Dear Editor,

Thank you for providing us with an opportunity to revise our manuscript.

We also thank both reviewers for taking the time to look at our manuscript and provide helpful suggestions for its improvement. We have addressed all the points below (our responses are in blue text). Changes to the manuscript are shown in red text on the pdf.

Referee(s)' Comments to Author:  
Referee: 1  
  
Comments to the Author  
Process-based modeling of microbial community dynamics in the human colon.  
Helen Kettle, Petra Louis, Harry J. Flint  
Royal Society Interface  
  
An existing model of microbial communities, previously published by H.K., P.L., and H.F., was adapted for use in simulating the microbial communities of the human proximal, transverse, and distal colon. The model consists of a system of ODEs describing the dynamics of 10 “functional groups” of bacteria (rather than specific taxa), four growth substrates (again using broad categories), and 10 metabolites. The model included simulation of inflow of growth substrates both continuously and in discrete “meals”, as well as outflow as “bowel movements” and absorption by the host. The model makes use of parameters determined in earlier work, and was verified by comparison to a set of 7 observable criteria. Finally, the model was used to simulate two experimental scenarios, with the results compared to two previously published studies. The simulations produced qualitative agreement with the data.  
  
Modeling the metabolic impact of the human colon microbiome is an important goal that can lead to improved treatments of various disorders, and so this work, which presents a useful model, is valuable. The choice to coarse-grain the microbiome into functional groups is an interesting one. This choice provides an important reduction in the possible complexity of the model, but may also make it difficult to understand differences in microbiomes across individuals, unless the presence/absence of an entire functional group is discovered.  
  
Major Comments:  
  
(1) The exact model used is not clearly presented. It is reasonable to leave out the full equations because they were previously published, but it would be helpful to at least explicitly list the state variables of the model in one place (I think I managed to list them by hunting around between table 1 and the Materials & methods section). In general, the paper is presented in such a way as to move the focus away from the methods and onto the results. While this is common in biological papers that are presenting novel results, the purpose of this paper is to present the method itself, and so this organization seemed confusing to me.

A table of the state variables has now been added (Table 2 in revised manuscript) and is referred to in the text at L73: “A graphical summary of the model is shown in Fig. 2 the microbial functional groups are shown in Table 1 and the model state variables are summarised in Table 2. “

As the reviewer points out the organisation of the paper is aimed towards biologists rather than modellers. This was chosen as the authors have written 2 preceding papers outlining the majority of the methods used here and felt that this was more an application of the model. Our overall aim is to produce a modelling tool that we hope biologists would use (and indeed, many biologists are using the R package microPop), rather than something that would always remain in the modelling arena, hence the choice of format with the Results before the Methods.

(2) The use of previously determined parameters makes the model user-friendly, but severely limits its adaptability. In fact, one way to distinguish microbiomes while keeping only 10 functional groups may be to inspect differences in these parameters. The authors comment that the user of the model may adjust the parameters, but a discussion of how the parameters can be found would be a helpful addition.

This is a good point, thank you. We have referred again to the getting started guide included in the Supp Info in the Methods section (L380) “Furthermore instructions on how to use the package are given in the supplementary file ‘gettingStartedWithMicroPopGut.pdf’”

and added the following text in the Methods section on L392:

“Although this application uses the microbial parameters (e.g. maximum growth rates, yields etc) that are in the package's intrinsic data frames, these can be easily changed by either modifying the dataframe in R or by providing a new dataframe - either as an input csv file or by creating one in R. One of the input arguments to the function microPopGut() is microbeNames which allows the user to also enter other microbial groups.”

And this to the Discussion (L311): “As well as adjusting the parameters for each group to represent inter-individual variation, groups can also be easily added or removed from the model through the input argument ‘microbeNames’.”

(3) The horizontal axis of the bar charts in figure 5 seem reversed from those in figures 6 & 8. I think that is certainly the case for comparison between figure 6 and the data from the Duncan study. In the data from the Walker study, it isn’t so clear if the groups can be ordered according to RS fraction, in particular the nature of the weight loss diet is not described.

The reviewer is correct about the horizontal axes – we have altered the bar charts in Fig. 5 to be in the same direction as Figs. 6 and 8 . The RS frac for the Walker data was computed using RSI/CI. The weight loss diet is detailed in the Table in Fig 5.   
  
(4) The use of functional groups rather than individual taxa is interesting, and I would be interested in more discussion of the advantages and disadvantages of this approach.

This is touched on in the Discussion but we have also added the following text (L288):

“However, it would also be possible to define completely different groupings that relate to other outputs (e.g. bile acid metabolism, or vitamin/ micronutrient supply) in order to address specific questions.”

And this (L301):

“In future it should become possible to define the relative abundance of functional groups (MFGs) and their relationship to phylogeny directly from genomic and metagenome analysis, by examining genes diagnostic for particular pathways and functions (e.g. Reichardt et al, 2014). “

Minor Comments:  
(1) Line 107 states “Having established the default model settings and parameters values…” but later (line 293) it is stated that “the work presented here did not attempt to fit particular parameters to data…”. This is a little confusing

Agreed! We were referring to the parameters controlling the processes to do with the colon rather than the microbial growth parameters but this wasn’t clear. We have rephrased this to (L113):

“The model settings which give the best fit to our criteria are shown in Table 4 (colon parameters and dietary inflow). The microbial group parameters are listed in Supp. Info. (section 3). These define our default model. From this we investigate...”  
  
(2) Many of the figures include sub-captions or legends that are hard to read (notably figure 1 C’s caption can be moved  into main caption).

We have moved much of Fig1C text into the caption and we have enlarged the rest of the text in the figure   
  
(3) Line 54 - The first two sentences of this paragraph seems awkwardly written to me. Especially the first sentence would be clearer if split into two separate sentences, as the clauses seem to be not so related (and the second clause seems to be more strongly related to the next sentence).

This has been changed to (L54):  
“Since aproximately 95% of the SCFA produced by the microbes during growth are absorbed by the host through the gut wall this represents a strong interaction between the microbes and the host. Indeed the ratio of the 3 main SCFAs (acetate, butyrate and propionate) is known to have a significant effect on human health.”

Referee: 2  
  
Comments to the Author  
In this manuscript, Kettle and colleagues aim to further develop a model that explains host-microbiota interactions in the context of diet. While the methods of modeling and proposed scenarios appear sound, the main weakness and concern of this manuscript is that there is no testing or validation of the product, only educated guessing of proximate values to place into the model in an attempt to reproduce specific outcome values/ranges. Furthermore, an explicit applicability to human health is not defined (minor, but vital), and it is unclear how this model would advance such investigations when the needed input data is currently beyond practical reach of current technologies. These limitations should be addressed prior to publication, as outlined in the following major and minor comments:

The modelling approach we describe here was in fact designed to use the type of input data that are most commonly available from human studies – 1) measurements of metabolites and the gut microbial community based on faecal samples; 2) data on dietary intake in terms of fibre, total carbohydrate, protein, and fat. It is true that concentrations within the colon are currently very hard to measure and are hardly ever reported, but the value of our modelling is that it attempts to predict how these unknowns are likely to relate to the data that can be measured. In our view this has considerable relevance to human health as it suggests the ranges in colonic metabolite concentrations and bacterial populations that may correspond to measured faecal outputs.

*[We have made some changes in the text with the aim of explaining these points better, as detailed below]*  
Major:  
1. The sole use of sudden death autopsy data (Cummings et al., 1987) to approximate physiological parameters neglects more recent data captured by remote collection and transmission devices in vivo. One such example: Mikolajczyk et al 2015 - doi: 10.1038/ctg.2015.22. This is likely true for microbial composition and nutrient flux. Can this be integrated or rationalized in comparison to post-mortem data?

Thank you for this comment. We acknowledge that some data are available through telemetry and remote sensing (e.g. for colonic pH). These are valuable and are now given more prominence in the paper. Mikolajczyk et aL (2015) (now cited, L88) report intestinal pH estimates in 11 volunteers based on use of the SmartPill device. We also trialled this device with a few volunteers in one of our own studies although we found it impractical to use routinely. Earlier telemetry work by Bown et al is also referred to. These results are helpful, but do not alter the assumptions that we have made on colonic pH.

*[(nb. We used SmartPill ourselves but did not find it an easy or reliable tool – hard to swallow and useless for transit measurement because of its large size! Also, while it records pH, the location where this is measured requires some guesswork)]*

2. Line 131, the authors state “Fig. 1C shows the average pH and TSCFA for the proximal, transverse and distal compartments. It can be seen that blue (meals) and red (continuous inflow) dots show the same basic trends.” This seems accurate for the proximal and distal regions, but the predictions for the transverse region actually show opposite trends. This is not explained or rationalized.

Thank you to the reviewer to pointing this out as we also noticed there was a labelling error in Fig 1C – red and blue dots are the other way around. This has now been corrected in the Fig caption.

We do not know why the trend with transit time is different for continuous inflow in the transverse section or why the trend is different between the two inflow types. This is a complex system and TSCFA is affected by both SCFA and water absorption rates as well as microbial growth rates which will differ with pH and substrate availability. Both of which alter along the colon as the contents of the previous compartment flow into the next. Similarly the meals inflow produces a different pattern of growth due to its intermittent nature, allowing different groups to flourish so this can also affect SCFA production. Due to the speculative nature about what is happening here we have modified the text as follows:

*L 132-133 Deleted the sentence “It can be seen that ….”*

Rephrased the next sentence (L138) to read “A decrease in TSCFA (and concomitant increase in pH) with longer transit time is predicted in the proximal colon both for meal feeding and continuous input and this is in broad agreement with experimental findings (Lewis and Heaton 1997); ….”

3. Line 161, the authors state “Comparing our simulations with data from human volunteer experiments is not straightforward since in order to run our model, ingested food must be translated to substrates reaching the colon. This is problematic due to unknown water consumption and transit times and uncertainties associated with the absorption rates of the ingested carbohydrate and protein higher up the digestive tract. Thus we do not attempt to reproduce human experiments.” but then proceed to compare results from their model to human dietary intervention studies where the certainty of these factors is even more problematic than if a well-controlled, small confirmation cohort was completed by the authors. This highlights the major weakness of the manuscript overall – how is this modeling ultimately related to real-world scenarios? What is the ultimate utility to researchers, particularly if they would need to estimate the vast majority of these factors (see major comment 4)? There is no validation of the generated model using a test cohort. It appears that the authors manipulated variables such as transit time, incoming nutrients, water content, and flow to fit the expectations of potentially flawed data (see major comment 1), rather than adjusting the parameters or computations of the model to fit real-world outcomes (guessing and checking what the numerical values to put into the model would be to fit the goal outcomes, rather than putting real-world data into the model to test its accuracy and applicability to known values).

We should explain that our two dietary studies discussed here were highly unusual in that they involved complete control and monitoring of dietary intake (which was provided entirely by our own kitchens) by the participating subjects throughout the duration of the studies (8-10 weeks). This was combined with detailed chemical analysis of the diets themselves. Indeed, we would argue that these are some of the most carefully controlled studies of this type ever to have been conducted and provided us with the best opportunity to examine the applicability of the theoretical model. Such studies are extremely laborious, time consuming and expensive. In recent months we have however published an overall analysis of data from 10 such studies involving 163 human subjects conducted over a 10-year period (LaBouyer et al 2022 Gut Microbiome). This revealed some highly significant relationships between total SCFA concentrations and the proportion of different SCFA (notably butyrate) in faecal samples. The observed increase in % butyrate with increasing total [SCFA] is in fact predicted by our model, which provides an important demonstration of the ’real-world’ relevance and validity of the modelling approach taken here. This finding also has important implications in view of the potential benefits of butyrate supply for the health of the colonic mucosa. This paper has now been added to the reference list, and some new text with supporting references added to the Discussion on this point.

*We have added (L208):*

*“Analysis of 10 human studies involving 163 subjects has shown a highly significant increase in percentage butyrate with increasing total SCFA concentration in faecal samples (LaBouyer et al, 2022)”*

*We have altered the Discussion as follows (L355):*

*“To conclude, our model helps to explain some important but poorly understood relationships that have been reported in human studies. including the increase in butyrate proportion with increasing total faecal SCFA (LaBouyer et al (2022).* *This phenomenon has important implications in view of the claimed benefits of butyrate supply for colorectal cancer prevention and the health of the colonic mucosa (Louis et al Nat Rev Microbiol 12:661-672 (2014), Hamer HM et al APT 27:104-119 (2008)). The model also predicts increasing total faecal SCFA with greater fibre intake and more rapid gut transit. Gut transit…..”*

*We have also altered the text (L172) to read:*

*“Thus we do not attempt to reproduce human experiments exactly but rather we run simulations based on variations to our standard model set up which are qualitatively similar and then compare our results with the trends in the available data.”*

4. While the use of “substrates entering the colon” would arguably lead to more accurate estimates than “substrates entering the GI tract” (i.e. diet), wouldn’t inclusion of nutrients consumed be a more realistic, quantifiable, and applicable way to develop the model (especially given the authors’ comments about inability to accurately predict and quantify nutrients passing the ileocecal valve)? In the near-future it would seem that accurately captured dietary intake would be more readily available for integration into the model than sampling incoming colon contents in real-life scenarios for validation.

Yes we very much agree with this. This is undoubtedly the way forward but would require the inclusion of a model of the GI tract preceding the colon to establish which parts of the diet actually reach the microbes. This is beyond the scope of this paper but we have added the following to the Discussion (L351):

“A longer term goal would be to model the processes in the gastrointestinal tract preceding the colon in order to simulate how substrates entering the colon relate to dietary intake. This would allow more accurate prediction of microbial metabolite production based on diet.”

Minor:  
1. Line 99, the authors state “…we believe this model represents a significant step forward in this field.” What field? The general weakness of the introduction overall is that it does not provide a rationale or potential applications for why these types of modeling are needed to advance what the authors define broadly as “human health”, simply that previous work on such models can be improved.

We have replaced the admittedly somewhat vague “in this field” by “in analysing this highly complex system”. (L106)

*We have also added a sentence to better explain our rationale with regard to human studies and health (L103): “An important goal of our modelling is to aid and inform the interpretation of data obtained, mostly from faecal samples, in studies on diet and health in humans”. ]*

2. The following statement requires citations: Line 55 “…and it is the ratio of the 3 main SCFAs … which is known to have a significant effect on human health.”

*Citations Morrison & Preston (2016), Louis et al (2014) have been added.*