Reviewer’s report

Title: Understanding fatigue during hepatitis C virus treatment: results of self-reported fatigue severity in two Phase IIb studies of simeprevir treatment in patients with hepatitis C virus genotype 1 infection

Version: 2 Date: 21 February 2014

Reviewer: Antoine Regnault

Reviewer’s report:

Overall comment

This paper describes the investigation of the benefit of the addition of simeprevir to peginterferon/ribavirin in terms of self-reported fatigue and health status in two phase II clinical trials (one in treatment-naïve patients and one in patients with null response, partial response of relapse to prior PegIFN/RBV treatment). The addition of simeprevir is not only showed not to worsen fatigue during the treatment period but also to allow patients to attain lower levels of fatigue quicker. This appears to be fairly clearly linked to the ability of this combined regimen to allow shorter treatment by the Response-Guided Therapy (RGT) approach.

Overall, the methods used are appropriate and the data support the conclusion made. Nonetheless I think the manuscript could be strengthened by some revisions.

Major compulsory revisions

1. The results presented in the paper present a very consistent pattern that supports the hypothesis that RGT is a key driver to the earlier alleviation of patient-reported fatigue and improvement in patient health status in the Simeprevir arm: both FSS and EQ-5D scores showed a clear difference between treatment arms from week 24 to 48 in the PILLAR study which is not observed in the ASPIRE study (in which no RGT was done). This key element would probably deserve more emphasis in the discussion of the paper. In addition, it may have been interesting to investigate further the relationships between PROs and treatment cessation according to RGT, for example in examining the self-reported fatigue and health status of the subgroup of patient who did not stop therapy at week 24 in the PILLAR study.

2. While the primary objective targeted by the authors (as reflected in the title of the paper) was related to fatigue, the EQ-5D data have a fairly important part in the discussion section. This perfectly makes sense as these data provide an interesting complement to the FSS data. However, the results of the EQ-5D actually presented in the paper are only partial. I suggest adding a table presenting the EQ-5D results to the paper. For instance, the authors could present the percentage of patients presenting some problems for each of the 5 domains at each visit, in each treatment arm of each trial.
3. An important question that is raised by the design and results of the trials (in particular PILLAR) is that of missing assessments. This question is composed of at least two distinct aspects. The first one is the discontinuation during the trial, the second is the missing baseline assessment of a subset of patient for whom the language version of the FSS was not available at the initiation of the study. In the discussion the authors state “the sensitivity analyses conducted found no evidence to suggest that the missing FSS data biased the results”. Now, the authors only mentioned some sensitivity analyses with LOCF. While this provides some elements to support the limited impact of discontinuation on the results, this does not address the question of the impact of having a substantial subgroup of patients that was recruited but not included in these analyses because of missing baseline assessment. This question probably deserves more than an endnote: it would be interesting to know whether this impacted the randomization ratios, whether the patients who did not have the FSS at baseline were comparable to those who had it in terms of baseline characteristics. The discussion could also be a little more detailed regarding this question.

Minor essential revisions

1. The authors present the results from two trials: ASPIRE and PILLAR. All along the paper, they present in parallel information on the 2 trials (in the material section, in the result section and in the discussion section). To help the reader, it is important that the order in which the information for each trial is presented is consistent (i.e. PILLAR consistently before ASPIRE or ASPIRE consistently before PILLAR), which is not the case in the current version. Also, only PILLAR is presented in the abstract; information about ASPIRE should be added.

2. The authors use the EQ-5D index, which can be obtained using different value sets exist to obtain this index. In the method section about the EQ-5D, the authors should describe which value set they actually used.

3. In the method section about the EQ-5D results, the authors present the results of the EQ-5D individual domain in different ways. They start presenting the percentage of patients with extreme problems, then they switch to patients “with problems” (I assume that this means at least some problems), then they present the proportion of patients with no problem. This may be confusing for the reader, hence presenting these results in a consistent way would probably be better.

4. In the results section on EQ-5D VAS results, the author discuss the fact that the slightly better VAS results in the PR arm of the ASPIRE study may be due to treatment failure in this group. This discussion point may be relevant for the interpretation of the other results of the ASPIRE trial and may be further developed in the article (in particular in the discussion section).

5. In the discussion the authors states that “The findings of the current study suggest that the FSS offers a reliable assessment of self-reported fatigue among patients with chronic HCV infection”. This claim is not supported by the data presented in this paper. I imagine that the authors refer to another set of (psychometric) analyses performed with these data. They should probably have this clearer and refer to another communication that supports this claim (reference 25??).
6. The FSS is central in this paper. Several questions related to the instrument arise and addressing them could be improved in the paper.

a. The presentation of the FSS in the methods section could be enhanced. Information on the instrument are provided in different sections of the manuscript (PRO in methods, FSS in methods, but also results and discussion). The manuscript would be improved if the information relevant to the FSS is gathered in a single section. For example it is not stated in this section that the instrument was originally developed and validated in other condition than Hepatitis C.

b. The normative values (which were apparently developed in the context of multiple sclerosis and Systemic lupus erythematosus) are given in the result section, while these may be interesting in particular to consider the level of fatigue at baseline, at EOT and EOF.

c. The authors mentioned 2 different recall periods for the FSS (one week for ASPIRE, 14 days for PILLAR). The potential impact of this difference may be relevant to discuss in the paper.

d. The authors have applied an interesting approach to inform the meaningfulness of a change in FSS score. They may have completed their analyses by a responder analysis based on the obtained values.

Discretionary revisions

1. In the introduction, the last sentence of the second paragraph “Furthermore…treatment-induced anemia” is not perfectly clear to me and should probably be reworded.

2. In the methods section, two sentences may be somewhat contradictory: “Trial-site staff reviewed the questionnaires after completion to ensure that all questions were completed. If required, patients were asked to complete any missing items” and “The 9-item responses in the FSS were combined into one total score per time point by calculating the mean of all non-missing items”. The later seems irrelevant since there should not be any missing item according to the process. In addition, the process poses the question of the bias in patient responses that may be induced by the fact that the responses were checked by the medical staff. The author may wish to discuss this point.

3. The AUC analysis is only briefly presented and the actual results are not presented. I would suggest to present the results more comprehensively.

4. The first sentence of the discussion states that the authors evaluated “patient-reported fatigue”, I suggest to add “and health status”.

5. The last sentence of the first paragraph of the discussion (“Of note…”) could be deleted as it conveys the exact same message than a sentence of the same paragraph a few lines above (“The findings in both treatment-naïve and treatment-experience patients …”).

6. In table 1, the number of decimal is not always consistent for a same data (Ex: mean baseline VAS in the PR group of PILLAR has 2 decimals while other mean baseline VAS only have one). This should be homogenized.

7. Redundant information is found in table 1 and 3 (description of BL FSS score).
Having it only in one of the 2 tables would be enough

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**
I declare that I have no competing interests