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.

This paper was submitted to a another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Diabetes care cascade and associated factors in 10,700 middle-aged adults in four sub-Saharan African countries-a cross-sectional study
AUTHORS	Wade, Alisha N.; Maposa, Innocent; Agongo, Godfred; Asiki, G; Boua, Palwende; Choma, Solomon SR; Gómez-Olivé, F. Xavier; Maimela, Eric; Micklesfield, Lisa; Mohamed, Shukri; Nonterah, Engelbert; Norris, Shane; Sorgho, Hermann; Ramsay, Michele; Crowther, Nigel John

VERSION 1 – REVIEW

REVIEWER	Jorgensen , Jutta M Adelin University of Copenhagen Department of Public Health
REVIEW RETURNED	01-Nov-2022

GENERAL COMMENTS	Thank you for letting me review this interesting paper. Kindly find my general and specific comments below in random order. COMMENT Key messages section should be rewritten elaborating on what is already know on this topic COMMENT Referring to progressing through the care cascade, I suggest to replace 'awareness' with 'been diagnosed' as you are referring to an individual's awareness of having been diagnosed with diabetes. At p 24 line 15 'awareness' seems to be used to describe the populations' awareness of a condition called diabetes, not having been diagnosed with it. COMMENT You will need to do a review of existing literature and update the text and references with newer literature - for instance, Jennifer Manne-Goehler et al (2018) and David Flood (2021) are investigating the diabetes care cascade disaggregated on world regions including SSA, using nationally representative cross-sectional data from DHS and STEPS surveys. COMMENT

	P 17 line 22 results referring to which table? COMMENT P 18 table 2 / line 7: What are the categories of age that is used in the regression analysis? The reference category should appear here COMMENT The variable 'HIV status' is a bit tricky since its definition vary between the study sites (diagnosed with HIV in some sites versus known with HIV or positive HIV-rapid test at other sites) - to avoid unclarity perhaps you should change the variable to read 'known with HIV' as this is what the variable measures, not whether they had HIV or not. COMMENT I wonder how the decision was reached to exclude the Soweto sample which makes up close to a quarter of the participants rather than excluding the variable 'Family history of diabetes' from the analysis - could you elaborate or consider running the analyses including Soweto? Would that give power to do more steps in the care cascade? COMMENT P 21 Table 3 / line 39: What is the reference category for BMI? Should appear here COMMENT P 23 line 24-34: to be revised after review of newer literature on diabetes care cascade in SSA COMMENT The diabetes treatment variable does not differentiate about current treatment or ever been on treatment, which is different from the definition of treatment used in eg WHO STEPS surveys. If possible to differentiate between those who currently are on dietary or medical treatment (= within last two weeks used antiglycemic treatment) and re-run the analysis with the new variable it would be preferred. If not, then perhaps 'ever received treatment for diabetes' is a more precise description than 'receiving treatment for diabetes'. COMMENT STROBE checklist could be included in the submission - however it seems like it has been followed in the manuscript.

REVIEWER	National Institute of Public Health, Health Policy and Systems Research Unit
REVIEW RETURNED	08-Nov-2022

GENERAL COMMENTS	This study on the diabetes cascade of care is useful for evaluation of health system performance in provision of care continuum for diabetes. Although the study objectives are clearly defined, the study should have investigated factors associated with the undiagnosed status (being unaware of the condition). The undiagnosed people are at higher risk of developing complications faster as they do not seek care promptly. The system should in fact provide targeted screening with the population groups who are usually unaware of the condition.

	The study design (cross-sectional) was appropriate to capture leakages of the care continuum and identify associated factors with being diagnosed (awareness), but for factors associated with the 'likelihood of having diabetes' the design should be an observational study. Regarding the study methods, there are several points need to be clarified: 1. What is the justification for the age limit (up to 60 only)? 2. What is the justification for including HIV status as an important variable for the analysis? According to the study, that required additional effort and resource for the rapid HIV testing, and in Ghana and Burkina Faso an assumption was geared toward negative status. 3. There was a claim on using standardized protocols, but in fact study procedures could not be carried out exactly the same across the six sites. What does it mean by saying 'harmonised data'? 4. What had been done to ensure the quality of data collected for the laboratory testing? From blood sample collection at the site to the analysis at the laboratory, any error at any stage could increase the unreliability of the data. This includes use of validated tools--no proper reference was given for each tool mentioned in the data collection procedure. Was the abdominal ultrasound performed on every participant across the six sites? 5. The use of either fasting plasma glucose or random plasma glucose to give the diagnosis should be justified with strong evidence. Since there was collection of venous blood samples for the analysis at a single laboratory, why wasn't it confirmed with HbA1c test? 6. Similar to point 5, the use of fasting plasma glucose to determine the optimal glycaemic control is another critical point that needs to be justified by strong evidence. Clinically, HbA1c is the recommended indicator. I am not quite convinced by the explanation in the limitation. 7. Regarding 'receiving the treatment', there was no indication of the time frame (in the past two weeks?) A study conducted by Stokes et al. on 'Prevalence and unmet need for diabetes care across the care continuum in a national sample of South African adults: evidence from the SANHANES-1, 2011-2012' showed that the reported prevalence was 10.1% among adults aged 15 and above. In this study, the reported prevalence was 5.5% among adults aged 40-60. My assumption is that the diabetes prevalence is likely to be higher with a sample concentrated on older groups of population. It would be clearer to see the different cascades for each site. In the discussion, the findings in other studies might not be based on the same definitions. Please pay more attention to the definitions. I hope the above comments would help improve the work.

REVIEWER	Tiwari, Swasti Sanjay Gandhi Post Graduate Institute of Medical Sciences
REVIEW RETURNED	09-Nov-2022

GENERAL COMMENTS	The global burden of diabetes follows an increasing trend and the authors have timely conceptualized the study highlighting the current scenario of diabetes care and management in Sub-

	Saharan African countries. However, there are the following major issues that need to be addressed by the authors: 1 The authors should mention the guidelines used to define diseases as well as other measures of metabolic syndrome. Single guideline for each item should be followed throughout the manuscript. 2 Introduction: “Sub-Saharan Africa……by 2045”. Reframe the sentence to write in terms of counts. 3 The care and management of diseases is different in High income countries (HICs) and low or middle income countries (LMICs). Is there any evidence for the application of care cascade model in LMICs? 4 The authors should report clear objective of the study. Was the care cascade studied regionally? Were the factors affecting care continuum identified regionally? Does the study focus on males and females separately? 5 The title of the study should clearly state the objective of the study. 6 Methodology including statistical analysis is confusing and difficult to follow. The section should be rewritten for a clear understanding of the study. 7 Study setting and participants: replace the term ‘Twenty+’. Mention the inclusion/exclusion criteria. Was HIV the inclusion criteria? Why only HIV was considered? 8 Statistical analysis: a) Mention the names of categorical characteristics/variables, and continuous characteristics/variables. Why were variables compared by sex and missingness? b) ‘Multivariable logistic regression……previous research’ is not clear and easy to follow. c) ‘Stata’ is the software and not STATA d) How were the variables selected for multivariable logistic regression? 9 Results: The section is not easy to understand. It should be scientifically and grammatically correct. Follow a similar pattern while reporting the results. P values should be reported to 2 decimal places. The percentage should be written along with CI. 10 Graphically represent the care cascade for better clarity 11 Discussion: The authors should comprehensively discuss the care cascade approach used in various LMICs to study diabetes/hypertension care. How the care cascade (each component separately and overall) is different from other countries (LMICs from other continents) to guide policymakers of SSA. 12 The authors should also discuss the probable causes behind the differences in the care cascade of diabetes

VERSION 1 – AUTHOR RESPONSE

Reviewer comments

1. Key messages section should be rewritten elaborating on what is already know on this topic. This section has been deleted in accordance with the editorial comments.

2. Referring to progressing through the care cascade, I suggest to replace 'awareness' with 'been diagnosed' as you are referring to an individual's awareness of having been diagnosed with diabetes. At p 24 line 15 'awareness' seems to be used to describe the populations' awareness of a condition called diabetes, not having been diagnosed with it.

We thank the reviewer for this comment and agree that we are referring to an individual's awareness of a diagnosis of diabetes as defined on pages 10- 11 of the manuscript. We believe "awareness" is a more appropriate term than "been diagnosed" in the latter parts of the manuscript as we define the term and are specifically interested in knowing whether an individual with diabetes knows that they have the disease. In these environments of low health literacy, it is possible that an individual may have a diagnosis of diabetes documented in clinical records but be unaware of this and therefore "awareness" and "been diagnosed" measure two different constructs.

We have reworded the sentence highlighted by the reviewer to clarify its meaning (page 21).

3. You will need to do a review of existing literature and update the text and references with newer literature - for instance, Jennifer Manne-Goehler et al (2018) and David Flood (2021) are investigating the diabetes care cascade disaggregated on world regions including SSA, using nationally representative cross-sectional data from DHS and STEPS surveys.
   We thank the reviewer for drawing our attention to these manuscripts and have updated our discussion (page 19) and references (included as references 2 and 3) accordingly.

4. P 17 line 22 results referring to which table?
   This refers to Table 1 and this has been clarified in the text (page 15).

5. P 18 table 2 / line 7: What are the categories of age that is used in the regression analysis? The reference category should appear here
   Age was entered as a continuous variable. This has been clarified in the table legend and text (pages 11, 16&17).

6. The variable 'HIV status' is a bit tricky since its definition vary between the study sites (diagnosed with HIV in some sites versus known with HIV or positive HIV-rapid test at other sites) - to avoid unclarity perhaps you should change the variable to read 'known with HIV' as this is what the variable measures, not whether they had HIV or not.
   We have made this change throughout the manuscript in line with the reviewer's suggestion.

7. I wonder how the decision was reached to exclude the Soweto sample which makes up close to a quarter of the participants rather than excluding the variable 'Family history of diabetes' from the analysis - could you elaborate or consider running the analyses including Soweto? Would that give power to do more steps in the care cascade?
   Soweto was excluded from the latter parts of the cascade i.e. ever receiving treatment and glycaemic control as the treatment variable was not collected at that site and the percentage achieving glycaemic control was calculated as a proportion of those receiving treatment. We have included this explanation in the methods (pages 11- 12) and clarified the explanation in the figure legend. The Soweto site was also excluded from the analysis of the factors associated with the odds of having diabetes. This was done as family history of diabetes has previously been demonstrated to be strongly associated with having diabetes, but this variable was not collected in Soweto (page 12).

   Of note, we do provide the diabetes care cascade per site in Table S2.

8. P 21 Table 3 / line 39: What is the reference category for BMI? Should appear here
   Body mass index was entered as a continuous variable. This has been clarified in the table legend and text (page 11).

9. P 23 line 24-34: to be revised after review of newer literature on diabetes care cascade in SSA
   This section has been revised as above.

10. The diabetes treatment variable does not differentiate about current treatment or ever been on treatment, which is different from the definition of treatment used in eg WHO STEPS surveys. If possible to differentiate between those who currently are on dietary or medical treatment (= within last two weeks used antiglycemic treatment) and re-run the analysis with

the new variable it would be preferred. If not, then perhaps 'ever received treatment for diabetes' is a more precise description than 'receiving treatment for diabetes'.
On pages 10-11, we define our treatment variable as ever having received treatment. We have revised the terminology throughout the manuscript to ensure it is consistent with this definition.

11. STROBE checklist could be included in the submission - however it seems like it has been followed in the manuscript
The STROBE checklist was uploaded with the original manuscript and should be available to the reviewer.

12. This study on the diabetes cascade of care is useful for evaluation of health system performance in provision of care continuum for diabetes. Although the study objectives are clearly defined, the study should have investigated factors associated with the undiagnosed status (being unaware of the condition). The undiagnosed people are at higher risk of developing complications faster as they do not seek care promptly. The system should in fact provide targeted screening with the population groups who are usually unaware of the condition.
We thank the reviewer for acknowledging the utility of our study. We draw their attention to the fact that we did indeed investigate and report factors associated with awareness of diabetes. We have added another sentence to the first paragraph of our discussion, summarising these findings (pages 17-18).

13. The study design (cross-sectional) was appropriate to capture leakages of the care continuum and identify associated factors with being diagnosed (awareness), but for factors associated with the 'likelihood of having diabetes' the design should be an observational study.
We agree that an observational study is appropriate for investigating factors associated with the likelihood of having diabetes. We draw the reviewer's attention to the fact that cross-sectional studies, such as the one we performed, are a subtype of observational study.

14. What is the justification for the age limit (up to 60 only)?
We selected individuals aged 40-60 for this study as this is a peak time for development for cardiometabolic diseases such as diabetes. We have included this point in our methods section (page 7).

15. What is the justification for including HIV status as an important variable for the analysis? According to the study, that required additional effort and resource for the rapid HIV testing, and in Ghana and Burkina Faso an assumption was geared toward negative status.
HIV prevalence is known to be high in four of our six sites. HIV positivity was an important variable in our analyses of factors associated with prevalence and awareness of diabetes given existing literature that suggests individuals with HIV are likely to have lower cardiometabolic disease risk and that individuals with HIV may have more contact with the health care system and therefore may be more aware of other chronic diseases. We have discussed these points and referred to the relevant literature in our discussion section (page 18&19; references 21, 22&25).

16. There was a claim on using standardized protocols, but in fact study procedures could not be carried out exactly the same across the six sites. What does it mean by saying 'harmonised data'?
We use the term "harmonised data" to illustrate that the same methods were used across all sites to collect a particular variable. In addition, the same make and model of ultrasound was used at all sites to measure visceral and subcutaneous fat (page 10). All field workers taking these measurements at all sites were centrally trained in Johannesburg (page 10). Furthermore, all glucose assays were performed at a central laboratory, also in Johannesburg (page 10). We do acknowledge that not all variables were collected at all sites and have reported where this was the case.

17. What had been done to ensure the quality of data collected for the laboratory testing? From blood sample collection at the site to the analysis at the laboratory, any error at any stage

could increase the unreliability of the data. This includes use of validated tools--no proper reference was given for each tool mentioned in the data collection procedure. Was the abdominal ultrasound performed on every participant across the six sites?
In our original manuscript, we detailed the procedure for the collection, handling and storage of the venous blood samples. We further provided details of the detection range and coefficient of variation of the assay that was used to analyse venous blood samples for glucose (page 10). As stated in the preceding comment, all glucose assays were performed at a central lab in Johannesburg (page 10). All sites were provided with a detailed protocol for taking and storing the blood samples and we have now noted this in our methods section (page 10). Similarly, we have provided the details for the rapid HIV tests used as well as the manufacturers for all the equipment used in the study, as is standard reporting practice (pages 9-10). The reference for the instrument used to collect physical activity data was provided (page 9). As described in the manuscript, we used pairwise deletion in our analysis and the individuals with missing ultrasound data were not included in analyses including the visceral and subcutaneous fat measures.

We draw the reviewer's attention to the methods section of the manuscript which we believe addresses the concerns raised.

18. The use of either fasting plasma glucose or random plasma glucose to give the diagnosis should be justified with strong evidence. Since there was collection of venous blood samples for the analysis at a single laboratory, why wasn't it confirmed with HbA1c test?
Either fasting or random plasma glucose are accepted by the American Diabetes Association and the World Health Organisation for the diagnosis of diabetes. We had already included the American Diabetes Association guidelines, but have now also included reference to the World Health Organisation guidelines and placed both of these references more prominently (page 10; references 11&12). The use of HbA_{1c} for the diagnosis of diabetes in African-ancestry individuals may overestimate the prevalence of diabetes as noted in reference 28.

19. Similar to point 5, the use of fasting plasma glucose to determine the optimal glycaemic control is another critical point that needs to be justified by strong evidence. Clinically, HbA1c is the recommended indicator. I am not quite convinced by the explanation in the limitation.
We acknowledge that HbA_{1c} is the clinical standard of care for determining glycaemic control in high income countries. However, a target fasting plasma glucose of 7.2 mmol/l is also cited by the American Diabetes Association and this has been referenced (page 11; reference 11). Several large epidemiological studies have also used glucose measures rather than HbA_{1c} to assess glycaemic control and we have now included this in our discussion (page 21; references 2&3).

As noted by the reviewer, we do acknowledge the limitations of this approach in our discussion (page 21).

20. Regarding 'receiving the treatment', there was no indication of the time frame (in the past two weeks?)
This has been clarified throughout the manuscript to indicate that we are referring to ever having received treatment.

21. A study conducted by Stokes et al. on 'Prevalence and unmet need for diabetes care across the care continuum in a national sample of South African adults: evidence from the SANHANES-1, 2011-2012' showed that the reported prevalence was 10.1% among adults aged 15 and above. In this study, the reported prevalence was 5.5% among adults aged 40-60. My assumption is that the diabetes prevalence is likely to be higher with a sample concentrated on older groups of population. It would be clearer to see the different cascades for each site. In the discussion, the findings in other studies might not be based on the same definitions. Please pay more attention to the definitions.
We thank the reviewer for this observation. Cascades for each site were provided in Table S2 in the supplementary material. Additionally, in the cited NHANES study, diabetes was diagnosed using HbA_{1c} which, as noted previously, may overestimate diabetes prevalence. In the meta-analyses we have cited, the included studies used varying definitions

of diabetes, a limitation we have now noted on page 20. We have therefore confined our comparisons to the overall composite findings of the meta-analyses.

22. The authors should mention the guidelines used to define diseases as well as other measures of metabolic syndrome. Single guideline for each item should be followed throughout the manuscript.
    We have further clarified the guidelines which have been used to define diabetes and hypertension (pages 9&10; references 9, 11&12). The two diabetes guidelines are consistent with each other and we have adhered to them throughout the manuscript.

23. Introduction: "Sub-Saharan Africa……by 2045". Reframe the sentence to write in terms of counts.
    This has been changed (page 6). The reference (reference 1) has also been updated to the most recent version of the IDF Atlas, with a consequent change in the cited mortality data (page 6).

24. The care and management of diseases is different in High income countries (HICs) and low or middle income countries (LMICs). Is there any evidence for the application of care cascade model in LMICs?
    While we agree that disease care and management differ between HIC and LMIC, the cascade of care model has universal application. We have however also included references to the use of the model in LMIC (page 6; references 2&3).

25. The authors should report clear objective of the study. Was the care cascade studied regionally? Were the factors affecting care continuum identified regionally? Does the study focus on males and females separately?
    As noted in our objective on page 7, we aimed to evaluate the cascade across four countries and did not study the cascade regionally. We performed exploratory analyses of the cascade by sex and study site and have now noted this as well (page 7).

26. The title of the study should clearly state the objective of the study.
    The title of the study has been rewritten in accordance with the journal's specifications.

27. Methodology including statistical analysis is confusing and difficult to follow. The section should be rewritten for a clear understanding of the study.
    We have addressed the specific comments provided below.

28. Study setting and participants: replace the term 'Twenty+'. Mention the inclusion/exclusion criteria. Was HIV the inclusion criteria? Why only HIV was considered?
    The name of the study referred to is "Birth to Twenty+", hence the term cannot be replaced.

    Given space limitations, we have referred to two manuscripts that report on the study methods in detail (page 7; references 5&6). However, we have now also included the inclusion and exclusion criteria in this manuscript (page 7). HIV status was neither an inclusion nor exclusion criterion.

29. Mention the names of categorical characteristics/variables, and continuous characteristics/variables. Why were variables compared by sex and missingness?
    The names of the categorical and continuous variables have been listed in the statistical analysis section (page 11).

    Variables were compared by sex to provide greater insight into our data as cardiometabolic disease and risk can differ by sex. Variables were compared by missingness to see if there were systematic differences between those who had missing data and those who did not. This has been clarified in the methods section (page 11).

30. 'Multivariable logistic regression……previous research' is not clear and easy to follow.
    This has been reworded for clarity (page 12).

31. ‘Stata’ is the software and not STATA
    This has been changed (page 13).

32. How were the variables selected for multivariable logistic regression?
    These were selected based on previous research studies as noted on page 12.

33. Results: The section is not easy to understand. It should be scientifically and grammatically correct. Follow a similar pattern while reporting the results. P values should be reported to 2 decimal places. The percentage should be written along with CI.
    We have changed our p values and they are now reported to two decimal places. We have carefully reviewed our results and confirmed that percentages are reported where appropriate and that the entire manuscript is grammatically correct. We would appreciate if the reviewer could draw specific examples to our attention.

34. Graphically represent the care cascade for better clarity
    We draw the reviewer’s attention to Figure 1 which graphically represents the cascade.

35. The authors should comprehensively discuss the care cascade approach used in various LMICs to study diabetes/hypertension care. How the care cascade (each component separately and overall) is different from other countries (LMICs from other continents) to guide policymakers of SSA. The authors should also discuss the probable causes behind the differences in the care cascade of diabetes
    We appreciate the reviewer’s comments. However, mindful of space constraints and given the limitations of our data, we feel the comprehensive discussion proposed by the reviewer is beyond the scope of this manuscript and it is more appropriate to focus our discussion on other studies of the diabetes care cascade within sub-Saharan Africa (see pages 17-20). We refer to the stages that should be targeted for intervention, based on our study and previous work (see page 21-22).

VERSION 2 – REVIEW

REVIEWER	Jorgensen , Jutta M Adelin University of Copenhagen Department of Public Health
REVIEW RETURNED	09-Mar-2023

GENERAL COMMENTS	Thank you for addressing my comments in the revised version.

REVIEWER	Te, Vannarath National Institute of Public Health, Health Policy and Systems Research Unit
REVIEW RETURNED	25-Mar-2023

GENERAL COMMENTS	Compared with the previous version, I have seen improvement in the current work. It is clearer; however, there are still points that need to be improved. Major comments: 1. In the title, would it be better to add the "outcome" of the associated factors and the study design "cross-sectional study"? 2. The cascade of care, to my understanding, is the main focus of this study and should be clearly illustrated with the cascade bars and definitions (perhaps in a table). In this study, there are only four bars: (1) prevalence bar, (2) awareness bar, (3) treatment, and (4) control. In other studies, there are more bars which also include testing. There could be a case that people get screened or tested in a community outreach campaign but have never been

	diagnosed by a healthcare professional. In this case, they might be aware but not yet confirmed through a proper diagnosis. I wonder why you do not consistently use the term "diagnosed" although it is self-reported. The denominator used (fixed or not) to generate the cascade should be clearly explained. I could not see Figure 1, although there is a description on the last page. 3. I find it difficult to understand this important sentence: "The proportion of those aware of having diabetes was calculated as a percentage of those with diabetes and similarly, the proportion of those ever receiving treatment for diabetes was calculated as a percentage of those aware of having diabetes; the proportion of those who had their diabetes controlled was calculated as a percentage of those who reported ever receiving treatment." 4. I am a bit concerned with the time span of this cross-sectional study (2013-2016). 5. The exclusion criteria are not quite clear to me: "Exclusion criteria were pregnancy and, given that one of the broader programme project objectives was to investigate genomic determinants of cardiometabolic disease, being closely related to an existing participant and recent immigration into the study site". Minor comments: 6. Please double check the punctuations whether they should be placed before or after the reference numbers. 7. No reference was given for this paragraph: "Studies have often been limited to diabetes prevalence and awareness and conducted in hospital-based populations, introducing selection bias, while multi-country studies that have reported on the entire cascade have meta-analysed data from heterogeneous studies with methodological differences in determining each cascade stage." 8. No reference was given for this questionnaire: "Physical activity was assessed using the Global Physical Activity Questionnaire". 9. On page 9, HbA1c is first mentioned and full words should be given. Please double check the spelling of the subsequent use of HbA1c.

VERSION 2 – AUTHOR RESPONSE

Reviewer comments

1. In the title, would it be better to add the "outcome" of the associated factors and the study design "cross-sectional study"?

   We draw the reviewer's attention to the fact that the study design is included in the title. In the interests of keeping the title as succinct as possible, we have elected not to include the outcomes of the associated factors in the title.

2. The cascade of care, to my understanding, is the main focus of this study and should be clearly illustrated with the cascade bars and definitions (perhaps in a table). In this study, there are only four bars: (1) prevalence bar, (2) awareness bar, (3) treatment, and (4) control. In other studies, there are more bars which also include testing. There could be a case that people get screened or tested in a community outreach campaign but have never been diagnosed by a healthcare professional. In this case, they might be aware but not yet confirmed through a proper diagnosis. I wonder why you do not consistently use the term "diagnosed" although it is self-reported. The denominator used (fixed or not) to generate the cascade should be clearly explained. I could not see Figure 1, although there is a description on the last page.

We draw the reviewer's attention to Figure 1 in which the cascade is clearly illustrated. We have confirmed that it has been accurately uploaded to the manuscript submission system and should therefore be available for review. The definitions for each stage of the cascade have been provided on pages 10-11 as well as on the key for Figure 1.

Our use of the term health care professional is not confined to doctors and includes those such as nurses or community health care workers who would participate in community outreach campaigns. This has been clarified on page 10.

We believe we have been clear and consistent in our definitions of "diagnosed" (page 10) and our denominators (pages 11-12) and would appreciate if the reviewer could point out any specific instances of confusion.

3. I find it difficult to understand this important sentence: "The proportion of those aware of having diabetes was calculated as a percentage of those with diabetes and similarly, the proportion of those ever receiving treatment for diabetes was calculated as a percentage of those aware of having diabetes; the proportion of those who had their diabetes controlled was calculated as a percentage of those who reported ever receiving treatment."

   This sentence explains what denominators were used at each stage of the cascade. We have divided it into two sentences to aid understanding (page 11).

4. I am a bit concerned with the time span of this cross-sectional study (2013-2016).

   We have acknowledged the limitations of the timing of the data collection on page 21.

5. The exclusion criteria are not quite clear to me: "Exclusion criteria were pregnancy and, given that one of the broader programme project objectives was to investigate genomic determinants of cardiometabolic disease, being closely related to an existing participant and recent immigration into the study site".

   This has been reworded for greater clarity (page 7).

6. Please double check the punctuations whether they should be placed before or after the reference numbers.

   The Journal's formatting instructions do not specify this, but our approach is in keeping with the format of articles which have been published in the Journal.

7. No reference was given for this paragraph: "Studies have often been limited to diabetes prevalence and awareness and conducted in hospital-based populations, introducing selection bias, while multi-country studies that have reported on the entire cascade have meta-analysed data from heterogeneous studies with methodological differences in determining each cascade stage."

   While a review of all of the existing literature is beyond the scope of this paper, we have now referenced meta-analyses which illustrate our point (page 7).

8. No reference was given for this questionnaire: "Physical activity was assessed using the Global Physical Activity Questionnaire".

   We draw the reviewer's attention to reference 10 which is cited on page 9.

9. On page 9, HbA1c is first mentioned and full words should be given. Please double check the spelling of the subsequent use of HbA1c.

   HbA_{1c} has now been fully spelt out and we have ensured its formatting is now consistent throughout the manuscript.