

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)	Cohort Profile: The ChicagO Multiethnic Prevention and Surveillance Study (COMPASS)
AUTHORS	Aschebrook-Kilfoy, Briseis; Kibriya, Muhammad; Jasmine, Farzana; Stepniak, Liz; Gopalakrishnan, Rajan; Craver, Andrew; Zakin, Paul; Tasmin, Saira; Kim, Karen; Goss, Kathleen; List, Marcy; LeBeau, Michelle; Ahsan, Habibul

VERSION 1 - REVIEW

REVIEWER	Ivo Muskens MD Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California.
REVIEW RETURNED	21-Apr-2020

GENERAL COMMENTS	<p>The authors present their extensive efforts in establishing a prospective cohort study in Chicago (COMPASS study) that aims to evaluate health outcomes in a population that has great heterogeneity regarding ethnicity, SES, and access to healthcare (amongst many other things). Although we know that these factors greatly influence health outcomes across the United States, much remains unknown and a better understanding is warranted and the authors are to be applauded for their efforts. It is great that they present a description of their study before publishing results, which will help transparency in reporting. I have some minor suggestions/questions for the authors to consider.</p> <p>Minor suggestions/questions:</p> <ul style="list-style-type: none">- The study is not registered at clinicaltrials.gov for instance. Do the authors have a reason for not doing so?- There are currently some other studies in the US that aim to evaluate healthcare outcomes by SES and ethnicity. Do the authors envision a collaboration with other multiethnic studies? This could for instance increase power in some analyses where the incidence of disease is low.- Why did the authors choose to only include inhabitants of Chicago older than 35? Some of the factors outlined by the authors might impact health outcomes at an earlier age (e.g. gun violence).- How long do the authors intend to perform follow-up on the participants?- The description of follow-up data collection is not very extensive and could perhaps benefit from a more extensive description. Especially follow-up data is often hard to collect due to drop out of participants. How do the authors envision to prevent this dropout? Also, as some participants were already included many years ago,
-------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	<p>how has the follow-up been thus far?</p> <p>- The authors have been very successful in involving the African American community. However, the Latino population appears to be underrepresented. Do the authors have some explanations for this or plan to involve more people from the Latino community?</p>
--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

REVIEWER	Victoria Cairnduff Queen's University Belfast, UK
REVIEW RETURNED	08-May-2020

GENERAL COMMENTS	<p>A well composed paper describing a very important study which will be provide an invaluable insight into health disparities in a traditionally hard to reach cohort.</p> <p>- The introduction frames the background and rationale for the study well, however it lacks evidence from the literature to back-up some of the statements, particularly sentences in lines 22-25, 25-26 and 49-53 on page 4 which I feel would benefit from being backed up by references.</p> <p>- It might be useful to move the eligibility criteria to earlier within the methodology section and I was wondering if a previous cancer diagnosis was considered when recruiting participants? Also it might be useful to briefly include the rationale for recruiting patients aged over 35 years for readers who are less familiar with the study.</p> <p>- It would be useful to have more information of the topics covered with the questionnaire in the main text and also would be useful to include information on any validated tools/questions e.g. EQ5D etc that were included in the questionnaire, particularly those that were used to assess anxiety, depression, stress, physical activity, self-reported health, barriers to health care access and pain status.</p> <p>- The figure of recruitment, exposure assessment, early detection, screening interventions etc. (page 21) needs a title.</p> <p>- paragraph 4 page 10 contains lots of information on what blood samples will be collected. If possible it would be useful to have more information on what these samples will be used to measure.</p> <p>- Will linkage to EMR (described in paragraph 3 page 12) take place at different time points or will this be done continuously?</p> <p>- Also when is follow-up (described in paragraph 4 page 12) planned to take place?</p> <p>- A flowchart showing the different stages of the study and when linkage to EMR will take would be really useful to help the reader follow the study plan.</p>
-------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 edits

2) The study is not registered at clinicaltrials.gov for instance. Do the authors have a reason for not doing so?

Thank you for this suggestion. We have now submitted COMPASS now for inclusion in clinicaltrials.gov.

3) There are currently some other studies in the US that aim to evaluate healthcare outcomes by SES and ethnicity. Do the authors envision a collaboration with other multiethnic studies? This could for instance increase power in some analyses where the incidence of disease is low.

Yes, we would welcome collaboration with other multiethnic studies.

4) Why did the authors choose to only include inhabitants of Chicago older than 35? Some of the factors outlined by the authors might impact health outcomes at an earlier age (e.g. gun violence).

We chose 35 for efficiency to study cancer outcomes (we are funded by our cancer center primarily). Most cancer studies begin recruitment at 40 or 45 and we tried to balance efficiency with exposure ascertainment as pointed out by the reviewer.

5) How long do the authors intend to perform follow-up on the participants?

We note in the consent that we plan to follow study participants for at least 10 years. This detail has been added as the last sentence in the paragraph on follow-up.

6) The description of follow-up data collection is not very extensive and could perhaps benefit from a more extensive description. Especially follow-up data is often hard to collect due to drop out of participants. How do the authors envision to prevent this dropout? Also, as some participants were already included many years ago, how has the follow-up been thus far?

We have been limited by funding to conduct follow-up on all participants. This is certainly a goal of the project. We have enhanced our data collection infrastructure such that we may text follow-up surveys to participants. In a recent effort to collect follow-up data, we were able to reach 25% of participants by phone. In the future, we will reach out via text and phone to enhance response rates. This is noted in the follow-up paragraph.

7) The authors have been very successful in involving the African American community. However, the Latino population appears to be underrepresented. Do the authors have some explanations for this or plan to involve more people from the Latino community?

Yes, there are many reasons for the lower recruitment in Hispanic communities. 1) we have oversampled persons in the communities surrounding the UCMC, which is predominantly African American; 2) our clinic-based recruitment does not reach Hispanics as less than 5% percent of our patient population is Hispanic compared to over 60% African American; 3) We have 2 field staff who are native Spanish speakers to accommodate any language barriers and surveys in Spanish, but the majority of our field staff are African American; 4) our engagement infrastructure has largely leveraged community partners in the African American community, however, a pilot project was awarded to a junior investigator to enhance community partnership in Hispanic communities and enhance Hispanic recruitment; 5) the response rate in Hispanic communities was 3X lower than in African American communities when field staff were seeking participation going door to door. The paper by Press details the difference in response rates by race/ethnicity.

Reviewer: 2 edits

8) The introduction frames the background and rationale for the study well, however it lacks evidence from the literature to back-up some of the statements, particularly sentences in lines 22-25, 25-26 and 49-53 on page 4 which I feel would benefit from being backed up by references.

References have been added to these passages as needed, and the wording has been revised for clarity.

9) It might be useful to move the eligibility criteria to earlier within the methodology section and I was wondering if a previous cancer diagnosis was considered when recruiting participants? Also it might be useful to briefly include the rationale for recruiting patients aged over 35 years for readers who are less familiar with the study.

Eligibility criteria have been moved above “Targeted community-based recruitment” section. A note has been added to criteria 2 to clarify that Age 35 at the point of contact was to enhance efficiency to obtain cancer outcomes.

10) It would be useful to have more information of the topics covered with the questionnaire in the main text and also would be useful to include information on any validated tools/questions e.g. EQ5D etc that were included in the questionnaire, particularly those that were used to assess anxiety, depression, stress, physical activity, self-reported health, barriers to health care access and pain status.

We conducted an extensive review of questionnaires used in other large cancer studies, including the American Cancer Societies Cancer Prevention Study and National Cancer Institute cohorts (including PLCO, NIH-AARP, and the Agricultural Health Study) in addition to the NHANES questionnaires, such that we would be able to harmonize our data in cohort consortium collaboration and maximize our ability to enhance study power for key diseases and populations. Our data dictionary can be found at compass.uchicago.edu. This detail has been added to the Data Collection section.

11) The figure of recruitment, exposure assessment, early detection, screening interventions etc. (page 21) needs a title.

It is titled as “Figure 1. COMPASS Scientific Focus Area”

12) paragraph 4 page 10 contains lots of information on what blood samples will be collected. If possible it would be useful to have more information on what these samples will be used to measure.

The lavender top collection accommodates assays of whole blood DNA, PBMC – RNA, lymphocyte markers, cytokines, cardiac/CVD markers, metabolic panels, cancer panels, and investigation of viral & bacterial load (PCR). The green top collection can accommodate assays including thyroid parameters, drug screening, and other biochemical tests. The gold top collection accommodates a variety of hormone assays, lithium, iron, vitamin B12, folate, immunoglobulins, autoantibodies and antibiotic assays. The blue top tubes can accommodate a variety of environmental measurements, including some metals. These details have been added after the description of blood collection approaches.

13) Will linkage to EMR (described in paragraph 3 page 12) take place at different time points or will this be done continuously?

We have added the following statement to address this point. The collection of medical records will be completed continuously, when available.

14) Also when is follow-up (described in paragraph 4 page 12) planned to take place?

Clarification has been added to note “Follow-up survey data collection will be requested annually, and frequency of in-person biospecimen collection will be funding dependent.”

15) A flowchart showing the different stages of the study and when linkage to EMR will take would be really useful to help the reader follow the study plan.

As noted in the follow-up section in response to Q13, EMR addition will be continuous and follow-up will occur regularly for some parameters but in subsets for other (environmental data, nutrition data, activity data from an app, etc). We considered a flowchart but thought that it would be too difficult to add the varying contexts and data collection over time. That it would overgeneralize the plan to be visually appealing. Unless the editor disagrees, we would prefer not to add such a chart.

VERSION 2 – REVIEW

REVIEWER	Victoria Cairnduff Centre for Public Health Queen's University Belfast N.Ireland
REVIEW RETURNED	03-Jul-2020

GENERAL COMMENTS	All my comments have been addressed. Thank you.
-------------------------	-------------------------------------------------