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 MULTI sTUdy Diabetes rEsearch (MULTITUDE) Consortium
AUTHORS	Pino, Elizabeth; Zuo, Yi; Maciel De Olivera, Camila; Mahalingaiah, Shruthi; Keiser, Olivia; Moore, Lynn; Li, Feng; Vasan, Ramachandran; Corkey, Barbara; Kalesan, B

VERSION 1 – REVIEW

REVIEWER	Jonathan R Treadwell ECRI Institute, USA
REVIEW RETURNED	12-Dec-2017

GENERAL COMMENTS	BMJ review 12-12-17 Cohort Profile: The MULTI sTUdy Diabetes rEsearch (MULTITUDE) Consortium Jonathan R Treadwell, ECRI Institute My comments are divided into major and minor to assist the BMJ in making their decision. Major comments This manuscript is essentially answering the questions “How did we develop this database?” and “What types of patients are in the database?” I suspect that for BMJ readers, particularly clinicians, these are not very pressing questions. The manuscript has two major omissions: 1) no citations or discussion of several other big-data projects related to diabetes with much the same goals, such as the project at Indiana Biosciences Research Institute (800,000 diabetics), Helen Colhoun’s project at the Univ of Edinburgh, or the Diabetes Collaborative Registry, or even the Archimedes project. I would have liked it if they authors had said something about how their database fits in with those projects, and how it offers something that the others don’t. 2) How they chose these 17 trials. I could not find any list of inclusion criteria. Is there a process for including the next large diabetes trial that comes out? Minor comments A little odd in the first sentence of the abstract for a BMJ paper that there is a focus on the US. Why not the UK? Or the world in general? To avoid the impression that the database is static, you might call it 135, 156 “ongoing” participants. Or something like that.

BMJ Open: first published as 10.1136/bmjopen-2017-020640 on 5 May 2018. Downloaded from http://bmjopen.bmj.com/ on September 27, 2021 by guest. Protected by copyright.

	Another strength is the ability to analyze rare subgroups of patients Many hospitals have databases of diabetes patients. But I suppose you are only including published trials? There should be verbiage about published data vs the" big data" that is routinely gathered and that you find the former to be more useful for your database. Page 8 line 9, you already mentioned large N, so mentioning higher statistical power is rather redundant Page 8 line 14, instead of "recently", just state the year when it was established. The FHS is rather old…today's treatment availability is very different. Might those data from the 1950s and 60s be tainting whatever inferences one would want to draw from a combined registry? Page 10 line 49 mentions "the research questions addressed" so it made me want a list of your specific questions. Page 11 line 11 mentions "target variables" but only later do you list them (I suspect Table 4). So either mention them later, or cite table 4 here. I didn't see if you said whether you consulted the published papers for additional information about how they collected their data, or whether you called the study authors etc.

REVIEWER	Freya MacMillan Western Sydney University, Australia
REVIEW RETURNED	14-Dec-2017

GENERAL COMMENTS	Thank you for the opportunity to review this manuscript, which reports on a consortium of cohort studies and trials of those with and without diabetes. Pooling of the data included in this consortium is important to advance the field in terms of understanding the determinants, risk factor and outcomes of diabetes and the authors should be congratulated on their efforts to bring together data from 17 studies thus far. The manuscript also provides a concise summary of key initial findings and provides a balanced discussion of the strengths and weaknesses of the consortium. Generally the paper is well written and I have only a few minor comments: - The introduction highlights the importance of population cohort studies but does not summarise what has already been published in the literature in this area. To highlight how the current study is filling the gaps in existing research, there needs to be a critique of existing research. - One would assume that each included study has sought ethical approval. A statement in this manuscript to confirm that all studies have ethical approval for their conduction and sharing of their data in this consortium is missing. - Do not use the term 'diabetics' - the correct, acceptable term is 'person/people with diabetes.' - page 7, line 30 - insert 'of' before 'morbidities.' - page 11, line 3 - 'self report' rather than 'self reported.' Line 36 - 'headers' rather than 'header.'

BMJ Open: first published as 10.1136/bmjopen-2017-020640 on 5 May 2018. Downloaded from http://bmjopen.bmj.com/ on September 27, 2021 by guest. Protected by copyright.

REVIEWER	Edwin Amalraj Raja University of Aberdeen Aberdeen, UK
REVIEW RETURNED	08-Jan-2018

GENERAL COMMENTS	1. When reporting mean/appropriate average please provide SD/ respective measure of variation 2. Please provide HR and its 95% CI 3. Please present p-value to the accuracy of p<0.001 (3 digits accuracy) 4. In tables, you may consider to provide % in the column heading and avoid % symbols in each number 5. You may provide number of missing for each variable in a column and calculate % for responses for each variable after excluding missing 6. You may consider to provide the title name of fig 2 as 'cumulative incidence of all-cause mortality …' instead of 'Kaplan Meier curve for all-cause mortality .. ' 7. There were no mention of statistical methods used and checked the assumption of statistical measures or statistical methods/tests 8. Are there any other cohorts/ clinical trials which could not be brought under consortium? And reason

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Jonathan R Treadwell
Institution and Country: ECRI Institute, USA
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
My comments are divided into major and minor to assist the BMJ in making their decision.

Major comments

Comment 1: This manuscript is essentially answering the questions "How did we develop this database?" and "What types of patients are in the database?" I suspect that for BMJ readers, particularly clinicians, these are not very pressing questions.
Response 1: Our manuscript is submitted under "Cohort Profiles" to BMJ Open. According to the author guidelines for Cohort Profiles, they must include:
"Describe methods of data collection and follow-up, and any external data sources used.
Give characteristics of study participants (e.g. demographic, clinical, social) and information on exposures and potential confounders. Indicate number of participants with missing data for each variable of interest."
We also believe that the extensive Data Collection Methods section will be helpful and instructive to BMJ Open readers who are clinical researchers and data scientists endeavoring to construct a similar harmonized data consortium.

Comment 2: The manuscript has two major omissions: 1) no citations or discussion of several other big-data projects related to diabetes with much the same goals, such as the project at Indiana Biosciences Research Institute (800,000 diabetics), Helen Colhoun's project at the Univ of Edinburgh, or the Diabetes Collaborative Registry, or even the Archimedes project. I would have liked it if they authors had said something about how their database fits in with those projects, and how it offers

BMJ Open: first published as 10.1136/bmjopen-2017-020640 on 5 May 2018. Downloaded from http://bmjopen.bmj.com/ on September 27, 2021 by guest. Protected by copyright.

something that the others don't. 2) How they chose these 17 trials. I could not find any list of inclusion criteria. Is there a process for including the next large diabetes trial that comes out?
Response 2: We thank the reviewer for pointing out our omission in presenting either the inclusion criteria for our consortium or summarizing the current relevant published literature.
We have added the following paragraph in the introduction, with several references to published cohort studies relevant to type 2 diabetes (i.e. Archimedes project, Diabetes Collaborative Registry):
"A comparable large-scale data harmonization effort11 has already been undertaken to better understand the long-term risks for cardiovascular disease (CVD) and to examine patterns of CVD development over the adult life course. However, similar projects focused on T2DM have thus far been limited in their sample size12 13, participant demographic makeup12-14, years of follow-up12, or with their focus on improvements in care15 16 rather than on determining risk of diagnosis and adverse outcomes. Additionally, while multiple risk models17-19 aimed at early identification of patients at high risk of developing T2DM are already widely used in the clinical settings, these models typically only consider the patient's current state at the time of the assessment, ignoring the complex trajectory of events that leads up to the disease state. The MULTITUDE consortium aims to address these limitations."
We have added the following to the Cohort Description, Study inclusion and follow-up time, paragraph 1:
"The 17 cohort studies and clinical trials in the MULTITUDE consortium were included based on their availability as open source data from the NHLBI Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) and their relevance to T2DM risk and outcomes."
We have added the following to the Conclusions, paragraph 1. In this sentence we cite both the project at Indiana Biosciences Research Institute and Helen Colhoun's project at the University of Edinburgh. As neither of those projects have been published to date, we did not include them in the introduction:
"Using the same harmonization principles, this data resource can be extended to include a larger number of studies to provide a more comprehensive data infrastructure as relevant data is added to the BioLINCC repository. Several promising large-scale retrospective data analyses focused on gaining a better understanding of T2DM risk and outcomes are currently underway.27 28"

Minor comments
Comment 3: A little odd in the first sentence of the abstract for a BMJ paper that there is a focus on the US. Why not the UK? Or the world in general?
Response 3: We have changed this section of the abstract and the introduction to cite global statistics on T2DM prevalence:
"Globally, the age-standardized prevalence of T2DM has nearly doubled from 1980 to 2014, rising from 4·7% to 8·5% with an estimated 422 million adults living with the chronic disease."

However, in the introduction, we have included the statistics citing US prevalence and cost of treatment as we believe it is relevant to the US-based cohorts and clinical trials comprising the MULTITUDE consortium. We used data from the NHLBI BioLINCC repository which exclusively consists of US studies. We hope to provide this harmonized data consortium back to the BioLINCC repository for other researchers to use for their studies.

Comment 4: To avoid the impression that the database is static, you might call it 135, 156 "ongoing" participants. Or something like that.
Response 4: This sentence has been changed to:
"Among the 135,156 ongoing participants included in the consortium . . ."

Comment 5: Another strength is the ability to analyze rare subgroups of patients
Response 5: We have added to the Strengths and Limitations:

BMJ Open: first published as 10.1136/bmjopen-2017-020640 on 5 May 2018. Downloaded from http://bmjopen.bmj.com/ on September 27, 2021 by guest. Protected by copyright.

"Using the consortium data, we will be able to understand the variation in risk among different subgroups, including rare populations with T2DM . . . "

Comment 6: Many hospitals have databases of diabetes patients. But I suppose you are only including published trials? There should be verbiage about published data vs the" big data" that is routinely gathered and that you find the former to be more useful for your database.
Response 6: We have added the following to the Cohort Description, Study inclusion and follow-up time, paragraph 1:
"The 17 cohort studies and clinical trials in the MULTITUDE consortium were included based on their availability as open source data from the NHLBI Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) and their relevance to T2DM risk and outcomes."
We also previously state in the introduction that we believe, "An optimal approach to examining T2DM risk and disease progression involves the longitudinal examination of population-based cohorts."

Comment 7: Page 8 line 9, you already mentioned large N, so mentioning higher statistical power is rather redundant
Response 7: The mention of "increased statistical power for detecting effects" has been removed.

Comment 8: Page 8 line 14, instead of "recently", just state the year when it was established.
Response 8: The sentence was changed to:
". . . (MULTITUDE) Consortium was established in 2017 . . ."

Comment 9: The FHS is rather old…today's treatment availability is very different. Might those data from the 1950s and 60s be tainting whatever inferences one would want to draw from a combined registry?
Response 9: As we state in the introduction, one of the research goals of the consortium is:
". . . to determine the temporal patterns of T2DM and related morbidity and mortality in the US." We believe the FHS original cohort, in conjunction with the other FHS cohorts will help us understand the temporal patterns of T2DM in relation to new treatments and therapeutics.
We also state, under Data Harmonization:
"We have also adjusted each analysis for cohort, which may help attenuate any confounding due to measurement differences or varying calendar decades across cohorts."
Finally, in the second paragraph of Conclusions, we compare the hazard ratios for all-cause mortality between those with and without T2DM in the Jackson Heart Study and FHS. The FHS original cohort can be used to make complex comparisons to other studies:
"The risk ratio of all-cause mortality among JHS participants with T2DM more closely resembles the original FHS cohort, recruited in 1948 when CVD risk factors were largely unknown and medical interventions more limited, than the risk ratio from their contemporaries in FHS-gen3. This suggests that either individuals with T2DM are more protected from mortality in the JHS cohort or, perhaps, that all individuals in this cohort are more at risk for mortality compared to FHS-gen3. It is likely that there is a complex interplay between genetics, lifestyle, culture, and access to healthcare that remains to be explored further."

Comment 10: Page 10 line 49 mentions "the research questions addressed" so it made me want a list of your specific questions.
Response 10: In the introduction, paragraph 4, we state the main research objectives of the consortium:
"The main research objectives of this project are to determine the relationship between the lifetime risk of T2DM and associated major risk factors, the transition-specific risk of adverse outcomes from T2DM diagnosis through intermediate morbidity to eventual mortality, and to determine the temporal patterns of T2DM and related morbidity and mortality in the US."
In the first paragraph of the conclusion we state:

BMJ Open: first published as 10.1136/bmjopen-2017-020640 on 5 May 2018. Downloaded from http://bmjopen.bmj.com/ on September 27, 2021 by guest. Protected by copyright.

"MULTITUDE consortium will be a unique resource for conducting research to determine: 1) age, time period, and cohort differences in the incidence and progression of T2DM 2) the sequence of events or biomarkers prior to T2DM diagnosis 3) disease progression from T2DM to cardiovascular disease outcomes, T2DM complications and premature mortality, and 4) to assess race/ethnicity differences in the above associations."

Comment 11: Page 11 line 11 mentions "target variables" but only later do you list them (I suspect Table 4). So either mention them later, or cite table 4 here.
Response 11: We have changed this sentence to read:
"To enable harmonization, we ensured that all the study-specific data items required to generate the target variables (Tables 2-4, Supplementary Tables 3-5) were available and that the collected information was valid."

Comment 12: I didn't see if you said whether you consulted the published papers for additional information about how they collected their data, or whether you called the study authors etc.
Response 12: Under Data Collection Methods, we state:
"As well, we documented information such as study designs, sampling protocols, and data access policies in order to evaluate sources of study heterogeneity and feasibility of harmonization."
However, we did not directly contact study authors, since we were limited by Biolincc repository methods.

Reviewer: 2
Reviewer Name: Freya MacMillan
Institution and Country: Western Sydney University, Australia
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
Thank you for the opportunity to review this manuscript, which reports on a consortium of cohort studies and trials of those with and without diabetes. Pooling of the data included in this consortium is important to advance the field in terms of understanding the determinants, risk factor and outcomes of diabetes and the authors should be congratulated on their efforts to bring together data from 17 studies thus far. The manuscript also provides a concise summary of key initial findings and provides a balanced discussion of the strengths and weaknesses of the consortium. Generally the paper is well written and I have only a few minor comments:

Comment 13: The introduction highlights the importance of population cohort studies but does not summarise what has already been published in the literature in this area. To highlight how the current study is filling the gaps in existing research, there needs to be a critique of existing research.
Response 13: We thank the reviewer for pointing out our omission in summarizing the current relevant published literature.
We have added the following paragraph in the introduction, with several references to published cohort studies relevant to type 2 diabetes (i.e. Archimedes project, Diabetes Collaborative Registry):
"A comparable large-scale data harmonization effort11 has already been undertaken to better understand the long-term risks for cardiovascular disease (CVD) and to examine patterns of CVD development over the adult life course. However, similar projects focused on T2DM have thus far been limited in their sample size12 13, participant demographic makeup12-14, years of follow-up12,

BMJ Open: first published as 10.1136/bmjopen-2017-020640 on 5 May 2018. Downloaded from http://bmjopen.bmj.com/ on September 27, 2021 by guest. Protected by copyright.

or with their focus on improvements in care15 16 rather than on determining risk of diagnosis and adverse outcomes. Additionally, while multiple risk models17-19 aimed at early identification of patients at high risk of developing T2DM are already widely used in the clinical settings, these models typically only consider the patient's current state at the time of the assessment, ignoring the complex trajectory of events that leads up to the disease state. The MULTITUDE consortium aims to address these limitations."

Comment 14: One would assume that each included study has sought ethical approval. A statement in this manuscript to confirm that all studies have ethical approval for their conduction and sharing of their data in this consortium is missing.
Response 14: Under Cohort Description, first paragraph, we have added:
"All studies have been approved for the sharing of their data in this consortium per the NHLBI policy for data sharing from clinical trials and epidemiology studies (http://www.nhlbi.nih.gov/funding/datasharing.htm)."

Comment 15: Do not use the term 'diabetics' - the correct, acceptable term is 'person/people with diabetes.'
Response 15: All reference to "diabetic(s)" as a noun have been changed to the more acceptable term proposed by the reviewer.

Comment 16:
1. Page 7, line 30 - insert 'of' before 'morbidities.'
2. Page 11, line 3 - 'self report' rather than 'self reported.'
3. Line 36 - 'headers' rather than 'header.'

Response 16: All typos have been fixed.

Reviewer: 3
Reviewer Name: Edwin Amalraj Raja
Institution and Country: University of Aberdeen, Aberdeen, UK
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
Comment 17: When reporting mean/appropriate average please provide SD/ respective measure of variation
Response 17: All means reported in Table 5 are provided along with standard deviations.

Comment 18: Please provide HR and its 95% CI
Response 18: In Figure 2, hazard ratios and 95% confidence intervals are shown for all cohorts that included persons both with and without T2DM. HRs and 95% CIs are also reported in the text under Findings to Date:
"(Hazard ratios HR [95% CI]: ARIC=2.36 [2.21-2.53], FHS-cohort=1.28 [1.06-1.56], FHS-offspring=2.65 [2.06-3.41], FHS-gen3=2.83 [1.20-6.70], JHS=1.78 [1.52-2.10]) compared to lower risk among clinical trials (HR [95% CI]: AFFIRM=1.84 [1.55-2.18], ALLHAT=1.36 [1.29-1.43], CORAL=1.28 [0.90-1.81], MRFIT=1.12 [0.79-1.58], SPRINT-POP=1.47 [0.37-5.92])."

Comment 19: Please present p-value to the accuracy of $p<0.001$ (3 digits accuracy)
Response 19: All p-values are reported in Table 5 as well as in the text under Findings to Date.

BMJ Open: first published as 10.1136/bmjopen-2017-020640 on 5 May 2018. Downloaded from http://bmjopen.bmj.com/ on September 27, 2021 by guest. Protected by copyright.

Comment 20: In tables, you may consider to provide % in the column heading and avoid % symbols in each number
Response 20: This has been changed accordingly in Table 5.

Comment 21: You may provide number of missing for each variable in a column and calculate % for responses for each variable after excluding missing
Response 21: This has been changed accordingly in Table 5.

Comment 22: You may consider to provide the title name of fig 2 as 'cumulative incidence of all-cause mortality …' instead of 'Kaplan Meier curve for all-cause mortality .. '
Response 22: The title of Figure 2 has been changed to the reviewer-suggested title.

Comment 23: There were no mention of statistical methods used and checked the assumption of statistical measures or statistical methods/tests
Response 23: Statistical methods for Table 5 and Figure 2 are listed in the legends of each respective figure:
"Categorical variables were compared using chi-square tests and continuous variables were compared using the Student's t-test."
"Kaplan Meier curves represent all-cause mortality. Cox proportional hazards regression models were used to estimate hazard ratios (HR) and 95% confidence intervals (95% CI), adjusted for age."

Comment 24: Are there any other cohorts/ clinical trials which could not be brought under consortium? And reason
Response 24: We encountered no cohorts/ trials thus far that we have not been able to add to the consortium. However, select cohorts/ trials have limited utility within the consortium due to shorter follow-up time (FOCUS, OMNI Heart) or limited CVD and T2DM-related timed outcomes (FOCUS, NGHS, OMNI Heart). However, with the addition of new studies to the Biolincc, we plan on implementing a second phase where we may choose to add other studies to the MULTITUDE consortium.

VERSION 2 – REVIEW

REVIEWER	Edwin Amalraj Raja University of Aberdeen, UK
REVIEW RETURNED	06-Feb-2018

GENERAL COMMENTS	Nothing. Some of the questions need to be answered by subject specialist

BMJ Open: first published as 10.1136/bmjopen-2017-020640 on 5 May 2018. Downloaded from http://bmjopen.bmj.com/ on September 27, 2021 by guest. Protected by copyright.