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

TITLE (PROVISIONAL)	The Manitoba Personalized Lifestyle Research (TMPLR) study protocol: a multi-centre bi-directional observational cohort study with administrative health record linkage investigating the interactions between lifestyle and health in Manitoba, Canada
AUTHORS	Mackay, Dylan; Mollard, Rebecca; Granger, Matthew; Bruce, Sharon; Blewett, Heather; Carlberg, Jared; Duhamel, Todd; Eck, Peter; Faucher, Patrick; Hamm, Naomi; Khafipour, Ehsan; Lix, Lisa; Mcmillan, Diana; Myrie, Semone; Ravandi, Amir; Tangri, Navdeep; Azad, Meghan; Jones, PJ

VERSION 1 – REVIEW

REVIEWER	Ala'a Alkerwi, MD, PhD, Principal investigator Luxembourg Institute of Health (LIH), Luxembourg
REVIEW RETURNED	27-Apr-2018

GENERAL COMMENTS	1. The authors pointed out to multi-omic analyses; do they incorporate proteomics and metabolomics technologies in the TMPLR study? And if not, Why? 2. What is the rationale behind selecting this age group (30-46 years) which constitutes only a limited segment (16 years) in the adult age? 3. Would you please justify why choosing this stratification of the BMI(40% <25 and 60% >25 Kg/m2 (overweight or obesity)? 4. Please add the unit for the BMI (line 93) 5. Based on the above comment, it is not convincing that the set of stratification criteria (age, sex, BMI and geography) would help to overcome the non-random sampling bias, as the matched groups are not homogeneous. Please explain. 6. Please provide more information with regards to specific primary hypothesis (line 123). 7. Please define “deep phenotyping” and add reference if possible? Please provide an example 8. Line 163; please indicate the period of follow up and tracking of the participants in the future? 9. Line 166-170; please elaborate more on how “the findings would be able to address why some people are more successful than others to changing their lifestyle”?... “facilitate the design and testing of personalized health promotion strategies” 10. What is the rationale behind choosing a period of 5 years residence in Manitoba? Why not more or less than 5 years? Do the study investigates specific environment-related factors, for example: pollution, exposure to certain pathogens..etc. 11. The sample is mainly patient-based rather than general population sample? Are hospitalized persons recruited or not?

	Please clarify. 12. Line 203, please revised. 13. Please elaborate more on the selection of four, different tools to collect nutritional and dietary habits data, indicating the purpose of each questionnaire. 14. Why fasting blood sampling is done in two days? What are the bio markers measured in the first sample and in the second sample? Please elaborate more in the text of the manuscript. 15. Please regroup the paragraphs concerning dietary assessment (line 299-308) and questionnaire (220-224). It is confusing and redundant to present dietary data in two sections. 16. Line 312; mothers of TMPLR study participants are asked to complete mother questionnaire… Considering the age of the participants (30-46 years), what measures done in case of deceased mothers? 17. Please extricate the data collected via questionnaire than the data retrieved from the PHIN, in two separated paragraphs with sub-headings. The present text is bit confusing. 18. Please indicate the full text for each abbreviation in the first citation. 19. Line 411; what kind of reimbursement is given to the participant? 20. Line 448-450, please provide an example. 21. Please indicate measures taken or planned, in case of potential risk generated in the phase of data collection. 22. Please clearly define all outcomes in a separate paragraph.

REVIEWER	Alexandra Zhernakova University of Groningen, University Medical Center Groningen, Genetics
REVIEW RETURNED	25-Jun-2018

GENERAL COMMENTS	The paper of MakKay et al describes the study protocol of The Manitoba Personalized Lifestyle Research (TMPLR) study, which aims to understand how lifestyle factors interact with each other, genetics and gut microbiome, to influence health. In general, the study is well described and the protocols are clear. Please find my comments below: The authors should realize and acknowledge, that the sample size of 840 participants is rather low for answering the proposed questions. In line with that, and given the fact that several other similar studies have been performed in other countries, it will be valuable to mention these studies and discuss the similarities and potential for data harmonization and cross-replication. The authors start the paper with the description of disease frequencies in the Manitoba population, but it will be informative to add a few sentences to describe the Manitoba population (origin, size, location); and to add how disease frequencies in Manitoba are related to the average disease frequencies in the Canadian population. “After establishing the baseline characteristics of this study cohort, administrative health records will be used retrospectively to examine the developmental origins of health and disease” It is unclear what the authors mean with “the developmental origins” It is not entirely clear why additional participants with reduced

	kidney function were included. The authors say: “because it is expected that very few of the 800 Manitobans who join TMPLR study from the general public will have reduced kidney function….”. Since this is the population study, the frequency of most diseases will be relatively low, so why particularly select non-Manitobans with reduced kidney function, and not with other diseases? Provision of results to participants: what is the policy on providing genetic and microbiome information to the participants? The difference between main factors and additional factors is unclear.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Ala'a Alkerwi, MD, PhD, Principal investigator
Institution and Country: Luxembourg Institute of Health (LIH), Luxembourg
Please state any competing interests or state ‘None declared’: none declared

Please leave your comments for the authors below
1. The authors pointed out to multi-omic analyses; do they incorporate proteomics and metabolomics technologies in the TMPLR study? And if not, Why?
Response: We are measuring the gut microbiome and genetics, we do not currently have a plan to measure proteomics and metabolomics, but are looking into ways to fund and conduct theses analyses.

2. What is the rationale behind selecting this age group (30-46 years) which constitutes only a limited segment (16 years) in the adult age?
Response: The selection of a smaller window of age was due our desire to select a period of time in which we think there is the potential for the development of chronic diseases or the protection from chronic disease based on lifestyle. We also wanted to avoid overlap with the age of the Canadian Longitudinal Study on Aging (45 to 80) which is an ongoing prospective cohort study with a site in Winnipeg, Manitoba. We also wanted to have a younger age range because we are interested in turning this cross-sectional study into a cohort if we can secure funding.

3. Would you please justify why choosing this stratification of the BMI (40% <25 and 60% >25 Kg/m2 (overweight or obesity)?
Response: This stratification is based on the current BMI ranges in the province of Manitoba. See Provincial statistics at
https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310009620&pickMembers%5B0%5D=1.8&pickMembers%5B1%5D=3.1

4. Please add the unit for the BMI (line 93)
Response: We have added the units.

5. Based on the above comment, it is not convincing that the set of stratification criteria (age, sex, BMI and geography) would help to overcome the non-random sampling bias, as the matched groups are not homogeneous. Please explain.

Response: We agree that this stratification is not likely to overcome the non-random sampling bias, we just wanted to end up with a final population that was closer to the overall demographics of the province of Manitoba than if we had not done any stratification.

6. Please provide more information with regards to specific primary hypothesis (line 123).
Response: This study was not set up to look at a specific primary hypothesis. It is an exploratory study which we have highlighted on line 184. Also as stated in line 123, the sample size of TMPLR study was selected based on considerations of feasibility of recruitment, costs, and logistics. It is an exploratory cross-sectional study.

7. Please define “deep phenotyping” and add reference if possible? Please provide an example
Response: We have changed the deeply to extensively, we are referencing the extensive testing that each participant undergoes, including gut microbiome, actigraphy (sleep and physical activity), clinical chemistry and other blood biomarkers, genetics, dietary intake assessments, and anthropometric measures including DXA.

8. Line 163; please indicate the period of follow up and tracking of the participants in the future?
Response: We have now specified that the follow up using the administrative health records will start after 5 years.

9. Line 166-170; please elaborate more on how “the findings would be able to address why some people are more successful than others to changing their lifestyle”?... “facilitate the design and testing of personalized health promotion strategies”
Response: We have added an example to the end of the introduction “For example, if we are able to identify interactions between lifestyle factors and disease risk, such as a genetic variant that associates with poor health measure with short sleep, a study could be designed looking to improve sleep hygiene specifically in the group with the risk variant. “ We have deleted the line related to addressing why “some people are more successful”

10. What is the rationale behind choosing a period of 5 years residence in Manitoba? Why not more or less than 5 years? Do the study investigates specific environment-related factors, for example: pollution, exposure to certain pathogens..etc.
Response: We were hoping that a 5-year time window would provide a minimum of 5-years of administrative health data and assure a certain connection to Manitoba to increase the likelihood of participants still living in Manitoba for the prospective follow-up via administrative health records. The number of years was determined by the investigators as a balance between not excluding too many participants who wanted to participate, while establishing the participants connection to Manitoba.

11. The sample is mainly patient-based rather than general population sample? Are hospitalized persons recruited or not? Please clarify.
Response: The sample is not patient based, they are recruited from the general population. Except in the case of the sub-set of individuals with ESRD, who are recruited from the Manitoba Renal Clinic. No hospitalized patients are recruited.

12. Line 203, please revised.
Response: We have removed the double words from fixed line 203

13. Please elaborate more on the selection of four, different tools to collect nutritional and dietary habits data, indicating the purpose of each questionnaire.
Response: The diet history questionnaire is a food frequency questionnaire which we selected to get dietary pattern of a participant’s intakes over the last year. The 3 24-hr recalls obtain a more acute snapshot of three days of participant food intake. The Three factor eating does not capture intake

data but measures a participant's dietary restraint, disinhibition and hunger in relation to eating. The mindful eating questionnaire also does not measure dietary intake but measures a participant's awareness of the physical and emotional sensations associated with eating.

14. Why fasting blood sampling is done in two days? What are the bio markers measured in the first sample and in the second sample? Please elaborate more in the text of the manuscript.
Response: Blood samples were collected on two consecutive days for the measurement of cholesterol and fatty acid synthesis using the stable isotopic deuterium incorporation method. Having two days sampling also allowed us a 2nd chance to get a blood sample from participants who may not have been able to provide blood on the 1st day due to issues with phlebotomy.

15. Please regroup the paragraphs concerning dietary assessment (line 299-308) and questionnaire (220-224). It is confusing and redundant to present dietary data in two sections.
Response: We have combined all the dietary assessment into one section as recommended.

16. Line 312; mothers of TMPLR study participants are asked to complete mother questionnaire… Considering the age of the participants (30-46 years), what measures done in case of deceased mothers?
Response: If a mother is deceased then we do not obtain that data set, however, participants are asked to complete a childhood retrospective questionnaire which contains some of the questions we ask the mothers. We want to validate the self reported answers from the childhood questionnaire against the mother's questionnaire in those who have both completed.

17. Please extricate the data collected via questionnaire than the data retrieved from the PHIN, in two separated paragraphs with sub-headings. The present text is bit confusing.
Response: We have added "Administrative health data will provide method of birth, gestational age, birth weight, diagnosis codes for post-delivery hospitalization, and post-delivery drug prescriptions."

18. Please indicate the full text for each abbreviation in the first citation.
Response: I have added the full text for PHIN on the first use in the manuscript

19. Line 411; what kind of reimbursement is given to the participant?
Response: The full remuneration for study participation is $100 Canadian dollars provided as cash or as a gift card.

20. Line 448-450, please provide an example.
Response: It is not clear to us what type of example is requested.

21. Please indicate measures taken or planned, in case of potential risk generated in the phase of data collection.
Response: It is not clear what information is being requested here.

22. Please clearly define all outcomes in a separate paragraph.
Response: It is not clear what information is being requested here.

Reviewer: 2
Reviewer Name: Alexandra Zhernakova
Institution and Country: University Medical Center Groningen, Department of Genetics, CB50, Building 3211; PO Box 30001, 9700 RB Groningen; The Netherlands
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
The paper of MacKay et al describes the study protocol of The Manitoba Personalized Lifestyle Research (TMPLR) study, which aims to understand how lifestyle factors interact with each other, genetics and gut microbiome, to influence health. In general, the study is well described and the protocols are clear. Please find my comments below:

The authors should realize and acknowledge, that the sample size of 840 participants is rather low for answering the proposed questions. In line with that, and given the fact that several other similar studies have been performed in other countries, it will be valuable to mention these studies and discuss the similarities and potential for data harmonization and cross-replication.
Response: We have included a section in the discussion “Given a projected sample size of 840 participants may be low for some of research questions that will be investigated, therefore harmonization and linking of data across multiple cohorts may be required. We will be looking to other studies which have undertaken overlapping measurements in order to increase sample sizes. The Canadian Longitudinal Study on Aging , the Toronto Nutrigenomics and Health, and The LifeLines DEEP studies among others will be approached regarding the potential of data harmonization and cross-replication. TMPLR study will also be available to other researchers who are interested in collaboration or using the data for cross-replication.” We have also included the line “We will be looking to collaborate with other existing studies with overlapping measures to replicate such findings, or increase sample size.” to the introduction.
The authors start the paper with the description of disease frequencies in the Manitoba population, but it will be informative to add a few sentences to describe the Manitoba population (origin, size, location); and to add how disease frequencies in Manitoba are related to the average disease frequencies in the Canadian population.

Response: We have added some details on the Manitoba population and disease frequencies relative to Canada to the introduction

“After establishing the baseline characteristics of this study cohort, administrative health records will be used retrospectively to examine the developmental origins of health and disease” It is unclear what the authors mean with “the developmental origins”
Response: We are referring to the “developmental origins” in the context of the Developmental Origins of Health and Disease”. Which relates to the evidence that events occurring in the early stages of human development may influence the occurrence of chronic disease like diabetes, cardiovascular disease. We have added a reference in order to clarify what we mean.

It is not entirely clear why additional participants with reduced kidney function were included. The authors say: “because it is expected that very few of the 800 Manitobans who join TMPLR study from the general public will have reduced kidney function….”. Since this is the population study, the frequency of most diseases will be relatively low, so why particularly select non-Manitobans with reduced kidney function, and not with other diseases?
Response: As you have suggested we have added details in the introduction about Manitoba, which identify Manitoba as having the highest rate of end stage renal disease in Canada. We have selected more participants with reduced kidney function to because of Manitoba’s uniquely poor kidney health status. We have also made it clearer that the patients with reduced kidney function will also be from Manitoba, not non-Manitobans.

Provision of results to participants: what is the policy on providing genetic and microbiome information to the participants?
Response: We are not providing genetic or microbiome information to the participants. We have added this to the manuscript.

The difference between main factors and additional factors is unclear.
Response: It is not clear what information is being requested here.

VERSION 2 – REVIEW

REVIEWER	Ala'a Alkerwi Centre de Recherch Public-Santé, Public Health
REVIEW RETURNED	03-Sep-2018

GENERAL COMMENTS	The authors would still need to define the outcomes of the study as requested by the check list of journal. Reading the manuscript, it is not clear at which stage the study is! does the data collection phase is already finished or on-going? therefore, it would be interesting to precise the date and duration of data collection, also to mention any precaution measures to be taken in case of inability to collect the data. If the journal is convinced with the answers, it will be fine for me to accept the article for publication.

REVIEWER	Alexandra Zhernakova University of Groningen, University Medical Center Groningen, Genetics
REVIEW RETURNED	03-Sep-2018

GENERAL COMMENTS	The authors replied to all my comments, except for the last one, which is minor: several times in the paper the authors mention "additional" measurements; for or example: "Lifestyle factors assessed will include dietary pattern, physical activity, cardiovascular fitness and sleep. Additional factors such as medical history, socio-economic status, alcohol and tobacco consumption, cognition, stress and anxiety, and early life experiences will also be documented." It was not entirely clear what the definition of "additional" factors is, and if they will be collected in all participants.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Ala'a Alkerwi

Institution and Country: Luxembourg Institute of Health, Department of Population Health, 1A-B, rue Thomas Edison, L-1445 Strassen, Luxembourg Please state any competing interests or state ‘None declared’: None declared

The authors would still need to define the outcomes of the study as requested by the check list of journal.

- We have highlighted that we did not specifically power out study for a specific outcome and therefore this study is exploratory for hypothesis generation. This is in the bullet points line 126, and in the design line 184. We also discuss what our power for specific outcomes are on line 215 and in Table 2.

Reading the manuscript, it is not clear at which stage the study is! does the data collection phase is already finished or on-going? therefore, it would be interesting to precise the date and duration of data collection, also to mention any precaution measures to be taken in case of inability to collect the data.

If the journal is convinced with the answers, it will be fine for me to accept the article for publication.

- Data collection started in March 2016 and is still ongoing, this is outlined in the study status section on line 461. In regards to the inability to collect data from certain participants, we have updated the statistical analyses section to state “Non-response bias or inability to collect certain data, may affect the validity of analyses for survey data or biological measures, necessitating the use of multiple imputation methods if the pattern of missing data is deemed to be ignorable [60].” Line 377

Reviewer: 2

Reviewer Name: Alexandra Zhernakova

Institution and Country: University Medical Center Groningen, Department of Genetics, CB50, Building 3211; PO Box 30001, 9700 RB Groningen; The Netherlands Please state any competing interests or state ‘None declared’: None declared

The authors replied to all my comments, except for the last one, which is minor: several times in the paper the authors mention "additional" measurements; for or example: "Lifestyle factors assessed will include dietary pattern, physical activity, cardiovascular fitness and sleep. Additional factors such as medical history, socio-economic status, alcohol and tobacco consumption, cognition, stress and anxiety, and early life experiences will also be documented." It was not entirely clear what the definition of "additional" factors is, and if they will be collected in all participants.

- We have removed the word “additional” which was just a language choice and not a classification of measures that are only collected in a subset. All measured are planned to be collected in all participants.