was used to assess survival and survival curves were compared using the Log-Rank test. Results are presented with 95% confidence intervals (95% CI). Number needed to treat calculations were made using the inverse of the absolute risk reduction.

RESULTS

Data on 1068 patients recruited to the STOPAH trial was available for analysis: of whom 534 received prednisolone. Patient characteristics are shown in Table 1. GIB and/or sepsis were features of initial presentation in 199 patients leaving 869 patients who presented without either of these complications. At Day 7 data was available on 720 patients to calculate the dynamic scores indicating corticosteroid effect: GIB or sepsis were presenting features of 143 patients who had Day 7 data available, leaving 577 who presented without either of these complications (Supplementary Figure 1).

Assessment of Prognosis

Mortality data up to and including day 90 were available for all patients. Overall mortality was 16.3% and 26.7% at 28 and 90 days respectively. For those presenting with GIB or sepsis initially the mortalities were 18.1% and 29.2% compared with 15.9% and 26.1% for those without these features.

Analysis of Baseline 'Static' Scores

The AUC analyses are shown in Figure 1. For prediction of both 28-day and 90-day outcomes the MELD, ABIC and GAHS had similar values without any significant differences. For 28-day outcome the values were 0.732, 0.747 and 0.753, and for 90-day outcome 0.704, 0.726 and 0.713 respectively. However all three of these 'static' scores were superior to the DF at both time-points (Table 2) which had AUC values of 0.673 and 0.670 for 28-day and 90-day outcome respectively.

Calibration of 'Static' Scores

For the ABIC and GAHS scores there are established cut-offs to identify prognostic groups. The MELD cut-off used to determine prognosis varies between publications. A recommended cut-off of 18 encompassed 99% of the STOPAH patients with a DF greater than or equal to 32 and was therefore thought to lack adequate specificity. The

optimal cut-off using the STOPAH data was 25 (*J*=0.36 for 28-day and *J*=0.32 for 90-day outcomes) This gave high negative predictive values of 91.4% and 83.2% for 28-day and 90-day outcomes respectively. The positive predictive values were more modest at 29.2% for 28-day outcome but improved to 43.2% for 90-day outcome. Comparison with the established cut-offs for the GAHS and ABIC scores are shown in Supplementary Table 1.

Of the patients studied, 61% had a MELD less than 25, 14% had an ABIC less than 6.71 and 53% had a GAHS less than 9 at baseline. Overall accuracy in identifying outcome at Day 28 was 66% for MELD greater than or equal to 25, 30% for ABIC greater than or equal to 6.71 and 61% for GAHS greater than or equal to 9.

Day 7 Scores: Analysis of Evolution of 'Static' Scores and 'Dynamic' Scores

Comparison of the change in 'static' score by Day 7 with the 'dynamic' scores indicated the Lille Score to have the highest AUC at 0.732 and 0.722 respectively for 28- and 90-day outcome. The AUC values obtained for either the 'dynamic' scores or the changes in 'static' scores were not significantly different. Comparison between these scores as well as with the initial baseline 'static' scores for those patients with Day 7 data available is shown in Table 3.

Some patients with an initial low baseline score showed an increase to a higher score category by Day 7. Of patients with an initial GAHS less than 9, 11.8% developed a GAHS greater than or equal to 9 on Day 7. For patients with an initial MELD less than 25, 4.5% subsequently developed a score greater than 25 and 2.6% of patients with an ABIC less than 6.71 showed a rise to greater than 6.7 by Day 7.

Lille Score stratification of corticosteroid treated patients into three categories of complete, partial or null responders showed 28-day mortalities for each group of 6.3%, 4.6% and 27.4%, and the 90-day mortalities 12.6%, 17.2% and 46.4% respectively.

Mortalities at Day 28 and Day 90 using the original cut point of greater than or equal to 0.45 were 27.8% and 43.6%, not significantly different from the null responders.

Application to Treatment Strategies

Outcome of Patients with Low Baseline 'Static' Scores (Supplementary Table 2)

The mortality rates after 28 days for those with low initial 'static' scores not treated with prednisolone were 11.7%, 6.3% and 9.3% for low MELD, ABIC and GAHS values respectively. The equivalent mortality figures for those treated with prednisolone were 6.3%, 1.4% and 4.4%. The differences were significant for the MELD and GAHS values. However, any difference was not sustained to Day 90.

Of this group of patients, those whose scores remained consistently below these thresholds had 28-day mortality rates that did not differ between those treated and not treated with prednisolone: treated 5.7%, 1.4% and 3.3% for low MELD, ABIC and GAHS values respectively; untreated 8.6%, 6.6% and 5.9%. Similarly there were no differences in 90-day mortality in this sub-group between steroid-treated and untreated patients.

Those patients whose initial 'static' score was low, but then rose above these thresholds over 7 days had a poor outcome. For those whose GAHS rose to greater than or equal to 9, the 28-day and 90-day mortalities were 27.0% and 41.7% respectively. A rise in MELD above 25 gave 28-day and 90-day mortalities were 52.2% and 60.9% respectively.

Outcome of Patients with High Baseline 'Static' Scores (Table 4)

Analysis of all patients with high initial baseline scores did not show any survival benefit with prednisolone at either 28 or 90 days. However with the exclusion of those who present with either GIB or sepsis the 28-day mortality of patients with an initial ABIC score greater than 6.7 treated with prednisolone was 14.6% compared to 21.0%

for untreated patients (p=0.02; 95%CI 1.06, 2.13). Similarly for patients with a GAHS greater than 8 the 28-day mortality was 21.0% with prednisolone and 29.3% for those untreated (p=0.04; 95%CI 1.02, 2.24). Patients with a MELD greater than 25 did not show a significant reduction in mortality at Day 28 with prednisolone treatment even with exclusion of GIB or septic patients. By Day 90 any difference in mortality between treated and untreated patients identified by any score had disappeared. The number needed to treat with prednisolone to prevent an individual death at 28 days was 12 for patients with a GAHS greater than 8, and 16 for those with an ABIC greater than 6.7.

<u>'Dynamic' Scores and Evolution of 'Static' Scores in Prednisolone Treated Patients.</u>

As 28-day survival benefit was seen in patients who did not present with GIB or sepsis with high 'static' score categories, these patients were analysed with regards to the 'dynamic' scores and evolution of 'static' scores.

When applied to prednisolone treated patients with an ABIC score greater than 6.70, the highest AUC value for 28-day outcome was 0.737 for ΔABIC and for 90-day outcome 0.698 for ΔGAHS. For patients with a GAHS greater than 8, the highest AUC value for 28-day outcome was 0.722 for ΔABIC and for 90-day outcome 0.688 for ΔGAHS. For patients with a MELD greater than 25, the highest AUC value for both 28-day outcome and for 90-day outcome was the Lille score at 0.712 and 0.681 respectively. There were no significant differences between any of these AUC values (Figure 3). Youden Index analysis identified a fall in ABIC by greater than or equal to 0.29, GAHS by greater than or equal 1 or MELD by greater than or equal 2.6 to be indicative of response.

Prednisolone treatment led to a greater percentage fall in serum bilirubin by Day 7. The proportions of patients with high baseline 'static' scores who could be classified

favourably after 7 days were greater for those patients who were treated with prednisolone. (Table 5)

Combination of Static and Dynamic Scores

Assessment of different combinations of baseline 'static' scores, with a 'dynamic' scores or a change in 'static' scores over 7 days demonstrated differences between different therapeutic strategies. Combining a baseline ('static') score with a Day 7 score ('dynamic' score or change in initial 'static' score over 7 days) allows separation of patients into three main groups: 1) a consistently low 'static' score; 2) a high initial 'static' score treated with corticosteroids with a favourable Day-7 response; 3) a high initial 'static' score treated with corticosteroids with an unfavourable Day 7 response (Supplementary Figure 2). Compared with the standard approach of treating all patients with a DF greater than 32 with subsequent Lille assessment, such strategies led to fewer patients being exposed to prednisolone with a trend to an overall improvement in 90-day survival (Table 6). Only the GAHS/ΔGAHS strategy led to a significant reduction in overall 90-day mortality (19.2% compared with 28.2%: p=0.026: 95%CI 0.63% to 14.72%). Survival curves for the DF/ Lille, MELD/ Lille, ABIC/ Lille, GAHS/Lille and GAHS/ΔGAHS strategies are shown (Figure 2).

DISCUSSION

Prognostic scores should be clinically useful as well as statistically sound. The DF has provided a consistency to clinical treatment and research in alcoholic hepatitis. However it is a sensitive score with less impressive specificity and alternative scores have been proposed: 'static' scores based on variables at a single timepoint such as MELD, ABIC and GAHS[1,12,13] and 'dynamic' scores based upon evolution of serum bilirubin over the first week of corticosteroid treatment such as ECBL, %ΔBili and the Lille Score[18,19,20].

This analysis of STOPAH trial data shows that all three new 'static' scores performed similarly in predicting outcome at 28 and 90 days and with greater discriminatory power than the DF in this population all of whom had a DF greater than or equal to 32. Therefore a re-calibration to a higher cut off of DF to make it more specific would still fall short of the newer scores' discriminatory ability. The analysis also indicates that a MELD threshold of 25 was optimal as previously described[15]. Analysis of the 'dynamic scores indicates that the Lille score has most useful overall prognostic capability, although the other dynamic scores and the 7 day change in 'static' scores also perform reasonably well. Sub-stratification of the Lille into Complete, Partial and Null Responders did not appear to add any additional useful prognostic information compared with the original description of a 0.45 cut off[6]. Overall the newer scores are more accurate but still have relatively modest discriminatory power.

Whilst the three newer 'static' scores had similar prognostic capabilities, there were differences in the application of these scores and identification of patients who might benefit from corticosteroid treatment. An overall treatment benefit from prednisolone was seen in patients with low 'static' scores at baseline. However, the conduct of the STOPAH trial in this regard would not reflect clinical stratification based upon a second

risk threshold. Sub-group analysis of patients with consistently low scores is justified by the dynamic nature of alcoholic hepatitis, with patients presenting at different stages of their illness. On account of this and the high sensitivity of the DF, patients with low values of newer more specific scores may move to higher values as their condition evolves irrespective of treatment. In clinical practice the movement of these patients to a poor prognostic group would trigger consideration of prednisolone treatment.

In this retrospective sub-group analysis patients with consistently low 'static' scores which did not rise above their threshold for severe disease had a favourable outcome irrespective of whether they received prednisolone or not, with no statistically significant additional benefit from prednisolone treatment. However despite having consistently low scores these patients still need optimal general clinical management including nutritional support and the surveillance and treatment of infection as indicated. They also require monitoring as an increase in the values of these scores above their threshold of severity is associated with a high mortality. In such circumstances corticosteroid treatment may be considered, essentially using the cut-off threshold of the 'static' prognostic score as a trigger to intervention.

Patients with high baseline 'static' scores initially did not appear to benefit from prednisolone treatment. However on excluding those patients who presented initially with either GIB or sepsis an improvement in 28-day survival for those with high GAHS (greater than 8) and ABIC scores (greater than 6.70) was seen with prednisolone. There is evidence that the natural history of apparent alcoholic hepatitis in patients who present with these complications may be different. Patients presenting with GIB may have a more favourable outcome[23] perhaps related to the routine use of antibiotics in this clinical circumstance resulting in fewer infections[10]. Despite a previous study indicating that treated infection had no impact upon subsequent

corticosteroid effect[24], more recent data from the STOPAH trial indicates that those patients with sepsis at presentation do not benefit from corticosteroids unless combined with a continuing course of antibiotics[25]. Therefore patients with such complications at presentation should be considered as specific groups for whom corticosteroid monotherapy may not be suitable.

However, even with additional stratification using either the GAHS and ABIC scores, it was only possible to identify patients deriving short-term (28-day) benefit from prednisolone. However, when used in combination with a Day 7 score (a 'dynamic' score or change in 'static' score), it may be possible to identify a sub-group of patients who derive benefit at 90 days. In the context of prednisolone treated patients with high baseline 'static' scores, all the Day-7 scores (the 'dynamic' scores and change in 'static' scores) performed similarly. The benefit from prednisolone appeared to be related to a greater fall in bilirubin leading to more favourable Day 7 scores.

The combination of these results indicates the scores can be applied clinically, such that consistently low 'static scored patients do not receive prednisolone but high 'static' score patients receive treatment with subsequent assessment at of response at Day 7. Compared to the standard approach of treating all patients with a DF greater than 32 and assessing by a Lille response at Day 7, use of the new scores led to fewer patients receiving prednisolone, more specific identification of corticosteroid non-responders and a reduction in overall mortality. This appeared to be particularly so using the GAHS as a baseline 'static' score and change in GAHS to assess response to treatment.

Whilst these retrospective sub-group analyses reflect 'real-world' management of this group of patients, caution should be used in interpreting their results. The STOPAH trial was designed to prospectively evaluate the benefit derived from prednisolone

treatment in patients with a DF greater than or equal to 32; consequently re-casting analyses based upon treatment instituted at a second, higher threshold of severity comes with an attendant risk of introducing bias. Nonetheless, it seems apparent that the use of a newer baseline score and a Day 7 score to stratify treatment reduces exposure to prednisolone without detriment to patient outcome and merits prospective evaluation.

In conclusion, application of existing prognostic scores to the largest prospective study of alcoholic hepatitis shows that the more recently advocated ‘static’ scores (ABIC, GAHS and MELD) are superior to the DF in determining mortality risk. Consistently low ‘static’ scores identify a sub-group with such a low event rate that any potential beneficial effect of prednisolone is difficult to establish. Prednisolone can be offered to those with high ‘static’ scores, excluding those who present initially with sepsis or GIB, with subsequent response assessed by a variety of measures after 7 days. The approach suggested in this paper reduces the number of patients exposed to corticosteroids, reserving this for those with detectable day 90 mortality benefit, and allows for identification of those with no response at Day 7 who should be considered for new interventional therapies. Whilst this approach can be used clinically, predictive abilities remain modest and there is a need for more accurate identification of patients who will respond to corticosteroid, ideally using information available at baseline, rather than at Day 7.

REFERENCES

1. Forrest E H, Evans C D J, Stewart S, et al Analysis of factors related to mortality in alcoholic hepatitis and the derivation and validation of the Glasgow alcoholic hepatitis score. Gut 2005541174–1179.1179.
2. Wyatt JC, Altman DG. Prognostic Models: clinically useful or quickly forgotten? British Medical Journal 1995; 311: 3159.
3. Maddrey WC, Boitnott JK, Bedine MS, Weber FL, Mezey E, White RI. Corticosteroid therapy of alcoholic hepatitis. Gastroenterology 1978; 75: 193-199.
4. Carithers JRL, Herlong HF, Diehl AM, Shaw EW, Coombes B, Fallon HJ, Maddrey WC. Methylprednisolone therapy in patients with severe alcoholic hepatitis: a randomized multicenter trial. Ann Intern Med 1989; 110: 685-690.
5. Robert A, Chazouillères O. Prothrombin time in liver failure: time, ratio, activity percentage, or international normalized ratio? Hepatology 1996; 24: 1392–4.
6. Mathurin P, O'Grady J, Carithers RL, et al. Corticosteroids improve short-term survival in patients with severe alcoholic hepatitis: meta-analysis of individual patient data. Gut 2011; 60: 255-260.
7. Mendenhall CL, Anderson S, Garcia-Pont P, Goldberg S, Kiernan T, Seef LB, Sorrell M, Tamburro C, Weesner R, Zetterman R, Chedid A, Chen T, Rabin L. Short-term and long-term survival in patients with alcoholic hepatitis treated with oxdandrolone and prednisolone. NEJM 1984; 311: 1464-1470.
8. Ramond MJ, Poynard T, Rueff B, Mathurin P, Theodore C, Chaput JC, Benhamou JP. A randomized trial of Prednisolone in patients with severe alcoholic hepatitis. NEJM 1992; 326:507-512.

9. Park SH, Kim DJ, Kim YS, et al. Pentoxifylline vs. corticosteroid to treat severe alcoholic hepatitis: a randomised, non-inferiority, open trial. Journal of Hepatology 2014; 61: 792-798.
10. Rudler M, Mouri S, Charlotte F, et al. Prognosis of treated severe alcoholic hepatitis in patients with gastrointestinal bleeding. Journal of Hepatology 2015; 62: 816-821.
11. Thursz MR, Richardson P, Allison M, et al. Prednisolone or pentoxifylline for alcoholic hepatitis. N Engl J Med 2015; 372: 1619–1628.
12. Dominguez M, Rincon D, Abraldes JG, Miquel R, Colomenero J, Bellot P, Garcia-Pagan J-C, Fernandez R, Moreno M, Banares R, Arroyo V, Caballeria J, Gines P, Bataller R. A new scoring system for prognostic stratification of patients with alcoholic hepatitis. American Journal of Gastroenterology 2008; 103: 2747-2756.
13. Dunn W, Jamil LH, Brown LS, Wiesner RH, Kim WR, Menon KVN, Malinchoc M, Kamath PS, Shah V. MELD accurately predicts mortality in patients with alcoholic hepatitis. Hepatology 2005; 41: 353-358.
14. Srikureja W, Kyulo NL, Runyon BA, Hu K-Q. MELD score is a better prognostic model than Child-Turcotte-Pugh score or Discriminant Function score in patients with alcoholic hepatitis. Journal of Hepatology; 2005; 42: 700-706.
15. Rincon D, Lo locano O, Ripoll C et al. Prognostic value of hepatic venous pressure gradient for in-hospital mortality of patients with severe acute alcoholic hepatitis. Aliment Pharmacol Ther 2007; 25: 841-8.
16. Soultati AS, Dourakis SP, Alexopoulo A, Deutsch M, Vasileva L, Archimandritis AJ. Predicting utility of a model for end stage liver disease in alcoholic liver disease. World Journal Gastroenterology 2006: 12; 4020-4025.

17. Forrest EH. Morris AJ. Stewart S. Phillips M. Oo YH. Fisher NC. Haydon G. O'Grady J. Day CP. The Glasgow alcoholic hepatitis score identifies patients who may benefit from corticosteroids. Gut 2007; 56:1743-6.
18. Mathurin P, Abdelnour M, Ramond MJ, et al. Early Change in bilirubin levels is an important prognostic factor in sever alcoholic hepatitis treated with prednisolone. Hepatology 2003; 38: 1363-9.
19. Morris JM, Forrest EH. Bilirubin response to corticosteroids in alcoholic hepatitis. Eur J Gastroenterol Hepatol 2005; 17: 759–62.
20. Louvet A, Naveau S, Abdelnour M, et al. The Lille model: a new tool for therapeutic strategy in patients with severe alcoholic hepatitis treated with steroids. Hepatology 2007; 45: 1348–54.
21. Louvet A, Labreuche J, Artu F, et al. Combining Data From Liver Disease Scoring Systems Better Predicts Outcomes of Patients With Alcoholic Hepatitis. Gastroenterology 2015; 149: 398-406.
22. Forrest EH, Mellor J, Stanton L, et al. Steroids or pentoxifylline for alcoholic hepatitis (STOPAH): study protocol for a randomised controlled trial. Trials 2013; 14: 262.
23. Lafferty HD, Johnson V, Keith R et al. 'Alcoholic Hepatitis' after Gastro-intestinal bleeding has a good prognosis. Journal of Hepatology 2007; 46: S271-2.
24. Louvet A, Wartel F, Castel H, et al. Infection in patients with severe alcoholic hepatitis treated with steroids: early response to therapy is the key factor. Gastroenterology 2009; 137(2):541-8.
25. Vergis N, Atkinson SR, Knapp S, et al. Patients with severe alcoholic hepatitis given prednisolone therapy who have high circulating levels of bacterial DNA

are at increased risk for developing infections. Gastroenterology 2016; 152: 1068-1077.

Figure and Table Legends

Figure 1: AUC for MELD, GAHS and ABIC Scores

Figure 2: Kaplan-Meier Survival Probability for Patients Stratified by 'Static' Scores (only high baseline scores treated with prednisolone) combined with 'Dynamic' Scores (excluding initial presentation with GIB or sepsis). R: Responder; NR: Non-responder
A) DF/Lille; B) MELD/Lille; C) ABIC/ Lille; D) GAHS/Lille; E) GAHS/ΔGAHS

Table 1: Patient characteristics

Table 2: Comparison of AUCs: MELD, ABIC and GAHS compared with DF for both 28 and 90-day Outcome

Table 3: Comparison of AUCs: 'Dynamic' Scores (Lille, ECBL, %ΔBili), change in 'static' scores (ΔMELD, ΔABIC, ΔGAHS) and baseline 'static' scores (MELD, ABIC, GAHS: for patients with Day 7 data available only) for both 28 and 90-day Outcome

Table 4: A) Mortality in all Patients with High Static Scores; B) Mortality in Patients with High Static Scores (excluding patients presenting with GIB or Sepsis)

Table 5: Effect of Prednisolone upon Day 7 bilirubin and Day 7 scores of Response.

Table 6: Application of Scores to Different Therapeutic Strategies and 90-Day Mortality.

Supplementary Figure 1: Flowchart of patients analysed.

Supplementary Figure 2: Supplementary Figure 2: Flowchart showing Stratified Management of Patients with Alcoholic Hepatitis in the STOPAH Trial.

Supplementary Table 1: Comparison of Cut Offs for A) MELD, B) ABIC, C) GAHS, D) ECBL, E) %ΔBili, F) Lille Score

Supplementary Table 2: A) Mortality in Patients with Low ‘Static’ Scores; B) Mortality in Patients with Consistently Low ‘Static’ Scores

Supplementary Figure 1: Flowchart of patients analysed.

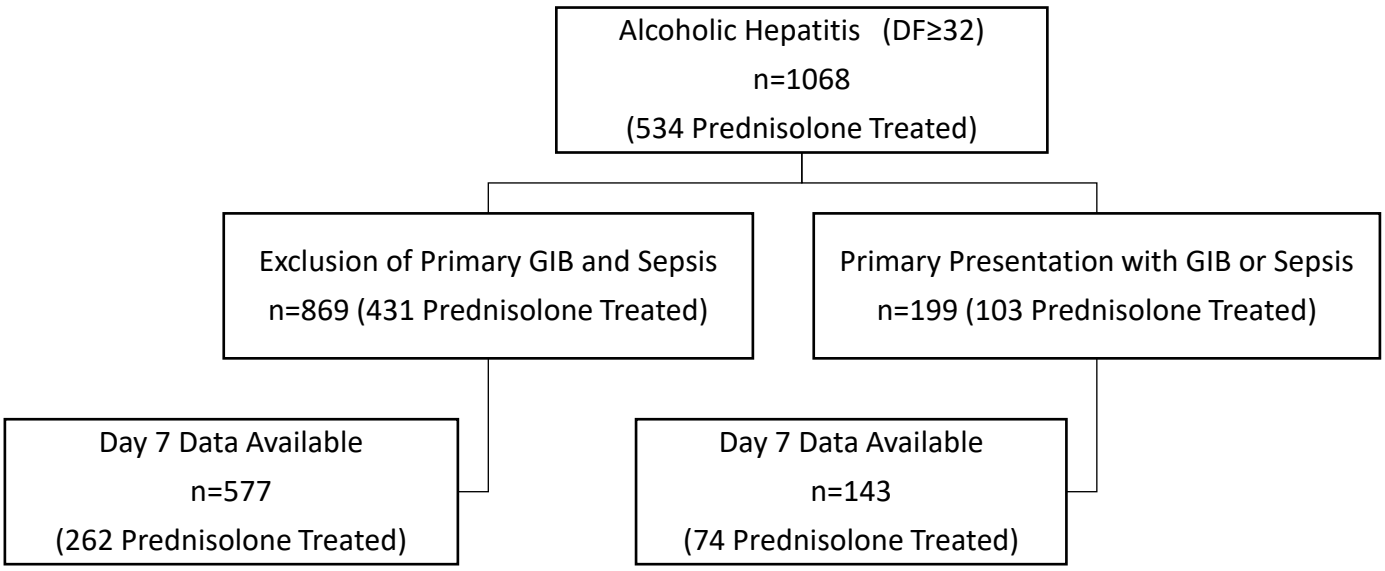
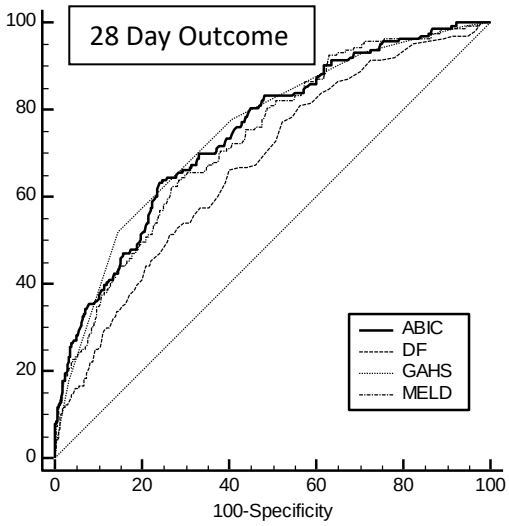


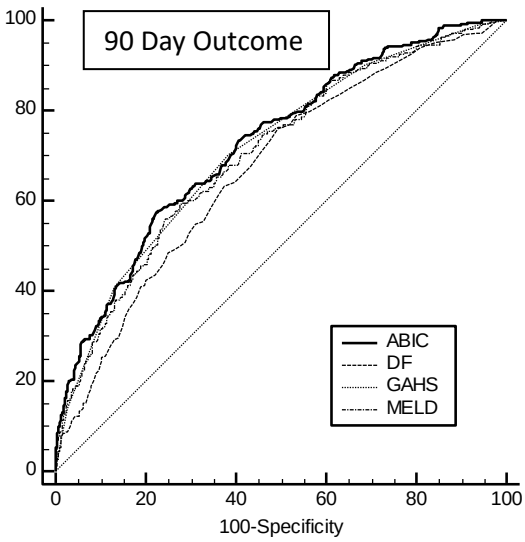
Table 1: Patient characteristics

	Prednisolone Treated	No Prednisolone
Age	49.4 (48.5, 50.3)	48.8 (47.9, 49.6)
Bilirubin (μmol/l)	308.3 (294.9, 321.7)	303.4 (290.6, 316.1)
INR	1.86 (1.82, 1.90)	1.90 (1.86, 1.94)
WCC (10^9/l)	10.1 (9.7, 10.6)	10.0 (9.5, 10.5)
Creatinine (μmol/l)	81.6 (77.3, 85.9)	76.3 (72.4, 80.1)
ABIC≥6.71	435 (86%)	435 (85%)
ABIC≥9.0	141 (28%)	119 (23%)
GAHS≥9	256 (48%)	250 (47%)
MELD≥25	206 (40%)	199 (39%)
Sepsis or GIB	103 (19%)	96 (18%)

In parentheses: 95% Confidence Intervals or percentage of available data
No significant differences between Prednisolone treated and untreated patients.

Figure 1: AUC for MELD, GAHS and ABIC Scores





	28-Day Outcome		90-Day Outcome	
	AUC	95% CI	AUC	95% CI
ABIC	0.747	0.719 to 0.774	0.726	0.697 to 0.753
DF	0.673	0.643 to 0.702	0.670	0.640 to 0.699
GAHS	0.753	0.725 to 0.779	0.713	0.684 to 0.741
MELD	0.732	0.704 to 0.759	0.704	0.675 to 0.732

Table 2: Comparison of AUCs: MELD, ABIC and GAHS compared with DF for both 28 and 90-day Outcome

	28-Day Outcome	90-Day Outcome
ABIC ~ DF_		
95% Confidence Interval	0.020 to 0.128	0.009 to 0.103
z statistic	2.673	2.333
Significance level	P = 0.008	P = 0.020
DF ~ GAHS		
95% Confidence Interval	0.032 to 0.128	0.002 to 0.084
z statistic	3.273	2.037
Significance level	P = 0.001	P = 0.042
DF ~ MELD		
95% Confidence Interval	0.026 to 0.093	0.006 to 0.062
z statistic	3.453	2.352
Significance level	P = 0.001	P = 0.019

Supplementary Table 1: Comparison of Cut Offs for A) MELD, B) ABIC, C) GAHS, D) ECBL, E) %ΔBili, F) Lille Score

A) MELD

	Cut Off	Sensitivity	Specificity	+PV	95% CI	-PV	95% CI	Accuracy
28-Day Outcome	≥25	66.06	69.19	29.2	24.7-34.1	91.4	88.9-93.4	65.5%
90-Day Outcome	≥25	59.63	71.81	43.2	38.1-48.4	83.2	80.1-86.0	66.2%

B) ABIC

	Cut Off	Sensitivity	Specificity	+PV	95% CI	-PV	95% CI	Accuracy
28-Day Outcome	<6.70	96.36	17.11	18.3	15.7-21.0	96.1	91.7-98.5	29.8%
	≥9	52.73	79.86	33.5	27.8-39.6	89.8	87.4-91.8	73.2%
90-Day Outcome	<6.70	95.56	18.62	29.7	26.6-32.8	92.1	86.6-95.9	38.9%
	≥9	47.41	82.45	49.2	43.0-55.5	81.4	78.4-84.1	73.2%

C) GAHS

	Cut Off	Sensitivity	Specificity	+PV	95% CI	-PV	95% CI	Accuracy
28-Day Outcome	≥9	77.51	58.00	26.1	22.4-30.2	93.1	90.6-95.1	61.2%
90-Day Outcome	≥9	70.61	60.57	39.3	35.0-43.8	85.1	81.8-87.9	63.2%

D) ECBL

	Cut Off	Sensitivity	Specificity	+PV	95% CI	-PV	95% CI	Accuracy
28 Day Outcome	>0	47.06	83.84	30.2	18.3-44.3	91.4	86.8-94.8	70.1%
90 Day Outcome	>0	36.23	85.57	47.2	33.3-61.4	79.0	72.9-84.3	68.0%%

E) %ΔBili

	Cut Off	Sensitivity	Specificity	+PV	95% CI	-PV	95% CI	Accuracy
28 Day Outcome	<25	79.41	56.58	21.4	14.6-29.6	94.9	89.7-97.9	49.6%
90 Day Outcome	<25	73.91	61.14	40.5	31.8-49.6	86.8	79.9-92.0	56.2%

F) Lille Score

	Cut Off	Sensitivity	Specificity	+PV	95% CI	-PV	95% CI	Accuracy
28 Day Outcome	≥0.45	78.12	65.14	24.8	16.7-34.3	95.3	90.6-98.1	62.4%
90 Day Outcome	≥0.45	65.67	68.85	43.6	33.7-53.8	84.6	77.7-90.0	65.0%

Accuracy indicates the percentage of correctly predicted outcomes using the cut offs described.
+/-: positive/ negative; PV: Predictive Value

Table 3: Comparison of AUCs: 'Dynamic' Scores (Lille, ECBL, %ΔBili), change in 'static' scores (ΔMELD, ΔABIC, ΔGAHS) and baseline 'static' scores (MELD, ABIC, GAHS: for patients with day 7 data available only) for both 28 and 90 day Outcome

	28-Day Outcome		90-Day Outcome	
	AUC	**95% CI**	**AUC**	**95% CI**
ΔABIC	0.725	0.687 to 0.761	0.682	0.642 to 0.719
ΔGAHS	0.688	0.649 to 0.726	0.678	0.639 to 0.716
ΔMELD	0.714	0.675 to 0.750	0.678	0.639 to 0.716
ECBL	0.690	0.651 to 0.727	0.668	0.629 to 0.706
%ΔBili	0.703	0.664 to 0.740	0.692	0.653 to 0.730
Lille	0.732	0.694 to 0.768	0.722	0.684 to 0.758
ABIC	0.695	0.658 to 0.731	0.703	0.666 to 0.739
GAHS	0.698	0.661 to 0.733	0.677	0.640 to 0.714
MELD	0.672	0.634 to 0.708	0.668	0.630 to 0.705

Supplementary Table 2:

A) Mortality in Patients with Low Static Scores

		28-Day Outcome		90-Day Outcome	
		Mortality	**Significance; 95%CI**	**Mortality**	**Significance; 95%CI**
MELD<25	Pred	6.3%	P = 0.016	16.2%	P = 0.492
	Non-Pred	11.8%	0.305 to 0.869	17.8%	0.597, 1.282
ABIC<6.71	Pred	1.4%	P = 0.115	7.6%	P = 0.939
	Non-Pred	6.3%	0.0424 to 1.041	8.2%	0.337, 3.242
GAHS<9	Pred	4.4%	P = 0. 025	14.4%	P = 0.660
	Non-Pred	9.3%	0.247 to 0.881	15.4%	0.589, 1.399

B) Mortality in Patients with Consistently Low Static Scores

		28-Day Outcome		90-Day Outcome	
		Mortality	**Significance; 95%CI**	**Mortality**	**Significance; 95%CI**
MELD<25	Pred	5.7%	P = 0.162	15.8%	P = 0.740
	Non-Pred	8.6%	0.353 to 1.185	14.4%	0.708, 1.626
ABIC<6.71	Pred	1.4%	P = 0.108	8.3%	P = 0.976
	Non-Pred	6.6%	0.0413 to 1.014	7.9%	0.328, 3.155
GAHS<9	Pred	3.3%	P = 0.175	11.8%	P = 0.789
	Non-Pred	5.9%	0.240 to 1.276	11.2%	0.634, 1.824

Pred: Prednisolone treatment; Non-Pred: Not treated with Prednisolone; 95%CI: 95% Confidence Interval

Table 4:

A) Mortality in all Patients with High Static Scores

		28-Day Outcome		90-Day Outcome	
		Mortality	**Significance; 95%CI**	**Mortality**	**Significance; 95%CI**
MELD≥25	Pred	24.8%	P=0.330	42.2%	P=0.71
	Non-Pred	29.1%	0.828 to 1.762	39.2%	0.694 to 1.282
ABIC≥6.71	Pred	15.9%	P=0.067	29.9%	P=0.894
	Non-Pred	20.7%	0.980 to 1.829	29.3%	0.796 to 1.298
GAHS≥9	Pred	24.2%	P=0.180	40.2%	P=0.960
	Non-Pred	28.8%	0.895 to 1.774	38.8%	0.762 to 1.331

B) Mortality in Patients with High Static Scores (excluding patients presenting with GIB or Sepsis)

		28-Day Outcome		90-Day Outcome	
		Mortality	**Significance; 95%CI**	**Mortality**	**Significance; 95%CI**
MELD≥25	Pred	22.2%	P = 0.130	41.3%	P = 0.970
	Non-Pred	28.9%	0.903, 2.150	39.6%	0.714, 1.418
ABIC≥6.71	Pred	14.6%	P = 0.021	29.6%	P=0.698
	Non-Pred	21.0%	1.063, 2.131	29.4%	0.805, 1.383
GAHS≥9	Pred	21.0%	P = 0.039	38.5%	P = 0.640
	Non-Pred	29.3%	1.022, 2.242	38.4%	0.785, 1.482

Pred: Prednisolone treatment; Non-Pred: Not treated with Prednisolone; 95%CI: 95% Confidence Interval

Table 5: Effect of Prednisolone upon Day 7 bilirubin and Day 7 scores of Response.

		%Change in Bilirubin (95%CI)	**Proportion with Favourable Day 7 Score**		
			Lille <0.45	%ΔBili ≥25%	Change in 'static' score*
GAHS>8	No Prednisolone	-1.6% (-6.4, 3.2)	27.5%	20.8%	37.0%
	Prednisolone	-17.5% (-22.4, -12.6)	45.4%	43.3%	41.6%
ABIC≥6.7	No Prednisolone	-4.0% (-8.2, -0.2)	40.4%	25.0%	42.5%
	Prednisolone	-20.4% (-24.1, -16.7)	55.4%	49.8%	68.8%
MELD≥25	No Prednisolone	-0.2 (-5.6, 5.2)	29.1%	18.3%	34.8%
	Prednisolone	-12.0 (-17.5, -6.6)	48.7%	33.6%	59.2%

*refers to change in relevant 'static' score

Comparison of Prednisolone treated and untreated patients:

Difference in percentage change in bilirubin:

GAHS>8: $p<0.0001$ (95%CI 9.07, 22.81)

ABIC>6.7: $p=0.001$ (95%CI 22.06, 10.79)

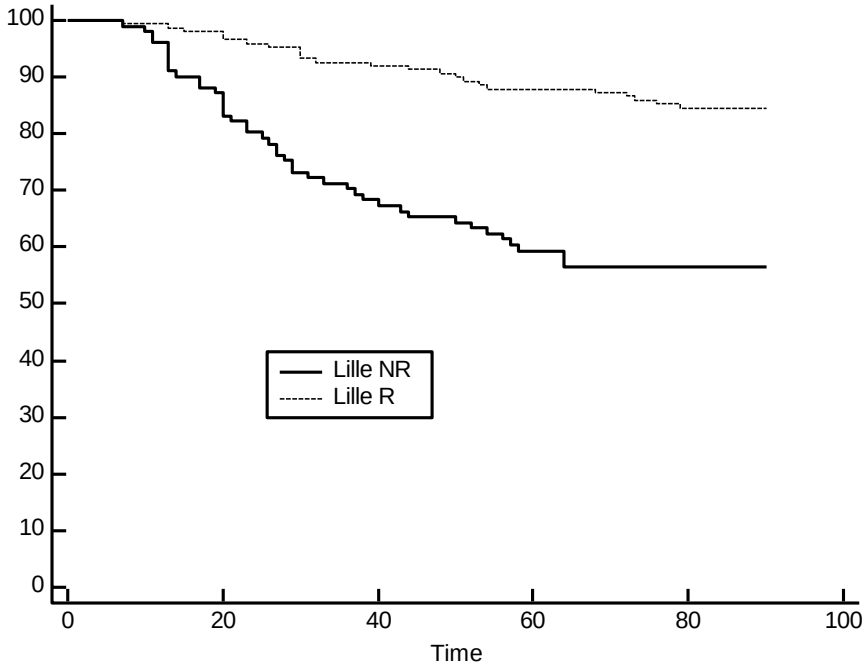
MELD>25; $p=0.0026$ (95%CI 4.14, 19.5)

All differences in proportions with a favourable Day 7 score significant ($p<0.05$) except the ΔGAHS at Day-7 for GAHS>8.

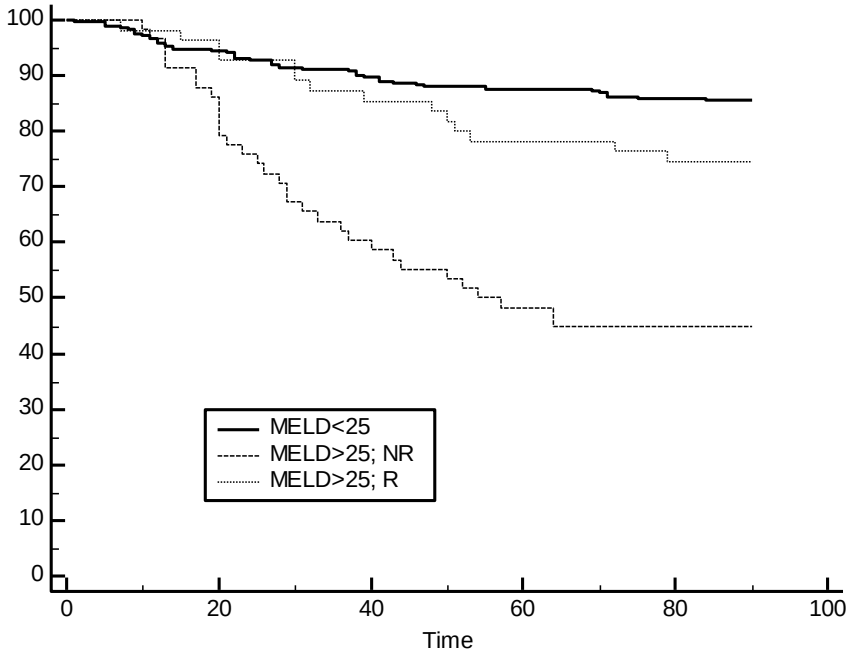
Figure 2: Kaplan-Meier Survival Probability for Patients Stratified by 'Static' Scores (only high baseline scores treated with prednisolone) combined with 'Dynamic' Scores (excluding initial presentation with GIB or sepsis). R: Responder; NR: Non-responder

A) DF/Lille; B) MELD/Lille; C) ABIC/ Lille; D) GAHS/Lille; E) GAHS/ΔGAHS

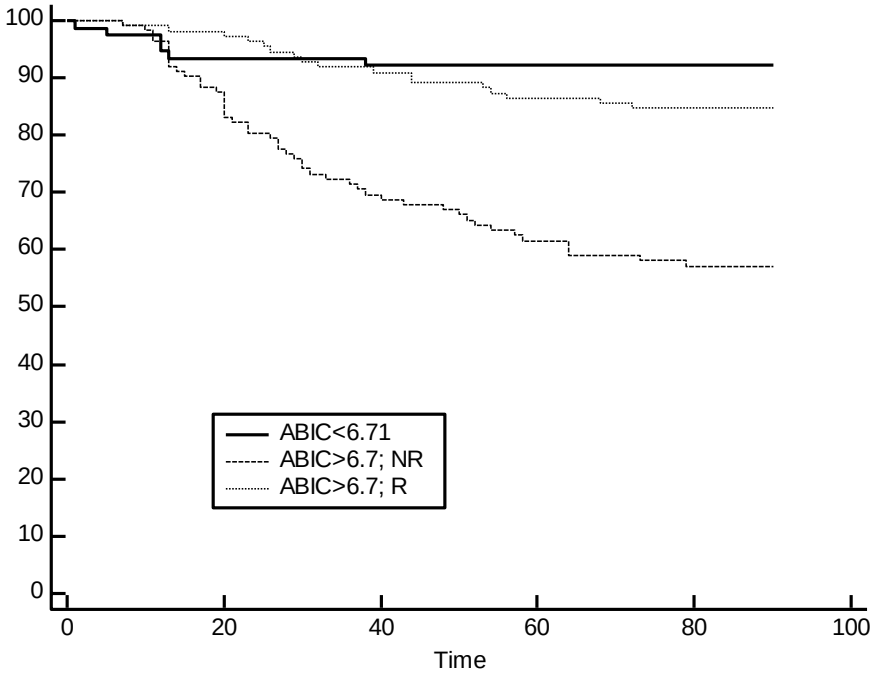
A) DF/Lille



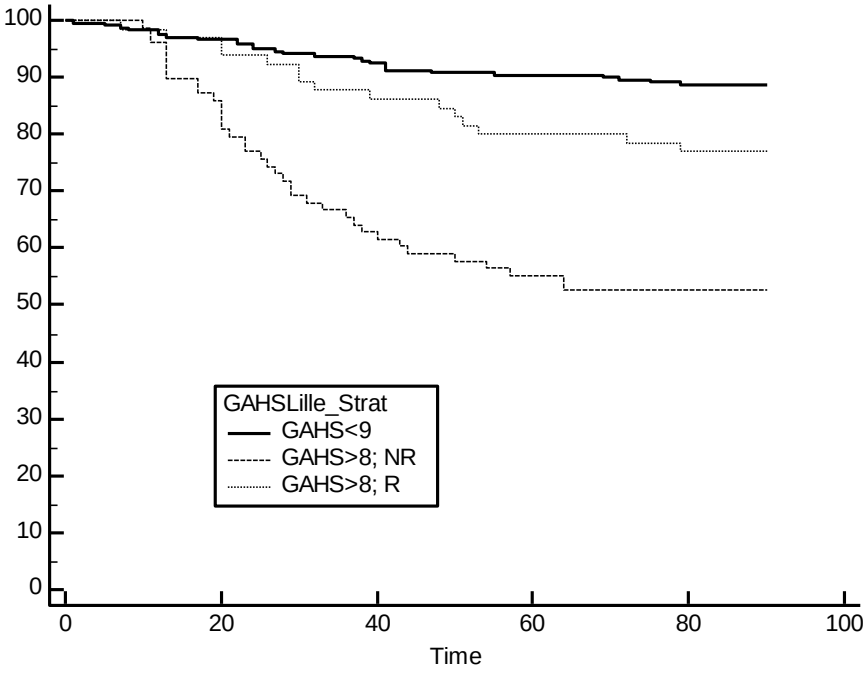
B) MELD/Lille



C) ABIC/Lille



D) GAHS/Lille



E) GAHS/ΔGAHS

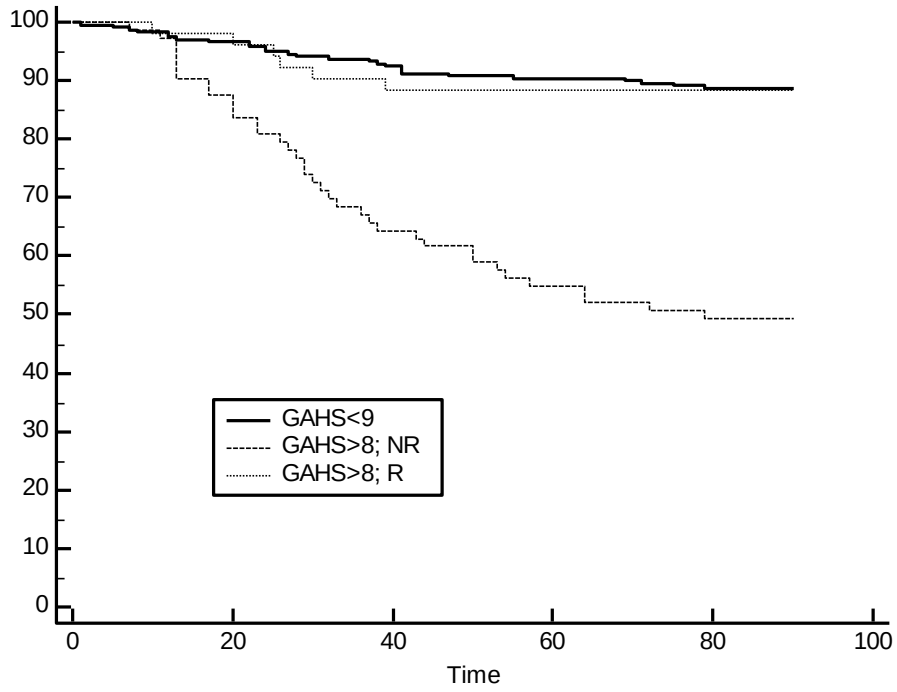

100
90
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
Time
GAHS<9
GAHS>8; NR
GAHS>8; R

Table 6: Application of Scores to Different Therapeutic Strategies and 90-Day Mortality. Prednisolone treatment offered to those not presenting primarily with either Gastrointestinal bleeding or sepsis initially.

	Low Score: No Prednisolone	High Score: Prednisolone Treated			TOTAL CUMULATIVE 90-DAY MORTALITY
	Mortality	Proportion	Responder Mortality	Non-Responder Mortality	
DF/Lille	-	100%	15.4%	43.5%	26.8%*
MELD/Lille	14.4%	39.0%	25.4%	55.2%	21.8%
MELD/%ΔBili			26.3%	48.0%	21.8%
MELD/ΔMELD			24.6%	66.7%	21.6%
ABIC/Lille	7.9%	85.7%	15.3%	42.9%	23.7%
ABIC/%ΔBili			17.1%	44.4%	23.8%
ABIC/ΔABIC			19.1%	54.7%	24.2%
GAHS/Lille	11.2%	47.0%	23.1%	47.4%	20.6%
GAHS/%ΔBili			18.5%	47.1%	20.3%
GAHS/ΔGAHS			11.5%	50.7%	19.2%*

*Difference in mortality: p=0.026: 95%CI 0.63% to 14.72%

Supplementary Figure 2: Flowchart showing the Stratification of Patients with Alcoholic Hepatitis in the STOPAH Trial.