

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

## ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Money-oriented risk-takers or deliberate decision makers; a cross-sectional survey study of participants in controlled human infection trials
<b>AUTHORS</b>	Hoogerwerf, Marie-Astrid; de Vries, Martine; Roestenberg, Meta

## VERSION 1 – REVIEW

<b>REVIEWER</b>	Anuradha Rose Professor Dept of Bioethics Dept of Community Health Christian Medical College, Vellore, S India
<b>REVIEW RETURNED</b>	08-Oct-2019

<b>GENERAL COMMENTS</b>	If the two groups are different in terms of risk taking behavior, can the reasons to participate be extrapolated to the rest of society. Will it be possible to match participants and non participants in CHIM studies for RPS and analyze?
-------------------------	--

<b>REVIEWER</b>	José Gerardo Gonzalez-Gonzalez Univerdad Autónoma de Nuevo León Mexico
<b>REVIEW RETURNED</b>	28-Oct-2019

<b>GENERAL COMMENTS</b>	<p>Hoogerwerf et al. report a cross-sectional descriptive survey which aims to determine the motivation and perceptions influencing the decision to get enrolled in controlled-human infection trials (CHI-trials). This study is important to enhance the knowledge of factors that drive subjects to participate in CHI-trials, as well as their opinions about these studies. In addition, there is a gap of knowledge regarding this issue. This interesting paper could be a milestone for further studies that could improve the setting about participants of CHI trials. When this study was evaluated using the NIH "Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies" there are some aspects that could lead to find a moderated risk of bias (e.g. the sample size justification and the eligibility criteria). The article is well written. However, there are some issues that may need clarification for the manuscript to improve:</p> <p>1. The method of selection of the subjects is not very clear, "an anonymous paper survey was distributed" gives us very little information about how the surveys were delivered. It would be important to give more details about the survey distribution and implementation.</p>
-------------------------	--

	<p>2. In page 4 participant section it would be valuable that authors state the number of possible participants that composed the three CHI trials to give a clear sight of the number of subjects considered and contrast it with the number analyzed at the end of your study.</p> <p>3. The authors do not give so much detail about if they performed a sample size formula calculation and whether the numbers they considered are supported by the scientific literature or not.</p> <p>4. In the participant section page 4, It would be important that authors please define the inclusion and exclusion criteria. Without these criteria it is unclear why only 66 previous participants were eligible.</p> <p>5. In the survey section it would be helpful that authors describe what aspects were addressed in the CHI-Trials description aimed to inform the control group. In addition, it is important to state the reasons for excluding the Schistosomiasis CHI-trial description to the control group.</p> <p>6. There is no clear division between survey section and statistical analysis section, it would be valuable to create a section to describe the statistical plan analysis used in this paper.</p> <p>7. In page 5 line 54 it is important to the authors to state what LUMC means, as well as, IBR.</p> <p>8. In page 6 motivation section, It would be important to state the proportion of important or very important motivation factors, When the authors show both as just one category they give less information that what your survey could show.</p> <p>9. In page 7 line 34 and 38, in the "Assessment of symptoms and risks" section, it is important to report the results as a proportion in order to be congruent with the rest of the data shown. (e.g. "The majority of PP (93%) considered the trial to be of no or little risk and the majority was not afraid of symptoms before the start of the trial (80%).")</p> <p>10. In page 7 line 42, in the Assessment of symptoms and risks section. Was a statistical hypothesis test used to assume the statement "with no significant differences between CHI-models." If this was the case, could authors please report the "p" value.</p> <p>11. In page 7 line 58 and 59, it could be valuable to report the proportion of this statement "participants PP described to be glad to have been offered that proposition and was proud to have completed the study after all."</p> <p>12. It would be important to state as a possible limitation that the control group was not comparable with the study group since the first one belongs to a student's cohort who may have important cofounders that were not adjusted in comparison to the CHI trials participants.</p> <p>13. In the conclusion section it could be a strong asseveration to say, "we conclude that the current image of the CHI-participant as 'money-oriented risk- taker' is not accurate and should therefore be nuanced to the CHI-participant as 'deliberate decision- maker'." Even when the current study brings valuable evidence, further</p>
--	--

	<p>studies are needed to draw this conclusion.</p> <p>14. The table 1. Could improve its format by using the same format and reviewing the data:</p> <p>a. On the response of the question, "Employed in healthcare or healthcare related study?", the authors present the percentages with one decimal digit whereas the other percentages are rounded.</p> <p>b. In the same question, for the controls the percentage presented for the "Yes" answers is 80%, but if the author rounds it up as all of the other results presented this should be 81%. The same applies for the "No" answers for the controls, in this case the number shouldn't have been rounded, being the result 19.23%.</p> <p>c. I would recommend reviewing the percentages presented as results of the controls on the sex section; The percentage for 35/156 male participants should be 22% and the same applies for the female participants, 98/156 accounts for 63%.</p> <p>15. It would be recommendable that authors display the proportion of the survey answers in the figures section to make easier for the readers to analyze the data.</p>
--	---

<b>REVIEWER</b>	Alana Cattapan University of Waterloo, Canada
<b>REVIEW RETURNED</b>	03-Nov-2019

<b>GENERAL COMMENTS</b>	<p>The objective of the paper is clear, and the general study the paper provides useful information about the motivations, and decision-making, and experiences of participants in the relevant studies. At the same time, the study does more than it needs to do to achieve that objective. The use of a control groups, is of particular concern, as it was unclear that the motivations of people who have participated can be weighed against people who have thought about doing so. Reorganizing the text to focus on the self-reported motivations of actual participants and then to identify the relationship of each of the findings for the PP group in relation to the relevant literature even more would strengthen this paper considerably. As I mention below, it felt to me as if the findings vis-à-vis the two parts of the control group could be its own study/publication.</p> <p>I am not an expert in quantitative methods, and cannot comment on the use of statistics in this paper.</p> <p>Major concerns</p> <ul style="list-style-type: none"> <li>The purpose of the control group in this study is not entirely clear to me. Identifying the motivation of someone who has participated in a study feels like a very different exercise from identifying the potential motivation of someone who might or might not potentially participate in one--it is unclear what comparing prospective and retrospective motivation (including motivation not to participate) serves. In other words, it is unclear if people who would participate (or would not) is a good control. This concern is apparent throughout the paper, and the division of the analysis into CN, CP, and PP suggests an interest in the answers of all, rather than the motivations of participants... (The homogeneity of the control group also raises some concerns about its use. The control group was composed entirely of students who were close in age, and who were largely actively employed in or engaged in the study of health care. Research has shown that people working in health care are more</li> </ul>
-------------------------	--

	<p>likely to participate in health research, and that there are biases associated with age, and other demographic elements that were relatively homogenous in the control group.)</p> <ul style="list-style-type: none"> <li>• The description of the methods on pages 3 and 4 could use more clarity about how the survey instrument was designed and distributed. The references to “previously published research” and “topics of ethical debate” do not give readers a clear enough sense of the way the survey was structured, or how the questions were developed. How does Castor EDC work – by phone? By email? Did follow-ups occur if people don’t respond a first time? More information about the design of the study (i.e., inclusion/exclusion of demographic characteristics, inclusion/exclusion of potential motivations from the literature, inclusion/exclusion of “other” category in certain places, space for qualitative comments, development/use of risk-propensity measures...). While the survey itself (Supplement A) helps to answer some of these questions, more description of the survey development within the text would help readers understand the relationship of this work to the relevant literature on participant motivations in healthy-volunteer/low reward studies.</li> <li>• I mentioned this above, but the discussion section largely focuses on the relationships between PP and CC (CP/CN) rather than the motivations and experiences of PP and situating it within the relevant literature – identifying whether the motivations and experiences of PP in CHI trials are studies are similar to those of people in other kinds of relevant studies (i.e. Phase 1 studies). The introduction and abstract led me to believe, as a reader, that this was a critical part of doing this study, that is, to identify whether the motivations of CHI trial participants were similar to those of people in Phase 1 studies. There is a brief discussion of the relevance of Grady et. al., to the study but money isn’t the only crossover in motivations between phase 1 and CHI participants, and this could be much better described.</li> <li>• Further, while the discussion of the Kenyan and American CHI studies on page 9-10 is helpful, but this would be more effective if it was its own paragraph(s), rather than mixed in with the discussion of Phase I trial volunteers. Discussing the relationship between the present study and Phase I trials is one significant point. The relationship between the present study and other CHI trials is another significant point and they would both be more effective if more clearly differentiated.</li> </ul> <p>Other concerns</p> <ul style="list-style-type: none"> <li>• (Page 1) Lines 38-42 seem to be providing the justification for the rest of the paper – the need for “quantitative data on motivations and experiences” of CHI-volunteers, on a large scale. However, the scale of the research is relatively small (medium?) – only 66 potential participants in total. Some rephrasing in the introduction and anywhere else scale is discussed to emphasize the need for quantitative data (rather than the scale of the work) would be helpful.</li> <li>• (Page 1) From lines 20-24, greater clarity about the personality-related research on risk-taking and participation in</li> </ul>
--	--

	<p>clinical trials, is needed. It was not clear to me as a reader that participants in phase-1 trials have a “reckless lifestyle” – rather than people slightly more likely to take risks than others. Also the term “sensation-seeking” could be better described...risk-taking? Thrill-seeking? Greater precision with language in this section of the introduction would help set the stage for readers.</p> <ul style="list-style-type: none"> <li>• (Page 1) At line 20, the use of the word “Alike” should just be “Like.” Further, on page 1, in the sentence beginning “Qualitative data,” there seems to be a missing verb....(</li> <li>• (Page 1). At lines 30-31, there is a citation for the “recent publication”, but not for the “public discussion” that followed. There are other places in the text (some others being page 8, line 21 “contrasting public belief” and page 10, line 37 – “than currently given credit to”) where there are general, unsubstantiated references to public opinion that could use more context or citation.</li> <li>• As I was rereading, I was wondering if the nature of the infection had any role to play in the motivations of participants. Did it matter to PP if the study was malaria, hookworm, or schistosomiasis?</li> </ul>
--	--

## VERSION 1 – AUTHOR RESPONSE

### Reviewer 1

If the two groups are different in terms of risk taking behavior, can the reasons to participate be extrapolated to the rest of society. Will it be possible to match participants and non participants in CHIM studies for RPS and analyze?

We thank the reviewer for this suggestion. We agree that this study shows that there are differences in risk-taking behavior between participants and non-participants. Investigating this was one of the aims of the survey. Knowledge on risk-taking behavior is important and may need to be addressed in recruitment. Interestingly, our study shows that controls who would participate in a CHI-study have a higher RPS than those who would not. Controls who would participate have an RPS similar to the actual participants. This is clarified in the discussion of the manuscript as “Both PP and CP scored higher on the RPS as compared to CN” Thus, matching is not needed, the willingness to participate seems to be higher among those with higher RPS. Indeed, the motivations of the controls who would potentially participate (with higher RPS) are different from the actual participants. This may reflect the actual practice of recruitment, whereby some eventually decide not to participate, or the motivation of participants may change over time as the selection process continues.

### Reviewer 2

Hoogerwerf et al. report a cross-sectional descriptive survey which aims to determine the motivation and perceptions influencing the decision to get enrolled in controlled-human infection trials (CHI-trials). This study is important to enhance the knowledge of factors that drive subjects to participate in CHI-trials, as well as their opinions about these studies. In addition, there is a gap of knowledge regarding this issue. This interesting paper could be a milestone for further studies that could improve the setting about participants of CHI trials. When this study was evaluated using the NIH “Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies” there are some aspects that could lead to find a moderated risk of bias (e.g. the sample size justification and the eligibility criteria). The article is well written. However, there are some issues that may need clarification for the manuscript to improve:



1. The method of selection of the subjects is not very clear, “an anonymous paper survey was distributed” gives us very little information about how the surveys were delivered. It would be important to give more details about the survey distribution and implementation.

The method section has been adjusted to more clearly reflect this:

“Students from the local university were recruited as control subjects. This group has been selected as the majority of participants in CHI-studies at the study center is recruited from this population in a similar manner, providing the best reflection of the majority of participants. Students were handed an anonymous paper survey by the researchers during lectures at the medical faculty and during meetings of local student societies. Surveys were collected afterwards.”

2. In page 4 participant section it would be valuable that authors state the number of possible participants that composed the three CHI trials to give a clear sight of the number of subjects considered and contrast it with the number analyzed at the end of your study.

The number of participants has been made more clear in the sentence “All 66 previous participants were eligible for inclusion”.

The number of eligible participants per trial has been added to table 1.

3. The authors do not give so much detail about if they performed a sample size formula calculation and whether the numbers they considered are supported by the scientific literature or not.

Because the study is primarily descriptive, a formal sample size calculation was not performed.

Because we were limited in number given the requirement of previous participation in the CHI trial, we have tried to recruit as many volunteers as possible from this group. Estimation of willingness to participate was based on our anecdotal experience in recruiting volunteers.

We wanted to match the participants with an equal number of controls willing to participate in order to compare the motivational factors. From our experience in recruiting we estimated that about one third of student controls would be willing to participate. In order to match our expected survey return by participants we therefore aimed to include 150 controls. Although not all questions were answered, all questionnaires were returned by the students.

The paragraph in the methods has been clarified:

“With an expected response rate of 80% we estimated that around 50 previous participants would return the survey. Based on experiences in recruiting we estimated that one-third of students would be willing to participate in a CHI-trial, so to include an equal number of controls willing to participate to actual participants we aimed to include 150 controls.”

4. In the participant section page 4, It would be important that authors please define the inclusion and exclusion criteria. Without these criteria it is unclear why only 66 previous participants were eligible.

The method section has been adjusted to clarify the in- and exclusion criteria and the number of eligible participants with the following:

“Inclusion criteria were having undergone controlled human infection and having previously consented to be contacted again for further studies. There were no exclusion criteria. All 66 previous participants were thus eligible for inclusion.”

5. In the survey section it would be helpful that authors describe what aspects were addressed in the CHI-Trials description aimed to inform the control group. In addition, it is important to state the reasons for excluding the Schistosomiasis CHI-trial description to the control group.

We agree that the information provided is important to obtain appropriate answers. The following description has been added to the manuscript: (...) “whereas the control group (CC) were asked to consider participation in a malaria trial and a trial with hookworm to reflect the different types of studies conducted. CC were provided descriptions of the trials detailing study procedures, possible adverse events, number of visits and sample collections and the financial compensation (descriptions in supplement A).” In order to limit the amount of information and reading materials provided to the controls we decided to select representative information material. We considered the malaria and

hookworm trials to be most representative, because in our experience the recruitment for hookworm trials is most difficult and malaria is easiest. We thus expected less willingness to participate in the worm-trial compared to the malaria trial, which was confirmed in the survey.

6. There is no clear division between survey section and statistical analysis section, it would be valuable to create a section to describe the statistical plan analysis used in this paper. A separate section for the statistical analysis has been created in the methods section.

"A ranking order of motivational and decision-making factors was compiled, ranking from the factor with the highest percentage of 'important' or 'very important' to the lowest. RPS scores were analysed as described by Meertens.<sup>20</sup> Differences in mean scores were calculated using a two-sided t-test or one-way ANOVA and were adjusted for age and sex using a univariate analysis. Multiple choice questions on the experiences of PP and ethical issues were analysed using descriptive statistics. Differences in demographical characteristics were calculated using a Chi-square test, differences between CHI-models were calculated using a one-way ANOVA for continuous data and Fisher's exact or Kruskal-Wallis test for categorical data."

7. In page 5 line 54 it is important to the authors to state what LUMC means, as well as, IBR. These abbreviations have been clarified into the following: "The institutional review board of the Leiden University Medical Center, where the study was performed, has reviewed the protocol and provided ethical approval (P18.203)."

8. In page 6 motivation section, It would be important to state the proportion of important or very important motivation factors, When the authors show both as just one category they give less information that what your survey could show.  
We agree that it is important to show as much data as possible. We have thus added the numbers for the individual categories to the text as follows:  
"PP considered "contributing to science" as an important (43%) or very important (38%) motivating factor, followed by "contributing to developing countries" (41% important, 31% very important, and the financial compensation (25% and 38% respectively, figure 1). This contrasted with the motivation of CP, where the largest group found the financial compensation to be an important motivation to participate (39% important, 52% very important) followed by "contributing to science" (33% important, 39% very important) and "contributing to developing countries"(46% important, 26% very important).

9. In page 7 line 34 and 38, in the "Assessment of symptoms and risks" section, it is important to report the results as a proportion in order to be congruent with the rest of the data shown. (e.g. "The majority of PP (93%) considered the trial to be of no or little risk and the majority was not afraid of symptoms before the start of the trial (80%).)  
The percentages have been changed into proportions, as suggested.

10. In page 7 line 42, in the Assessment of symptoms and risks section. Was a statistical hypothesis test used to assume the statement "with no significant differences between CHI-models." If this was the case, could authors please report the "p" value.  
This was analysed using a one-way ANOVA test. The p-value has been added to the manuscript. With a p-value of 0.078 we have described this as no clear significant differences.

11. In page 7 line 58 and 59, it could be valuable to report the proportion of this statement "participants PP described to be glad to have been offered that proposition and was proud to have completed the study after all."  
We understand that this sentence is confusing. This statement concerns only one participant. This has been clarified in the manuscript.

12. It would be important to state as a possible limitation that the control group was not comparable with the study group since the first one belongs to a student's cohort who may have important cofounders that were not adjusted in comparison to the CHI trials participants.

We agree that the limitations of the comparison should be clear. We have thus added the following text to the discussion:

"The use of the control group has several limitations. The control group of students is more homogenous in age, education and healthcare background than the actual participants which impairs generalizability. Controls were furthermore offered a hypothetical participation, which may not be comparable to the actual decision to take part. However, participants are selected from the same population and this control group thus represents two-thirds of trial participants. We thus believe that the comparison is still of value."

We have added the following text to the article summary:

"Control group were students, a more homogeneous population than the participants which consist of roughly 2/3 students. This difference may hamper comparison."

13. In the conclusion section it could be a strong asseveration to say, "we conclude that the current image of the CHI-participant as 'money-oriented risk- taker' is not accurate and should therefore be nuanced to the CHI-participant as 'deliberate decision- maker'." Even when the current study brings valuable evidence, further studies are needed to draw this conclusion.

The wording of the statement has been altered as follows: "Based on these findings the current image of the CHI-participant as 'money-oriented risk-taker' may need adjustment to the CHI-participant as 'deliberate decision-maker'.

14. The table 1. Could improve its format by using the same format and reviewing the data:

- a. On the response of the question, "Employed in healthcare or healthcare related study?", the authors present the percentages with one decimal digit whereas the other percentages are rounded.
- b. In the same question, for the controls the percentage presented for the "Yes" answers is 80%, but if the author rounds it up as all of the other results presented this should be 81%. The same applies for the "No" answers for the controls, in this case the number shouldn't have been rounded, being the result 19.23%.
- c. I would recommend reviewing the percentages presented as results of the controls on the sex section; The percentage for 35/156 male participants should be 22% and the same applies for the female participants, 98/156 accounts for 63%.

Thank you for these corrections, the numbers have been adjusted accordingly.

15. It would be recommendable that authors display the proportion of the survey answers in the figures section to make easier for the readers to analyze the data.

We have adjusted the figures according to the reviewer's suggestion.

### Reviewer 3

The objective of the paper is clear, and the general study the paper provides useful information about the motivations, and decision-making, and experiences of participants in the relevant studies. At the same time, the study does more than it needs to do to achieve that objective. The use of a control groups, is of particular concern, as it was unclear that the motivations of people who have participated can be weighed against people who have thought about doing so. Reorganizing the text to focus on the self-reported motivations of actual participants and then to identify the relationship of each of the findings for the PP group in relation to the relevant literature even more would strengthen this paper considerably. As I mention below, it felt to me as if the findings vis-à-vis the two parts of the control group could be its own study/publication.

I am not an expert in quantitative methods, and cannot comment on the use of statistics in this paper.

Major concerns



- The purpose of the control group in this study is not entirely clear to me. Identifying the motivation of someone who has participated in a study feels like a very different exercise from identifying the potential motivation of someone who might or might not potentially participate in one—it is unclear what comparing prospective and retrospective motivation (including motivation not to participate) serves. In other words, it is unclear if people who would participate (or would not) is a good control. This concern is apparent throughout the paper, and the division of the analysis into CN, CP, and PP suggests an interest in the answers of all, rather than the motivations of participants... (The homogeneity of the control group also raises some concerns about its use. The control group was composed entirely of students who were close in age, and who were largely actively employed in or engaged in the study of health care. Research has shown that people working in health care are more likely to participate in health research, and that there are biases associated with age, and other demographic elements that were relatively homogenous in the control group.)

We understand that the control group purpose as well as its representation and limitations should be stated more clearly. We have therefore adjusted the manuscript throughout according to the suggestions:

- In the introduction, the purpose of the control group has been more explicitly stated.

“ In order to investigate whether participants in CHI-trials are different from the general population it is valuable to compare the participants to a control group. This also enables a longitudinal comparison of motivations and thought-processes of potential participants with those who have actually participated, providing a better insight into how volunteers come to their decision.”

- The results section has been restructured to focus primarily on the results of the CHI participants, with only a secondary discussion of the comparison with controls.

- In the discussion section we have made more explicit the differences and limitations of the control group:

“ The use of the control group has several limitations. The control of group of students is more homogenous in age, education and healthcare background than the actual participants which impairs generalizability. Controls were furthermore offered a hypothetical participation, which may not be comparable to the actual decision to take part. However, participants are largely selected from the same population and this control group represents two-thirds of trial participants. We thus believe that the comparison is still of value. “

- The description of the methods on pages 3 and 4 could use more clarity about how the survey instrument was designed and distributed. The references to “previously published research” and “topics of ethical debate” do not give readers a clear enough sense of the way the survey was structured, or how the questions were developed. How does Castor EDC work – by phone? By email? Did follow-ups occur if people don’t respond a first time? More information about the design of the study (i.e., inclusion/exclusion of demographic characteristics, inclusion/exclusion of potential motivations from the literature, inclusion/exclusion of “other” category in certain places, space for qualitative comments, development/use of risk-propensity measures...). While the survey itself (Supplement A) helps to answer some of these questions, more description of the survey development within the text would help readers understand the relationship of this work to the relevant literature on participant motivations in healthy-volunteer/low reward studies.

We agree that more information on the study design is needed. We have therefore extended the description of the study design:

- In the ‘Participants’ subsection: “Participants of previously conducted CHI-trials with malaria, hookworm or schistosomiasis were invited to participate in an anonymous survey. Inclusion criteria were having undergone controlled human infection and having previously consented to be contacted again for further studies. There were no exclusion criteria. All 66 previous participants were eligible for

inclusion. CHI-trials were conducted between November 2016 and September 2018. Surveys were distributed and collected via e-mail through the data management program Castor EDC. Participants who did not respond to the e-mail were sent one reminder to complete the survey.”

- In the ‘Survey’ subsection: “Motivational and decision-making factors were chosen based on the research by Grady et al and by identification of potential motivational factors through discussion with researchers involved in screening and recruitment of trial participants. Each question had an open option to allow participants to provide their own factors. Questions on ethical acceptability were formulated based on issues identified in literature as key concepts in CHI-trials”

- I mentioned this above, but the discussion section largely focuses on the relationships between PP and CC (CP/CN) rather than the motivations and experiences of PP and situating it within the relevant literature – identifying whether the motivations and experiences of PP in CHI trials are studies are similar to those of people in other kinds of relevant studies (i.e. Phase 1 studies). The introduction and abstract led me to believe, as a reader, that this was a critical part of doing this study, that is, to identify whether the motivations of CHI trial participants were similar to those of people in Phase 1 studies. There is a brief discussion of the relevance of Grady et. al., to the study but money isn’t the only crossover in motivations between phase 1 and CHI participants, and this could be much better described.

We have rearranged the discussion to reflect the reviewer’s comments and have limited the comparisons with the control group in the discussion. The paragraph on the comparison with phase 1 trials has been extended:

“The motivations of CHI participants seem to be concurrent with findings in volunteers of phase I drug trials. Stunkel and Grady describe in a 2011 systematic review that although the financial compensation is usually necessary, it is not sufficient for participation, and note that risk is the deciding factor in participation. However, other large-scale studies in phase I drug-research participants, noted that money is the most important motivator in 60% of individuals, which is clearly more than we found. Possibly, the population (students, gender and age) might play a role in motivating factors as well as the nature of the trial. A survey of the motivations of individuals participating in Ebola and influenza vaccines is a good example of the latter, whereby almost 90% of participants found contributing to the health of others important. It is possible that both CHI-trials, especially those researching vaccines for Neglected Tropical Diseases and phase 1 trials for vaccines with similar expected public health benefits may attract volunteers with more altruistic motivations compared to phase I drug research in general.”

- Further, while the discussion of the Kenyan and American CHI studies on page 9-10 is helpful, but this would be more effective if it was its own paragraph(s), rather than mixed in with the discussion of Phase I trial volunteers. Discussing the relationship between the present study and Phase I trials is one significant point. The relationship between the present study and other CHI trials is another significant point and they would both be more effective if more clearly differentiated.

We agree with the reviewer and have made this into separate paragraphs. The paragraph on phase 1 trials is quoted above. The paragraph on the comparison with other CHI-trials has been adjusted and separated:

“Differences in population may also be reflected within CHI-studies in different countries. Our Dutch PP were motivated by other factors than Kenyan participants of a controlled human malaria infection trial, who were most often driven by the financial compensation and the health care provided by the trial staff.<sup>18</sup> The Kenyans were rewarded the wage of a day’s work for each day of participation to make up for lost income. This was different for the Dutch PP, who have universal access to healthcare and receive compensation for time spent and travel expenses. Participants from both countries, however, showed little concern about trial risks and showed high levels of trust in the study team. In a qualitative study amongst US controlled human malaria infection participants<sup>17</sup> the

participants similarly describe little concerns about the risks, trust in the study team as important and mixed motivations for participation. The differences between the American, Kenyan and Dutch CHI-participants illustrate the influence of cultural differences and healthcare organization that remain important to address and separately investigate.'

Other concerns

- (Page 1) Lines 38-42 seem to be providing the justification for the rest of the paper – the need for “quantitative data on motivations and experiences” of CHI-volunteers, on a large scale. However, the scale of the research is relatively small (medium?) – only 66 potential participants in total. Some rephrasing in the introduction and anywhere else scale is discussed to emphasize the need for quantitative data (rather than the scale of the work) would be helpful. We have clarified the need for this study and it's differences in size in the Introduction: “However, these studies only included small groups of participants (16 and 36 respectively) in a malaria trial, and quantitative data on motivations and experiences is lacking. Given the ongoing debate on the ethics of CHI-trials, amore quantitative assessment of the experiences and motivation of participants in a broader group of volunteers is needed to gain a better insight into the profile of the CHI-volunteer, their motivations and experiences.”

We have furthermore adjusted the description of trial size in the discussion:

“Notwithstanding, this study has included a reasonably large number of CHI-participants compared to previous studies and covers several different CHI-models, thereby improving generalizability.”

- (Page 1) From lines 20-24, greater clarity about the personality-related research on risk-taking and participation in clinical trials, is needed. It was not clear to me as a reader that participants in phase-1 trials have a “reckless lifestyle” – rather than people slightly more likely to take risks than others. Also the term “sensation-seeking” could be better described...risk-taking? Thrillseeking? Greater precision with language in this section of the introduction would help set the stage for readers.

These sentences have been altered into the following phrases:

“Participants in phase I trials score higher on questionnaires examining sensation-seeking behaviours compared to age- and sex-matched controls, adding to the notion that these volunteers are more prone to take, possibly ill-considered, risks in their lives.”

- (Page 1) At line 20, the use of the word “Alike” should just be “Like.” Further, on page 1, in the sentence beginning “Qualitative data,” there seems to be a missing verb....( This has been corrected.

- (Page 1). At lines 30-31, there is a citation for the “recent publication”, but not for the “public discussion” that followed. There are other places in the text (some others being page 8, line 21 “contrasting public belief” and page 10, line 37 – “than currently given credit to”) where there are general, unsubstantiated references to public opinion that could use more context or citation.

The public opinion mentioned in lines 30-31 involved discussions in mainly Twitter and other social media following the referenced article, and responses to the article itself, posted alongside it. It is therefore difficult to give a single reference to an online discussion. The Introduction has been clarified to better show the origin of the discussion:

“In response to a recent publication public discussion, particularly on social media (...)”

References have been added to the lines in page 8 and wording has been altered to clarify the origin of the statement: “We have found that, contrary to commonly mentioned fears (...)”

The wording on page 10 has been altered to “We found that the motivation of CHI-participants is highly varied with significant importance for altruistic motivations.”

- As I was rereading, I was wondering if the nature of the infection had any role to play in the motivations of participants. Did it matter to PP if the study was malaria, hookworm, or

schistosomiasis?

We agree with the reviewer that this is an interesting question. This question is addressed now more prominently in the Result section, subsection Motivation: "There were no apparent differences in motivation for participants from different CHI-models." As we did not find any relevant differences between different models this is not further elaborated in the manuscript.

## VERSION 2 – REVIEW

<b>REVIEWER</b>	PhD MD Jose Gerardo Gonzalez Gonzalez Universidad Autonoma de Nuevo Leon
<b>REVIEW RETURNED</b>	05-Jan-2020

<b>GENERAL COMMENTS</b>	<p>Hoogerwerf et al. Performed a cross-sectional survey study describing the motivation, decision-making, and experience of previous participants of controlled human infection studies comparing these to a control group of university students. The conclusion of the study is important for the development of controlled human infection trials and fits the scope of the journal. The manuscript is well-written and understandable overall with the previous commentaries clearly answered. I further suggest the following:</p> <ol style="list-style-type: none"> <li>1. Could the authors expand further on the process undertaken to invite the students to participate in the study as the control group?</li> <li>2. It may be beneficial to specify the decision-making factors assessed in the "survey" paragraph for improving the clarity of the description of the survey.</li> <li>3. Could the authors present an example of the Risk Propensity Scale questionnaire given to the participants as supplementary material? Alternatively, Could the authors specify the items evaluated by the questionnaire?</li> <li>4.- It may be advisable to perform a statistical inference test to the main results to compare the proportion of patients in the motivational surveys between control and previous participants group.</li> <li>5.- The authors may want to consider adding a table or figure summarizing the results of the Risk Propensity Scale as this could make this section of the results easier to comprehend.</li> <li>6.- In the discussion section authors states that financial compensation is more important for the control group than the previous participants, however, panels C and D from figure 1 clearly show that financial compensation was ranked as the singled most important factor in a similar proportion in both groups.</li> </ol> <ol style="list-style-type: none"> <li>1. Could the authors expand further on the process undertaken to invite the students to participate in the study as the control group?</li> <li>2. It may be beneficial to specify the decision-making factors assessed in the "survey" paragraph for improving the clarity of the description of the survey.</li> <li>3. Could the authors present an example of the Risk Propensity Scale questionnaire given to the participants as supplementary material? Alternatively, Could the authors specify the items evaluated by the questionnaire?</li> <li>4.- It may be advisable to perform a statistical inference test to the main results to compare the proportion of patients in the motivational surveys between control and previous participants group.</li> <li>5.- The authors may want to consider adding a table or figure summarizing the results of the Risk Propensity Scale as this could make this section of the results easier to comprehend.</li> </ol>
-------------------------	---

	6.- In the discussion section authors states that financial compensation is more important for the control group than the previous participants, however, panels C and D from figure 1 clearly show that financial compensation was ranked as the singled most important factor in a similar proportion in both groups.
--	---

<b>REVIEWER</b>	Benjamin Mayer Ulm University
<b>REVIEW RETURNED</b>	12-Feb-2020

<b>GENERAL COMMENTS</b>	<p>I was asked to review this revised paper from a statistical perspective. Overall, I think the authors did a great job when addressing the comments raised during the initial review phase. The manuscript in its current form looks well-structured and its contents perfectly impart the specific piece of knowledge the authors would like to present. Some minor points will need clarification from my point of view and I will list them in the following:</p> <ul style="list-style-type: none"> <li>- page 5, lines 180 and 181: The authors stated that "...multiple choice questions...were analyzed using descriptive statistics". I am not sure what this means, since "descriptive statistics" include both, calculating observed frequencies for categorical variables as well as calculating means, standard deviations, etc. for continuously measured variables. I guess that the authors refer to the calculation of frequencies for categorical MC-answers, but this is just a guess. The authors may be more clear on that point.</li> <li>- page 5, line 183: The authors stated that they used a Kruskal-Wallis test for categorical data. I am not sure whether they finally used this test (I found no results indicating a non-parametric comparison of &gt;2 groups), and furthermore the authors should be aware of the fact that the Kruskal-Wallis test is only able to handle ordinally scaled categorical variables, not categorical variables in general.</li> <li>- page 5, paragraph on statistical analysis in general: I missed the information to which type 1 error the authors are referring to. I think the authors should add a sentence like "A p-value <math>\leq 0.05</math> was considered statistically significant."</li> <li>- page 7, lines 243 and 244: The authors reported an ANOVA p-value of 0.078 for the comparison of CHI-models. However, in view of the descriptive statistics reported in line 243 on that page, I am not sure whether it was appropriate to use ANOVA. The reported SD is relatively large when compared with the corresponding mean, which might be an indication of a skewed distribution. The authors may check the appropriateness of a non-parametric analysis.</li> </ul>
-------------------------	--

## VERSION 2 – AUTHOR RESPONSE

### Reviewer 2

1. Could the authors expand further on the process undertaken to invite the students to participate in the study as the control group?

The procedure has been clarified as follows:

“ Before lectures at the medical faculty the anonymous paper survey was distributed to all students present and collected afterwards. Surveys were furthermore distributed during two meetings of local



(non-medical) student societies, where the researchers handed students present the survey and collected them after completion.”

2. It may be beneficial to specify the decision-making factors assessed in the “survey” paragraph for improving the clarity of the description of the survey.

The following description has been added to the paragraph:

“Motivational factors in the survey were “curiosity”, “contributing to science”, “contributing to developing countries”, “financial compensation”, “interest in the subject” and “personal experience”. Factors in the decision making process were “Severity of possible symptoms”, “chance of developing symptoms”, “time investment”, “an easy way to make money”, “trust in the study team” and “it’s research about parasites”.”

3. Could the authors present an example of the Risk Propensity Scale questionnaire given to the participants as supplementary material? Alternatively, Could the authors specify the items evaluated by the questionnaire?

The Risk Propensity Score has been added as Supplementary Material B.

4.- It may be advisable to perform a statistical inference test to the main results to compare the proportion of patients in the motivational surveys between control and previous participants group. All previous participants were healthy volunteers who were selected based on their naivety to the pathogen under investigation. Since all studies concerned diseases not endemic in the Netherlands it is highly unlikely any of the student controls or previous participants would have been pre-exposed. We agree with the reviewer’s concerns that the control group may not be completely representative of the study population. This has been emphasised more clearly in the discussion:

“The control group of students may not be a complete representation of the participant population as it is more homogenous in age, education and healthcare background than the actual participants which impairs generalizability.””

5.- The authors may want to consider adding a table or figure summarizing the results of the Risk Propensity Scale as this could make this section of the results easier to comprehend.

We had initially limited the number of figures and tables to five, corresponding the BMJ Open submission guidelines. However, upon the reviewer’s suggestion we have added a figure with a graphical representation of the Risk Propensity Scale outcomes as figure 5.

6.- In the discussion section authors states that financial compensation is more important for the control group than the previous participants, however, panels C and D from figure 1 clearly show that financial compensation was ranked as the singled most important factor in a similar proportion in both groups.

The distinction between the single most important motivation and how often participants found a motivational factor important was unclear in this section. The sentence has been adjusted as follows to clarify:

“A larger group of CC found the compensation important compared to PP, although as a single most important motivation for participation proportions were similar.”

Reviewer 4

- page 5, lines 180 and 181: The authors stated that “...multiple choice questions...were analyzed using descriptive statistics”. I am not sure what this means, since “descriptive statistics” include both, calculating observed frequencies for categorical variables as well as calculating means, standard deviations, etc. for continuously measured variables. I guess that the authors refer to the calculation of frequencies for categorical MC-answers, but this is just a guess. The authors may be more clear on

that point.

The reviewer is correct, we indeed meant the calculation of frequencies for categorical multiple choice questions. We have adjusted the lines accordingly:

"Frequencies were calculated for the multiple-choice questions on the experiences of PP and ethical issues."

- page 5, line 183: The authors stated that they used a Kruskal-Wallis test for categorical data. I am not sure whether they finally used this test (I found no results indicating a non-parametric comparison of >2 groups), and furthermore the authors should be aware of the fact that the Kruskal-Wallis test is only able to handle ordinally scaled categorical variables, not categorical variables in general.

This is indeed an error in the description: the tests used were one-way ANOVA for continuous data and Chi-square for categorical data and should have included the Kruskal-Wallis test for continuous non-parametric data instead of categorical data. We have adjusted the lines accordingly:

"(...) differences between CHI-models were calculated using a one-way ANOVA for continuous parametric data and Kruskal-Wallis test for non-parametric data, and a Chi-square test for categorical data."

- page 5, paragraph on statistical analysis in general: I missed the information to which type 1 error the authors are referring to. I think the authors should add a sentence like "A p-value  $\leq 0.05$  was considered statistically significant."

We thank the reviewer for this addition and have added the suggested line to the paragraph.

- page 7, lines 243 and 244: The authors reported an ANOVA p-value of 0.078 for the comparison of CHI-models. However, in view of the descriptive statistics reported in line 243 on that page, I am not sure whether it was appropriate to use ANOVA. The reported SD is relatively large when compared with the corresponding mean, which might be an indication of a skewed distribution. The authors may check the appropriateness of a non-parametric analysis.

We had initially performed a one-way ANOVA based on the sample size, however given the reviewer's comments we have re-checked and do indeed see a skewed distribution to the left. We have re-run the analysis using a Kruskal-Wallis test for non-parametric data. This resulted in a p-value of 0.228. The lines 243 and 244 have been adjusted accordingly:

"with no clear significant differences between CHI-models ( $p=0.228$ )."

The Kruskal-Wallis test has been added to the description of the statistical analysis as described above.

### VERSION 3 – REVIEW

<b>REVIEWER</b>	PhD MD Jose Gerardo Gonzalez Gonzalez Universidad Autonoma de Nuevo Leon Mexico
<b>REVIEW RETURNED</b>	17-Mar-2020
<b>GENERAL COMMENTS</b>	<p>Line 186 Page 5 Author state they used one-way ANOVA and two-sided t-test to assess mean differences of risk propensity scores and were adjusted for age and sex. I am not sure if in both of these statistical tests cofounder variable adjustment can be done. Did the authors meant that they used a general linear model to adjust for age and sex?</p> <p>Having healthcare-associated work and being a medical student could have been a cofounder variable as those participants can have a different perspective of the risks associated with hookworms</p>

	<p>and malaria infections and it should have been taken in to account in the adjusted analysis.</p> <p>The authors should add as a limitation and possible cofounding variable that 69% of the control group would have not participated in any of the presented CHI trials, as this suggests that participants in the control group are not comparable to previous participants</p> <p>The results section's grammar should address minor revisions from an English native speaker.</p> <p>In Figure's 5 legend it should be included what does the error bars and figures estimate represent (standard deviation, median, mean, confidence intervals)</p> <p>Authors should add a statement that differences between control and previous participants' motivation and decision to participate reasons are merely descriptive as they did not perform a chi-square test (with its proper correction taking into account low expected value and counts) or another applicable statistical test to provide p-values and confidence intervals.</p>
<b>REVIEWER</b>	Benjamin Mayer Ulm University
<b>REVIEW RETURNED</b>	24-Feb-2020
<b>GENERAL COMMENTS</b>	The authors answered all of my questions raised through the most recent review round adequately. I recommend publication of this article in its current form.

### VERSION 3 – AUTHOR RESPONSE

#### Reviewer 2

Line 186 Page 5 Author state they used one-way ANOVA and two-sided t-test to assess mean differences of risk propensity scores and were adjusted for age and sex. I am not sure if in both of these statistical tests cofounder variable adjustment can be done. Did the authors meant that they used a general linear model to adjust for age and sex?

Having healthcare-associated work and being a medical student could have been a cofounder variable as those participants can have a different perspective of the risks associated with hookworms and malaria infections and it should have been taken in to account in the adjusted analysis.

Initially we did use a one-way ANOVA and t-test. However per the reviewer's comments we have re-run the analysis using a linear regression model, taking into account age, sex and having a health-care related education or job. The methods and results section have been adjusted accordingly.

#### Methods:

"RPS scores were analysed as described by Meertens.<sup>20</sup> Differences in mean scores were analysed using a linear regression model, adjusting for age, sex and health-care related education or job"

#### Results:

"For CP the financial compensation was most often important (39% important, 52% very important,  $p=0.001$  for comparison between PP and CP), followed by "contributing to science" (33% important, 39% very important,  $p=0.48$ ) and "contributing to developing countries" (46% important, 26% very important,  $p=0.9$ ). The single most important motivation was financial compensation for 41% of CP

and “contributing to science” and “interest in the subject” for 15% each. The single most important factors were not distributed significantly different between PP and CP.”

“CC most often considered the chance of developing symptoms and severity of symptoms important ( $p<0.001$  between PP and CC), with CP also considering the time investment and “an easy way to make money”. The severity of symptoms was the single most important factor (47% for CP, 53% for CN) (Figure 2), which is significantly more often than for PP ( $p<0.001$ ).”

The authors should add as a limitation and possible cofounding variable that 69% of the control group would have not participated in any of the presented CHI trials, as this suggests that participants in the control group are not comparable to previous participants.

One of the reasons we implemented the control group was to compare a group who would participate with a group who would not. This has given us valuable information on reasons not to participate and acceptance of important themes in controlled human infections such as the concept of deliberate infection and the right to withdraw in a population who themselves would not take part. The fact that these controls do not want to participate is an interesting aspect of the control group. We therefore do not consider this to be a limitation of the study. This was not clearly stated in the introduction. We have therefore added the following lines to the introduction:

“An additional benefit of a control group from the general population is there will be a proportion unwilling to participate. These controls provide a comparator in decisional factors and can give information on the acceptance of aspects of controlled human infections even by those unwilling to take part. “

The results section’s grammar should address minor revisions from an English native speaker. A native speaker of English has revised the Results section and has made several adjustments.

In Figure’s 5 legend it should be included what does the error bars and figures estimate represent (standard deviation, median, mean, confidence intervals).

Apologies for this omission, this has been added to the figure legend.

Authors should add a statement that differences between control and previous participants' motivation and decision to participate reasons are merely descriptive as they did not perform a chi-square test (with its proper correction taking into account low expected value and counts) or another applicable statistical test to provide p- values and confidence intervals.

We initially omitted the statistical comparison between the controls and participants, focussing mainly on the descriptive aspects. However as it is clear from the reviewer’s comments that this raises questions we have added the results from the Fisher’s exact analysis to the main different outcomes for motivation and decisional factors.

Methods:

“Differences between CC and PP were calculated using a Fisher’s exact test.”

“For CP the financial compensation was most often important (39% important, 52% very important,  $p=0.001$  for comparison between PP and CP), followed by “contributing to science” (33% important, 39% very important,  $p=0.48$ ) and “contributing to developing countries” (46% important, 26% very important,  $p=0.9$ ). The single most important motivation was financial compensation for 41% of CP and “contributing to science” and “interest in the subject” for 15% each. The single most important factors were not distributed significantly different between PP and CP.”

“CC most often considered the chance of developing symptoms and severity of symptoms important

( $p < 0.001$  for comparison between PP and CC), with CP also considering the time investment and “an easy way to make money”. The severity of symptoms was the single most important factor (47% for CP, 53% for CN) (Figure 2), which is significantly more often than for PP ( $p < 0.001$ ).”