PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Going to extremes: the Goldilocks/Lagom principle and data distribution
AUTHORS	Leese, Henry; Sathyapalan, Thozhukat; Allgar, Victoria; Brison, Daniel; Sturmey, Roger

VERSION 1 - REVIEW

REVIEWER	Adam Watkins University of Nottingham Division of Child Health, Obstetrics and Gynaecology School of Medicine Queens Medical Centre Nottingham UK NG7 2UH
REVIEW RETURNED	28-Mar-2019

GENERAL COMMENTS	The article by Leese et al., explores the concept that many biological data sets are presented incorrectly within the literature and therefore may present an inaccurate picture or analysis of the data. The authors propose that for a wider range of data there is a variability in which the distribution of the data may be both above and below the 'average' of a control group. As such, simply reducing any one data set to a mean +/- and error may both fail to show this and prevent the authors/readers from appreciating the natural spread in the data. The scope and nature of this article is warranted and will be of interest to a wide range of scientists dealing with the analysis of data sets, both large and small. The article is well written and appropriately referenced. As such, I only have minor comments to make. In line 26 of the article, the authors write about a 'range of expression' suggesting that the Goldilocks effect may only be relevant to analyse of gene/transcript expression. I am sure that the authors did not mean to suggest that such effects will not extend beyond the RNA level and so may wish to reconsider the use of the word 'expression'. I feel that lines 24-28 represent a very long sentence. The authors may wish to consider breaking up this sentence to make it easier to read? Apologies if I have missed what the '(see a*)' refers to in line 86, but it is not immediately evident to what this refers.

	In line 90, it may be advisable to state that the HbA1c refers to glycated haemoglobin which is related to the levels of glucose in the blood. As it currently stands it reads as though haemoglobin is a measure of glucose control. Line 95, did the authors mean that the target was to lower average HbA1c levels by 7% from existing levels? If not, then in relation to what measure was the study trying to achieve a level of 7%? Similarly, in line 97, would it make more sense to change the word 'low' to 'lower'? In line 103-105, could the authors indicate which 'further studies' they are refereeing to as non are cited. Also, is it possible to give values to what constitutes a 'high' and 'low' level of HbA1c or is it just any value higher or lower that then mean? Line 110 should be reworded as it current suggests that there are extremes of mortality. In line 158, I am not sure 'physiologically' is quite the correct word as all biological processes will proceed along a physiological pathway. Would something like 'proceeds optimally' be a more accurate phrase? I will happily leave it to the authors to choose their preferred phrasing. In the legend to Figure 1 it feels like the text is more akin to that within the body of the article rather than a legend to a figure. I am not sure they need the words '…is described in Figure 1.' In line 295. Also, I am not sure how the text below line 298 'explains how Figure 1 illustrates…' The authors may wish to review the text in the Legend to see if it can be made more succinct.

VERSION 1 – AUTHOR RESPONSE

In line 26 of the article, the authors write about a 'range of expression' suggesting that the Goldilocks effect may only be relevant to analyse of gene/transcript expression. I am sure that the authors did not mean to suggest that such effects will not extend beyond the RNA level and so may wish to reconsider the use of the word 'expression'.

We have reworded the sentence for clarity.

I feel that lines 24-28 represent a very long sentence. The authors may wish to consider breaking up this sentence to make it easier to read?

We have edited the sentence for clarity.

Apologies if I have missed what the '(see a*)' refers to in line 86, but it is not immediately evident to what this refers.

This has been deleted as it was a legacy from a earlier draft.

In line 90, it may be advisable to state that the HbA1c refers to glycated haemoglobin which is related to the levels of glucose in the blood. As it currently stands it reads as though haemoglobin is a measure of glucose control.

This has been reworded for clarity.

Line 95, did the authors mean that the target was to lower average HbA1c levels by 7% from existing levels? If not, then in relation to what measure was the study trying to achieve a level of 7%? Similarly, in line 97, would it make more sense to change the word 'low' to 'lower'?

We have rephrased this to avoid confusion.

In line 103-105, could the authors indicate which 'further studies' they are refereeing to as non are cited. Also, is it possible to give values to what constitutes a 'high' and 'low' level of HbA1c or is it just any value higher or lower that then mean?

This has been clarified and supporting studies have been included.

Line 110 should be reworded as it current suggests that there are extremes of mortality.

This has been reworded.

In line 158, I am not sure 'physiologically' is quite the correct word as all biological processes will proceed along a physiological pathway. Would something like 'proceeds optimally' be a more accurate phrase? I will happily leave it to the authors to choose their preferred phrasing.

We have rephrased this sentence.

In the legend to Figure 1 it feels like the text is more akin to that within the body of the article rather than a legend to a figure. I am not sure they need the words '…is described in Figure 1.' In line 295. Also, I am not sure how the text below line 298 'explains how Figure 1 illustrates…' The authors may wish to review the text in the Legend to see if it can be made more succinct.

We have included the text in an appropriate context within the main manuscript.