Heart-to-heart with ChatGPT: the impact of patients consulting AI for cardiovascular health advice

Anton Danholt Lautrup,1 Tobias Hyrup,1 Anna Schneider-Kamp,2 Marie Dahl,3 Jes Sanddal Lindholt,4 Peter Schneider-Kamp1 1

ABSTRACT

Objectives The advent of conversational artificial intelligence (AI) systems employing large language models such as ChatGPT has sparked public, professional and academic debates on the capabilities of such technologies. This mixed-methods study sets out to review and systematically explore the capabilities of ChatGPT to adequately provide health advice to patients when prompted regarding four topics from the field of cardiovascular diseases.

Methods As of 30 May 2023, 528 items on PubMed contained the term ChatGPT in their title and/or abstract, with 258 being classified as journal articles and included in our thematic state-of-the-art review. For the experimental part, we systematically developed and assessed 123 prompts across the four topics based on three classes of prompts formulated at higher and lower levels of health literacy generally yielded higher-quality answers. Counterintuitively, responses in a lower-resource language were often of higher quality.

Results The articles reviewed were fairly evenly distributed across discussing how ChatGPT could be used for medical publishing, in clinical practice and for education of medical personnel and/or patients. Quantitatively and qualitatively assessing the capability of ChatGPT on the 123 prompts demonstrated that, while the responses generally received above-average scores, they occupy a spectrum from the concise and correct via the absurd to what only can be described as hazardously incorrect and incomplete. Prompts formulated at higher levels of health literacy generally yielded higher-quality answers. Counterintuitively, responses in a lower-resource language were often of higher quality.

Conclusions The results emphasise the relationship between prompt and response quality and hint at potentially concerning futures in personalised medicine. The widespread use of large language models for health advice might amplify existing health inequalities and will increase the pressure on healthcare systems by providing easy access to many seemingly likely differential diagnoses and recommendations for seeing a doctor or even harmless ailments.

INTRODUCTION

The release of ChatGPT as a ‘research preview’ on 30 November 2022 has sparked a discussion on the opportunities and threats associated with current and upcoming technologies based on large language models (LLMs) for medical research, clinical practice and health communication and education.1,2 An as-yet limited body of research points LLMs as a potentially valuable tool for patients to access health advice,3-5 a scenario that likely already has become a reality as health inherently is a topic of great interest to all of us. Patients already use Google extensively for health advice,6 and the societal lockdowns during the COVID-19 pandemic have conditioned patients further to find advice online instead of visiting medical professionals.
The use of LLMs for health advice represents a double-edged sword. On the one hand, ChatGPT has been found to provide more informative responses than Google snippets in the context of patients and health professionals accessing information on cancer. ChatGPT has also been found to be good at reproducing consensus rather than fringe opinions, providing an opportunity to counteract the echo chamber effect of social media and other user-generated content. Comparing responses from ChatGPT to human responses, the former was perceived as more empathetic, opening avenues for more individualised, compassionate and scalable healthcare. On the other hand, ChatGPT has been found to provide dangerous advice in infection scenarios, showing limits by being unable to ask clarifying questions, raising questions regarding the quality of responses in safety-critical applications such as health advice.

In this article, we aim to provide an overview of the academic debate on the potential of LLMs and systematically and critically explore this potential for the case of patients searching for advice on cardiovascular conditions using ChatGPT. Cardiovascular diseases were chosen as they are the leading cause of death on a global scale, the academic debate already discusses the potential of LLMs to contribute to cardiovascular medicine, and initial explorations have investigated their use in educating citizens on cardiopulmonary resuscitation. We contribute to research on the use of large LLMs in medical fields by providing insight into how language, description depth and health literacy of the prompts impact the quality of the responses obtained. We also showcase a generic template for further systematic explorations and structured evaluations based on a 4C model for assessing correctness, comprehensiveness, conciseness and comprehensibility. Ultimately, we contribute to the wider academic debate on the opportunities and threats of employing LLMs in medical fields, pointing to mechanisms that might amplify existing health inequalities and increase pressure on healthcare systems.

**MATERIALS AND METHODS**

This mixed-methods study comprises two phases: (1) a systematic state-of-the-art review of the academic debate on ChatGPT and LLMs in health communication, education and research and (2) a systematic experiment with prompts and a multidimensional evaluation of the quality of the associated responses in the field of cardiovascular diseases.

**State-of-the-art review**

To keep the review of extant literature feasible in the face of the veritable explosion of publications related to ChatGPT (see figure 1), we considered all 528 items from the PubMed database as of 30 May 2023 where the title and/or abstract contained the search term ‘ChatGPT’, including all 258 journal articles and excluding 270 commentaries, editorials and other publication types.

We performed a thematic coding of the included articles, identifying the different potential and actual areas of application of LLMs discussed in different medical contexts.

**Systematic experimentation with ChatGPT**

Prior work has performed experiments using prompting and LLMs, but until now, few have performed systematic simulation experiments, often due to concerns of limited replicability and/or a lack of best practices to do so. Most works focus on a single chat session, or a handful of prompts that are run a manageable number of times and are then analysed and discussed by the authors, effectively providing some first impressions of the potential of LLMs in the chosen medical context. A few more extensive studies randomly sample real-world requests for health advice and corresponding human responses from online sources to compare human and AI-generated responses.

In our experiment, we define a latent space of prompts and systematically construct a variety of prompts in this space. We assess the responses by employing a hybrid approach where we first perform a qualitative evaluation, followed by a qualitative analysis. In particular, we score responses by how correct, concise, comprehensive and comprehensible they are and refer to these criteria as the 4Cs of LLM evaluation in medical fields (see table 3). Instead of running the same prompts multiple times, we opt for prompt diversity in each of the four topics outlined in table 1, to investigate whether and to what degree different ways of prompting about the same subject influence the results.

We choose four cardiovascular diseases that cover different degrees of urgency as our topics. Namely, for chats with the myocardial infarction (MI) topic, it is crucial to direct patients to emergency services if they or another real person in their immediate surroundings appear to be suffering from relevant acute symptoms. For peripheral arterial disease (PAD), critical symptoms should be recognised, but the risks and need for...
treatment in more mundane or chronic cases should not be exaggerated. Similarly, for varicose veins (VV) and cardiovascular prevention (CP), the LLM should also simply answer the questions of the patients as accurately as possible with proportional response in mind.

For each of these four topics, we systematically construct prompts from the latent prompt space distributed across two languages (English and Danish), two levels of description depth (low and high), and two levels of health literacy (low and high). As we do not combine low description depth with high health literacy, the space effectively comprises six classes of prompts presented in table 2. Extant work portrays the ability of conversational AI systems to meet patients on their level as one of their greatest potential strengths.5 9 The six combinations of language, language proficiency and health literacy allow us to investigate how well this strength is realised in practice.

After constructing at least five prompts for each combination of the six classes and four topics, we are left with the 123 prompts presented together with the responses in online supplemental materials. We ran all 123 prompts through the ChatGPT March 23 Version using the state-of-the-art GPT-4 model.15 Each prompt was entered into a new chat session with no repeats (ie, we did not rerun prompts to obtain alternative responses) or follow-ups (ie, we did not supply further prompts to clarify or extend responses). The responses were then subsequently scored on the 4Cs regarding correctness, conciseness, comprehensiveness and comprehensibility according to the criteria set forth in table 3. Reflections were noted down by the author team comprising experts in cardiovascular disease and nursing, machine learning, communication and patient perspectives. Correctness and comprehensiveness were scored by medical professionals while conciseness and comprehensibility were scored by communication and patient perspective experts. Each score was reviewed by at least two authors to ensure inter-rater reliability by consensus.

The main limitations of our research design are that we refrained from including patients in the assessment of prompt-response pairs and that we limited our systematic exploration to a selection of four cardiovascular topics. However, small-scale experimentation by the authors of this article suggests that a similar exploration of other common medical fields such as endocrinology and oncology is likely to yield comparable results.

### RESULTS

As of 30 May 2023, 528 items on PubMed contain the term ChatGPT in their title and/or abstract, with 258 being classified as journal articles and included in our thematic state-of-the-art review. The five main themes identified represent potential and actual application areas of LLMs and comprise clinical practice (73, clipra), scientific publishing (72, medpub), education of medical professionals (71, edupro), patient education (65, patedu) and medical research (39, medres), supplemented by minor themes of often discursive/philosophical nature (48). Many articles discussed more than one of these themes, with the distribution of themes to articles represented in figure 2.

Next in this section, we provide an overview of the experimental results. The full breakdown of every individual prompt-response pair is found in online supplemental materials, where we identify the pairs by topic (MI/PAD/VV/CP), health literacy (L/H for low/high), description depth (L/H), language (E/D for English/Danish), and a variant number (1-5). For example, PAD-LH-E2 identifies

<table>
<thead>
<tr>
<th>Topic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myocardial infarction</td>
<td>An acute and life-threatening situation mandating correct and comprehensive information, prospects of underplaying or overplaying risks and recommendations to seek immediate medical assistance.</td>
</tr>
<tr>
<td>Peripheral arterial disease</td>
<td>A potentially amputation-threatening chronic cardiovascular condition, slowly creeping symptoms from decreased activity radius to pain while lying, hard to diagnose and easy to overlook, prospects of overplaying risks.</td>
</tr>
<tr>
<td>Varicose veins</td>
<td>A typically harmless chronic cardiovascular condition, receiving a lot of questions in clinical practice, easy to diagnose but usually requires no treatment, prospects of overplaying recommendations to seek medical assistance.</td>
</tr>
<tr>
<td>Cardiovascular prevention</td>
<td>General and specific inquiries regarding how to prevent severe cardiovascular conditions such as heart attacks, atherosclerosis, etc., extensive topic with prospects of overloading patients with too much information.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Overview of the six classes of prompting that are employed across the four topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>Depth of description of prompt</td>
</tr>
<tr>
<td>English</td>
<td>Low</td>
</tr>
<tr>
<td>English</td>
<td>High</td>
</tr>
<tr>
<td>English</td>
<td>High</td>
</tr>
<tr>
<td>Danish</td>
<td>Low</td>
</tr>
<tr>
<td>Danish</td>
<td>High</td>
</tr>
<tr>
<td>Danish</td>
<td>High</td>
</tr>
<tr>
<td>Patients’ language, language proficiency and health literacy not only influence their ability to generate prompts and the responses obtained but also their ability to understand and interpret responses.</td>
<td></td>
</tr>
</tbody>
</table>
the second English, low-health literacy, high-description depth prompt variant for PAD.

Overall, the prompt responses score somewhere between three and four, with conciseness generally being the weakest of the 4Cs (see figure 3). This trend is somewhat consistent across various sortings and subdivisions of the data, though with some variance.

When sorting by topic (see table 4), it seems MI is among the more consistently correct and concise topics, while for comprehensiveness and comprehensibility, VV get better scores on average. CP generally seems to be doing the worst on all fronts, indicating that ChatGPT may be trying to weave together too many contradictory ideas or that scoring was in some way unintentionally biased.

When arranging the data by the different prompt classes (see table 5), we observe that, on average, the high-health literacy prompts gave more correct, comprehensive and

---

**Table 3** Overview of the 4Cs of large language model evaluation in medical fields detailing the scoring criteria that we subjected all responses generated by ChatGPT

<table>
<thead>
<tr>
<th></th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>Absolutely nothing wrong—this satisfies textbook standards.</td>
<td>Only few/minor simplifications to decomplexify and aid comprehension</td>
<td>Oversimplifications that affect the truth content of the answers</td>
<td>The answer contains serious mistakes but also several true statements</td>
<td>Nearly everything is wrong and not much makes any sense</td>
</tr>
<tr>
<td>Concise</td>
<td>Brief and to the point, with minimal use of unnecessary words/phrases.</td>
<td>Fairly brief but might profit from redacting some words/phrases.</td>
<td>Somewhat brief, but there are some lengthy parts that could be shortened.</td>
<td>Not brief, and several lengthy parts should be shortened significantly.</td>
<td>Elaborates on topics that are not or only tangentially relevant.</td>
</tr>
<tr>
<td>Comprehensive</td>
<td>All possible aspects have been covered, even the more esoteric ones</td>
<td>All the common and most significant aspects have been covered</td>
<td>The most common and significant aspects have been covered</td>
<td>Several common and significant aspects are absent from the answer</td>
<td>The answer provides no or nearly no meaningful aspects one might expect</td>
</tr>
<tr>
<td>Comprehensible</td>
<td>Extremely clear and easy-to-understand answer with no ambiguity.</td>
<td>Clear and easy-to-understand answer with only minimal ambiguity.</td>
<td>Generally clear answer that might be clarified and/or simplified.</td>
<td>Answer should be simplified and/or clarified to ease understanding.</td>
<td>Answer requires substantial simplification and clarification for legibility.</td>
</tr>
</tbody>
</table>

---

**Figure 2** Upset plot of the distribution of the five main themes found in the 258 articles from PubMed included.
comprehensible (for patients with high literacy) results compared with low-health literacy prompts, but they were also less concise. Description depth seems to have little effect, except that a more descriptive prompt sometimes gives a more comprehensible response, which seems sensible.

Finally, the prompts and responses in Danish scored better on all criteria, which is unexpected given that ChatGPT was trained predominantly on English language datasets. A likely explanation is that ChatGPT was trained on less user-generated and/or self-help material in Danish that might obscure and dilute the influence of official recommendations, textbooks and freely available professional doctor manuals. As an example, consider the absurd recommendation to avoid ‘tight clothing, particularly around your waist, groin and legs’ as this ‘can restrict blood flow’ (VV.HH.E1).

Figure 4 visualises the class imbalance-corrected, relative counts across health literacy L/H, description depth L/H and language E/D. Regarding health literacy, low-health literacy prompts receive worse scores on comprehensiveness and comprehensibility than high-health literacy prompts but slightly better scores on conciseness. Description depth presents somewhat more balanced, though high description depth also takes the majority of the comprehensible scores of five. Regarding language, Danish has both the most fives on all parameters and the fewest ones.

We conclude the quantitative assessment of the prompt-response pairs with an analysis of pairwise mutual information. Mutual information is the measure of information shared between two variables, that is, how much we can learn about one variable by knowing the other. If \( X, Y \) are random variables over the space \( X \times Y \), then their mutual information is:

\[
I(Y;X) = I(X;Y) = H(X) - H(X|Y)
\]

where \( H(X) \) is the entropy, and \( H(X|Y) \) is the conditional entropy of \( X \) given \( Y \). If \( X, Y \) are completely independent, then \( H(X|Y) = H(X) \). If \( X \) is completely determined by the value of \( Y \), then \( H(X|Y) = 0 \).

Figure 5 provides a heatmap of pairwise mutual information. Notably, comprehensibility and correctness are associated while the other criteria are not significantly influenced by each other. What matters the most for the scores are the response (and prompt) lengths, and to some degree the topic.

Qualitatively analysing the responses, we find that the verbs and nouns used in the prompts, as well as whether negation is used are highly influential on the style of the responses and their wording. The topics tend to receive ‘advice’ (ML.LL.E2) when asking what they are supposed to do and ‘general information’ (ML.LL.E1) when asking about the symptoms of a heart attack.

Interestingly, the description depth appears to have little bearing on the correctness and conciseness of the responses. In some instances, additional details in the prompt derail rather than improve responses, for example, when ChatGPT focuses on the (arbitrary) context of ‘fitness training’ (ML.LL.D5) mentioned, missing a potentially acute MI and providing extensive best practices of safe fitness training instead.

The correctness of health advice is also questionable in some of the responses, where ChatGPT seems to reproduce health beliefs rather than evidence-based medical consensus. In addition to ‘avoid tight clothing’ (V.HH.E1), ChatGPT also suggests ‘herbal remedies’ which it notes ‘are believed to improve blood circulation’ (V.LL.E2). It also suggests treating PAD by ‘sleeping on your side’ and to ‘elevate your legs while sleeping’. While the former is non-sensical, the latter should be considered harmful as it may result in falling pressure, triggering pain and, ultimately, ischaemic tissue loss. As noted above, such responses are more prolific in the English responses compared with the Danish.

Table 4 Average scores by topic with SE

<table>
<thead>
<tr>
<th></th>
<th>All together</th>
<th>Topic 1: MI</th>
<th>Topic 2: PAD</th>
<th>Topic 3: VV</th>
<th>Topic 4: CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>3.45±0.11</td>
<td>3.84±0.22</td>
<td>3.19±0.28</td>
<td>3.73±0.23</td>
<td>3.03±0.15</td>
</tr>
<tr>
<td>Concise</td>
<td>3.195±0.082</td>
<td>3.65±0.14</td>
<td>3.06±0.17</td>
<td>3.37±0.16</td>
<td>2.71±0.14</td>
</tr>
<tr>
<td>Comprehensive</td>
<td>3.52±0.10</td>
<td>3.55±0.23</td>
<td>3.19±0.28</td>
<td>4.13±0.17</td>
<td>3.10±0.15</td>
</tr>
<tr>
<td>Comprehensible</td>
<td>3.715±0.077</td>
<td>3.81±0.16</td>
<td>3.52±0.15</td>
<td>4.07±0.13</td>
<td>3.48±0.16</td>
</tr>
</tbody>
</table>

CP, cardiovascular prevention; MI, myocardial infarction; PAD, peripheral arterial disease; VV, varicose veins.
In many responses scoring low on conciseness, ChatGPT lists a multitude of symptoms, diagnosis and treatment options (PAD.LL.D5), or even other potential medical conditions (PAD.LL.E2). Despite these extensive responses, ChatGPT remains vague on the more quantitative aspects involved such as how many symptoms should a patient exhibit before medical assistance should be sought. When asked directly for quantitative information, ChatGPT usually just lists risk factors when asked to estimate risk (CP.HH.E5) or provides highly vague responses such as ‘80%–90%’ when asked for proportions of ‘preventable’ ‘heart attacks’ (CP.HH.E4). Finally, ChatGPT also provides some additional information without being prompted to do so, for example, by prominently including insights on how symptoms vary between individuals of differing gender (MI.HH.E5) or racial/ethnic backgrounds (T1.HH.D3).

DISCUSSION

Our quantitative analysis found average scores of just under four out of five for correctness. These results are well aligned with extant literature where subject professionals scored 100 and 96 prompt-response pairs, respectively, on a five-point grading scale, finding average scores around four out of five for validity/factual correctness.\(^\text{17}\)\(^\text{18}\) Our results demonstrate the value of considering three additional dimensions when evaluating prompt-response pairs: two additional independent dimensions (conciseness, comprehensibility), as well as one somewhat correlated but predominantly complementary dimension (comprehensiveness).

Our qualitative analysis uncovered some areas of concern that can be expected to have both direct and indirect implications for the use of LLMs for patient education, as well as potential for further improvements. While the correctness of the responses of LLMs is, of course, important, it is even more important to understand to what degree patients are believing the health information and advice acquired through LLMs,\(^\text{19}\) and how those beliefs are affecting their practices of diagnosis, treatments and interactions. Health information campaigns need to stress the ability of LLMs to make mistakes and the need to be critical of their responses and advice to use officially sanctioned health information. The development of carefully curated self-surveillance LLMs such as Med-PaLM-2 is a first step in this direction.\(^\text{20}\)

Furthermore, to increase validity and contextualise responses, AI systems such as ChatGPT need to be taught to find and/or derive quantitative estimates of risks, disease prevalence, biochemical thresholds, etc in addition to qualitative summaries and lists of symptoms, risk factors, diagnose and treatment options.

The organisations developing LLMs also need to critically reevaluate in which situations the responses should contain disclaimers, list long lists of diagnosis and treatment options that might or might not be available to the patients, or even nudge patients to seek medical assistance, and when this is neither required nor desirable.

<table>
<thead>
<tr>
<th>Table 5 Average scores grouped by different prompt classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
</tr>
<tr>
<td>Health literacy=low</td>
</tr>
<tr>
<td>Health literacy=high</td>
</tr>
<tr>
<td>Description depth=low</td>
</tr>
<tr>
<td>Description depth=high</td>
</tr>
<tr>
<td>Language=English</td>
</tr>
<tr>
<td>Language=Danish</td>
</tr>
</tbody>
</table>

Figure 4 Clustered relative population size histogram (corrected for class imbalance) visualising how the different characteristics influence the distributions of scores. If the filled and half-filled bars meet at 50% then there is no difference in the coverings, if they are tending to one side, one of the categories received fewer of this score than the other.
Here, LLMs could and arguably should be imbued with the capacity to interview the patient to better understand their situation and provide more personalised and selective responses.

For the time being, many of the responses we analysed leave an impression of shying away from such responsible considerations in favour of providing a combination of long authoritative sounding but only semicomprehensive responses to which disclaimers have been added as part of the human feedback integrated into the LLM development. Patients are confronted with unlikely but severe potential diseases, long lists of symptoms without guidance and many diagnosis and treatment options. This has the potential to both instigate fears and anxiety in patients, as well as to reinforce trends of self-diagnosis and overdiagnosis. Patients might also feel obliged to act on the advice obtained, with the alternative being to bear the responsibility of inactivity. Acknowledging the potential of LLMs like ChatGPT to get more patients the right diagnosis and necessary treatment in good time and, thus, to improve health outcomes, an important but non-trivial task will be to find the right balance between underplaying and overplaying risks and the need for action.

Last but not least, observing that the quality of the responses of ChatGPT nearly always reflects the quality of the prompts, LLMs might prove a hazard to rather than an aide of already disadvantaged populations. Not only do those who are less proficient at formulating prompts and possess less health literacy receive less valuable responses. The correlation between prompt and response quality also has the potential to amplify existing health inequalities, increasing the health literacy and agency of those who already possess relatively high levels and yielding little to no benefits for those who do not. Ultimately, this can be expected to amplify differences regarding key areas of contemporary health policy such as patient involvement, education and empowerment.

While the responses of ChatGPT might be perceived as more empathetic than human responses, medical and cross-disciplinary research on LLMs in medical fields needs to take into account how socioeconomic and cultural resources of the patients impact not only their interaction with LLMs but also their ability to benefit from this interaction. The very perception of an empathetic conversational partner in matters of health might exacerbate an already existing problem where low-resource patients shy away from interactions with medical professionals. Simultaneously, the ease of obtaining confident and extensive health advice might increase anxiety among high-resource patients.

Demanding consumer patients, who already equip themselves with printed results of Google searches when consulting medical professionals, might sooner than later start to demand to be diagnosed and treated using all the options known to and listed by LLMs that have already demonstrated surprising adeptness at differential diagnosis. Under current legislation, healthcare workers typically have no choice other than to comply with such demands to avoid liability and the risk of prosecution in case unlikely events do occur. This is a potential bomb under the healthcare systems, where insurance-driven systems need to raise the price of their insurances while tax-financed systems constantly given sparse financial...
resources will increasingly be forced to spend critical human resources on healthy citizens instead of patients actually requiring medical assistance.

Contributors ADL and TH curated the prompts and responses, performed the quantitative analysis and wrote the initial draft of the methods section. AS-K and MD provided input and feedback on the prompts and performed the qualitative analysis. TH and AS-K assessed conciseness and comprehensiveness. AS-K conceptualised the study, acts as guarantor, and, together with PS-K, wrote the initial and final drafts. ADL, TH and PS-K per-formed the state-of-the-art review. All authors provided comments and reflections on the responses and for the discussion.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as online supplemental information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC) 4.0 license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD Peter Schneider-Kamp http://orcid.org/0000-0003-4000-5570

REFERENCES