PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Impact of patient characteristics on efficacy and safety of once-weekly semaglutide versus dulaglutide: SUSTAIN 7 post hoc analyses
AUTHORS	Pratley, Richard E; Aroda, Vanita R; Catarig, Andrei-Mircea; Lingvay, Ildiko; Lüdemann, Jörg; Yildirim, Emre; Viljoen, Adie

VERSION 1 – REVIEW

REVIEWER	Young Min Cho Seoul National University College of Medicine I received grants from Sanofi, AstraZeneca, and LG, and consulting fees from Hanmi.
REVIEW RETURNED	24-Mar-2020

GENERAL COMMENTS	This manuscript is a post hoc analysis of SUSTAIN 7 study showing statistical superiority of once-weekly semaglutide vs once-weekly dulaglutide in HbA1c reduction and body weight reduction in people with type 2 diabetes across spectrum of key clinical characteristics (e.g. age, sex, diabetes duration, HbA1c and body mass index (BMI)). The analysis is comprehensive and the paper is well written. However, I have some concern about the conclusions in Line 425. “Understanding the impact of heterogeneity in clinical characteristics on the treatment differences between GLP-1RAs further supports patient-centred decision-making in clinical practice.” And Line 434 “This indicates that the efficacy of semaglutide vs dulaglutide is retained across a range of diverse clinical characteristics, thereby increasing the evidence base available to clinicians to guide care.” It sounds like recommending OW semaglutide over OW dulaglutide in major clinical situations related to age, sex, diabetes duration, HbA1c and BMI. Clinical outcomes such as cardiovascular events need to be considered when selecting one GLP-1RA over another. Both semaglutide and dulaglutide effectively improved such clinical outcomes. I think the conclusions need to be a little bit toned down or modified.

REVIEWER	Steve Kanters School of Population and Public Health, University of British Columbia, Vancouver, Canada
REVIEW RETURNED	05-Apr-2020

GENERAL COMMENTS	The authors have presented and summarized the results of a post hoc analysis of the SUSTAIN 7 trial that helps provide further insights into the robustness of the trial’s results. Namely, the study found that semaglutide was more efficient than dulaglutide with respect to changes in HBa1c and body weight over a variety of sub-

	groups. While comparisons across and within subgroups have been conducted in other trials involving semaglutide and dulaglutide individually, this is arguably the more important comparison to consider within subgroups, so its results are certainly of interest to clinicians and others. The authors should also be commended for their use of multiple imputation, which helps conserve randomization and reduce risk of bias. I do not have any major revisions, but do have a series of minor revisions and some comments 1. [Abstract] In the results, it may be useful to add the word statistical, as it is not clear at first read that greater reductions are only those that are statistically significant. 2. [Introduction] The references for the SUSTAIN programme are SUSTAIN 1-5. Should that not also include a reference to other SUSTAIN trials or to a reference for the programme as a whole? Methods 3. On line 166, the abbreviation ANCOVA should be introduced as the term is used in figures, yet never introduced. This would be the place to do so. 4. The first paragraph of the statistical analyses could benefit from some references, so that the curious reader could more easily look at the employed methods in more detail. 5. What statistical software was used to conduct the analyses? Please also specify specific packages if these were used. Results 6. The figures all seem to be of low DPI. This may be in the manner it was provided to us for revision. But please ensure that figures are of sufficient sharpness. Again, please ignore if this has already been addressed. 7. Personally, I think it would be more useful to have all the baseline tables in the supplementary materials and move Figure 1-3 into the main text. Otherwise, perhaps reducing to a single baseline characteristics table and 1 or 2 of these figures would also be helpful. 8. Lines 223-274: It's a bit confusing that the figures have p-values

	for the comparisons between treatment groups, but the results are describing differences between the groups. How are differences being determined? Visually or with a p-value? If there are p-values, some of them or a description of them (e.g., all <0.0001), would be useful to include 9. It would be useful to add the unit of measurement on the X-axis of Figure 1. Although, I admit that the unit is mentioned elsewhere, so this is not a necessary change. 10. As mentioned above, I'm quite happy with the use of multiple imputations. Some comments around this might be useful in the supplementary materials, with a small reference in the results. First, how many missing values were there? It would be informative to understand whether there are many or few missing values. Second, were the analyses conducted without MI and how did the results compare? I think some reporting of the missing values is required. Discussion 11. Use of multiple imputation to help conserve randomization could be considered a strength 12. May want to re-word the section on small sample sizes. First, adding an explicit statement about the consequence of small sample size and second rephrasing that this was exacerbated in the case of age group. [all optional] The following copy-editing should also be done (very minor) • Please be consistent in the use of dashes, such as over-weight and over weight. • The "as previously reported" on line 114 seems redundant. • The two sentences from line 147 to 150 should be combined as it currently reads as if weight loss response proportions are perhaps not predefined targets. • Colons should be used instead of commas in the table legends for abbreviations.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1		
1	This manuscript is a post hoc analysis of SUSTAIN 7 studyshowing statistical superiority of once-weekly semaglutide vs once-weekly dulaglutide in HbA_{1c} reduction andbody weight reduction in people with type 2 diabetes across spectrum of key clinical characteristics (e.g. age,sex, diabetes duration, HbA_{1c} and body mass index (BMI)). The analysis is comprehensive and the paper iswell written. However, I have some concern about the conclusions in Line 425. "Understanding the impact of heterogeneity in clinical characteristics on the treatment differences between GLP-1RAs further supports patient-centred decision-making in clinical practice." And Line 434 "This indicates that the efficacy of semaglutide vs dulaglutide is retained across a range of diverse clinical characteristics, thereby increasing the evidence base available to clinicians to guide care." It sounds like recommending OW semaglutide over OWdulaglutide in major clinical situations related to age, sex, diabetes duration, HbA_{1c} and BMI. Clinicaloutcomes such as cardiovascular events need to be considered when selecting one GLP-1RA over another. Bothsemaglutide and dulaglutide effectively improved such clinical outcomes. I think the conclusions need to be alittle bit toned down or modified.	The sentence from Line 425 (last paragraph of the Discussion) has been deleted (see strikeout of text below), and the importance of understanding the impact of patient characteristics has been referred to in more general terms at the beginning of the paragraph (see below in bold): Understanding the impact of heterogeneity in patient characteristics on treatment effects is important for clinical practice.This analysis provides insight on the influence of five of the most common and relevant patient-level factors from a clinical perspective and highlights semaglutide as an effective choice across these patient subgroups that are commonly encountered in clinical practice. ~~Understanding the impact of heterogeneity in clinical characteristics on the treatment differences between GLP-1RAs further supports patient-centred decision-making in clinical practice.~~ Line 434 (last sentence of the Conclusion) has been replaced, as shown below: Semaglutide was associated with superior efficacy to dulaglutide across various clinically relevant patient subgroups that are commonly encountered in clinical practice, with a safety profile similar to other GLP-1RAs and in line with previously published data for semaglutide. The treatment effect for semaglutide versus dulaglutide did not appear to be influenced by age, sex, diabetes duration, HbA_{1c} or BMI at baseline. Together with results from other studies and from experience in clinical practice, these findings support the efficacy of semaglutide across the continuum of care in a heterogeneous T2D population. ~~This indicates that the efficacy of semaglutide vs dulaglutide is retained across a range of diverse clinical characteristics, thereby increasing the evidence base available to clinicians to guide care.~~

Reviewer 2		
1	The authors have presented and summarized the results of apost hoc analysis of the SUSTAIN 7 trial that helps provide further insights into the robustness of the trial'sresults. Namely, the study found that semaglutide was more efficient than dulaglutide with respect to changes inHBa1c and body weight over a variety of sub-groups. While comparisons across and within subgroups havebeen conducted in other trials involving semaglutide and dulaglutide individually, this is arguably the moreimportant comparison to consider within subgroups, so its results are certainly of interest to clinicians andothers. The authors should also be commended for their use of multiple imputation, which helps conserverandomization and reduce risk of bias.I do not have any major revisions, but do have a series of minor revisions and some comments. [Abstract] In the results, it may be useful to add the word statistical, as it is not clear at first read thatgreater reductions are only those that are statistically significant.	The text shown below in bold has been added to the Results section of the Abstract: HbA_{1c} and BW reductions (estimated treatment difference ranges: –0.22 to –0.70%-point; –1.76 to –3.84 kg) and proportion of subjects achieving HbA_{1c} targets and weight-loss responses were statistically significantly greater for the majority of comparisons ofsemaglutide versus dulaglutide within each subgroup category and, excepting glycaemic control within the low-dose comparison in HbA_{1c} subgroups, this was irrespective of subgroup or dose comparison. As we were already at the word count limit for the Abstract, we have had to make edits to accommodate the above text, please see the tracked changes in the manuscript.
2	[Introduction] The references for the SUSTAIN programme are SUSTAIN 1-5. Should that not alsoinclude a reference to other SUSTAIN trials or to a reference for the programme as a whole?	Only SUSTAIN 1–5 (and SUSTAIN 6 in the next sentence) are cited in the Introduction, as the text refers to the global phase 3a clinical trials. The other SUSTAIN trials were either national-level trials or phase 3b. We have, therefore, not made any changes to the text and we hope this is acceptable.
3	[Methods] On line 166, the abbreviation ANCOVA should be introduced as the term is used in figures, yet neverintroduced. This would be the place to do so.	This recommendation has been implemented.

4	[Methods] The first paragraph of the statistical analyses could benefit from some references, so that the curiousreader could more easily look at the employed methods in more detail.	At the end of the paragraph mentioned by Reviewer 2, the following reference has now been cited: Little RJA, Rubin DB. Statistical analysis with missing data. New York: John Wiley & Sons, 1987.
5	[Methods] What statistical software was used to conduct the analyses? Please also specify specific packages if thesewere used.	SAS version 9.4 was used. At the end of the Statistical Analyses section of the Methods we have added the following text: Analyses were conducted using SAS version 9.4.
6	[Results] The figures all seem to be of low DPI. This may be in the manner it was provided to us for revision. Butplease ensure that figures are of sufficient sharpness. Again, please ignore if this has already been addressed.	We have rechecked the separate, individual figure files and all are in line with the journal guidance of 300 dpi tifs, but please let us know if a different file type/format is needed.
7	[Results] Personally, I think it would be more useful to have all the baseline tables in the supplementary materialsand move Figure 1-3 into the main text. Otherwise, perhaps reducing to a single baseline characteristics tableand 1 or 2 of these figures would also be helpful.	We thank the reviewer for this suggestion. We have now moved Supplementary Figures 2 and 3 into the main manuscript, and these have become: • Figure 1. Proportion of subjects achieving HbA_{1c}targets by subgroup • Figure 2. Proportion of subjects achieving weight-loss responses by subgroup To make space for these figures, Tables 2 and 4 have been moved from the main manuscript into the supplement and have become: • Supplementary Table 5. Subject demographics and baseline characteristics by BMI • Supplementary Table 10. Adverse events by BMI

8	[Results] Lines 223-274: It's a bit confusing that the figures have p-values for the comparisons between treatment groups, but the results are describing differences between the groups. How are differences being determined? Visually or with a p-value? If there are p-values, some of them or a description of them (e.g., all <0.0001), would be useful to include	Two types of p-values are provided in the manuscript: 1. p-values for comparisons between the semaglutide and dulaglutide treatment groups (ETD for change from baseline in HbA_{1c} and body weight, and ORs for the proportion of subjects achieving HbA_{1c}targets and weight-loss responses; low- and high-dose comparisons) within a subgroup category. These are presented in the figures in the manuscript. 2. p-values for the interaction between treatment effect and subgroup categories, i.e. to assess the impact of each clinical characteristic on the treatment effect. These are described in the Results section titled 'Treatment–subgroup interaction effects' (Lines 275–285 in clean copy of the updated manuscript). In Lines 223–274, the trends highlighted refer to the effect of treatment (semaglutide or dulaglutide) across the categories of a subgroup. No statistical analyses were performed for these comparisons and, to clarify this, throughout this section we have stated 'numerically', for example: …the proportion of elderly versus non-elderly subjects achieving glycaemic targets and weight-loss response of ≥5% was consistently numerically higher with both semaglutide and dulaglutide…
9	[Results] It would be useful to add the unit of measurement on the X-axis of Figure 1. Although, I admit that theunit is mentioned elsewhere, so this is not a necessary change.	This update has been made. In addition to adding the units to the x-axis, we have added 'ETD' prior to stating the units.
10	[Results] As mentioned above, I'm quite happy with the use of multiple imputations. Some comments around thismight be useful in the supplementary materials, with a small reference in the results. First, how many missingvalues were there? It would be informative to understand whether there are many or few missing values.Second, were the analyses conducted without MI and how did the results compare? I think some reporting of the missing values is required.	In the efficacy section of the Results, we have added the following text: Missing observations in the efficacy analyses were mainly due to subjects who discontinued treatment or received rescue medication. At week 40, between 81% and 86% of subjects were on treatment without initiation of rescue medication in the four treatment arms (Supplementary Figure 1). In the supplement, we have now included a subject disposition figure. We can confirm that the analyses have not been conducted without imputation of missing data.

11	[Discussion; optional] Use of multiple imputation to help conserve randomization could be considered a strength.	We thank the reviewer for this suggestion and have added the bold text shown below to the Discussion: A strength of the present analysis is the inclusion of comparator data, which allows for a more robust analysis and direct comparison of the differences in efficacy and safety of semaglutide versus dulaglutide across the subgroups and subgroup categories, and also the use of multiple imputation that helps to conserve randomisation.
12	[Discussion; optional] May want to re-word the section on small sample sizes. First, adding an explicit statement about theconsequence of small sample size and second rephrasing that this was exacerbated in the case of age group.	The bold text shown below has been added to this paragraph in the Discussion: A further limitation is the relatively small number of subjects in each subgroup category, which means that the findings should be interpreted with caution. Additionally, in the age subgroups, there was an imbalance in subject numbers (elderly versus non-elderly), with relatively few patients in the elderly subgroup (260; 22% of the analysis population). However, given the overall consistency of the age-subgroup analyses, as well as the general limitations of these post hoc analyses, the difference in subject numbers between the age subgroup categories seemed to have had little or no impact. Furthermore, elderly subjects in previous pooled analyses of the SUSTAIN 1–5 (26) and AWARD (30,32) trials have demonstrated similar efficacy and safety, supporting the results obtained here.
13	Please be consistent in the use of dashes, such as over-weight and over weight.	This recommendation has been implemented.
14	The “as previously reported” on line 114 seems redundant.	This text has been deleted and the sentence now reads: Semaglutide was administered subcutaneously via a prefilled injection device at one of two maintenance dose levels (0.5 mg or 1.0 mg OW), following a fixed-dose escalation regimen.
15	The two sentences from line 147 to 150 should be combined as it currently reads as if weight loss responseproportions are perhaps not predefined targets.	These sentences have been combined and the resulting sentence now reads: Predefined HbA_{1c} treatment targets (proportion of subjects achieving HbA_{1c} targets of <7% [53 mmol/mol] and ≤6.5% [48 mmol/mol]) and weight-loss responses (proportion of subjects achieving ≥5% and ≥10% weight loss) were also assessed.
16	Colons should be used instead of commas in the table legends for abbreviations.	This recommendation has been implemented and, for consistency, it has also been applied to the abbreviations in the footnotes to the figures.

VERSION 2 – REVIEW

REVIEWER	Young Min Cho Seoul National University College of Medicine, Seoul, Korea I received grants from Sanofi, AstraZeneca and LG; and consulting fees from Hanmi.
REVIEW RETURNED	06-Jul-2020

GENERAL COMMENTS	I concerns have been well addressed in this revised manuscript.

REVIEWER	Steve Kanters University of British Columbia Canada
REVIEW RETURNED	10-Jul-2020

GENERAL COMMENTS	I am very satisfied with all the changes and responses made by the authors and believe that this is ready for publication