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 CARTaGENE Cohort Nutrition Study (Quebec, Canada)
AUTHORS	Ho, Vikki; Csizmadi, Ilona; Boucher, Beatrice; McInerney, Maria; Boileau, Cahterine; Noisel, Nolwenn; Payette, Yves; Awadalla, Philip; Koushik, Anita

VERSION 1 – REVIEW

REVIEWER	Jeremy C.-H. Wang University of Waterloo, Systems Design Engineering
REVIEW RETURNED	02-Jun-2024

GENERAL COMMENTS	The paper presents an observational cohort study that collected dietary characteristics and potential predictors on a cohort (n = 9,379 after inclusion criteria applied) derived from the original CARTaGENE study of circa 2009/2010, which predominantly sampled 40 to 69 year-olds in Quebec at the original time of recruitment. Overall, the paper is well-written and relevant, and appears to meet the majority of STROBE criteria that one might apply to an epidemiological study. Furthermore, the CARTaGENE website offers a complete description of variables. I offer the following considerations for revision: 1. The selected characteristics described in Tables 1 and 4 are helpful for dataset validation (i.e. comparing to CCHS or established health-diet correlations) but a more complete summary of variables is necessary for readers intending to evaluate the dataset for use in future research. Short of this, readers may underappreciate or prematurely dismiss the completeness of the data, especially when several aspects such as ethnicity or alcohol consumption were also collected but are not summarized or discussed. If such a summary becomes too long to be included in the main body, then perhaps a supplement would be more fitting. 2. Further to the above, the paper has limited discussion on the potential limitations of the study. For example, what variables would the researchers have liked to collect, but were unable to? How does the present cohort compare to other nutritional cohorts, either in Canada or other regions?

VERSION 1 – AUTHOR RESPONSE

Responses to Reviewer #1

Comment #1: *The selected characteristics described in Tables 1 and 4 are helpful for dataset validation (i.e. comparing to CCHS or established health-diet correlations) but a more complete summary of variables is necessary for readers intending to evaluate the dataset for use in future research. Short of this, readers may underappreciate or prematurely dismiss the completeness of the data, especially when several aspects such as ethnicity or alcohol consumption were also collected but are not summarized or discussed. If such a summary becomes too long to be included in the main body, then perhaps a supplement would be more fitting.*

Reply: We appreciate the Reviewer recognizing the richness of the CARTaGENE cohort. There is indeed a multitude of factors, described in the cited previous publication of the full CARTaGENE cohort, which can be linked to the diet data in the CARTaGENE Cohort Nutrition Study. As we described in the introduction to our article, the motivation to collect dietary data in CARTaGENE was our interest in chronic diseases. Thus, in describing the CARTaGENE Cohort Nutrition Study here, we have highlighted general variables commonly known to have an important relation to chronic diseases, but also other outcomes. Furthermore, with the descriptive analysis presented in Table 4, we showed that diet quality varies across subgroups based on these variables as would be expected given past research in the context of chronic diseases. These results demonstrate the value of the CARTaGENE Cohort Nutrition Study for future studies of diet and health. Table 4 results reflect a simple multivariable analysis for descriptive purposes, which applies generally to chronic diseases. Future evaluations of the diet data according to other factors/diseases will of course take into consideration the specific topic and interrelated variables by the researcher carrying out the evaluation. Nonetheless, we agree that the richness of the CARTaGENE data can be further highlighted by indicating the other data collected. Thus, we have now included the following text in the “Findings to date” section, first paragraph, just after describing Table 1:

> As described in the CARTaGENE cohort profile paper,[6] information on various domains, such as sociodemographic factors, mental health, psychosocial environment, disease history, health care/ medication use, and reproductive health, was collected and can be linked to the dietary data presented here. Readers are also directed to the CARTaGENE website for an exhaustive list of variables (www.cartagene.qc.ca).

Comment #2: *Further to the above, the paper has limited discussion on the potential limitations of the study. For example, what variables would the researchers have liked to collect, but were unable to? How does the present cohort compare to other nutritional cohorts, either in Canada or other regions?.*

Reply: We appreciate this comment from the Reviewer, which allowed us to reflect further on potential limitations. In doing this, we realized the importance of including other strengths as well. We have added these in various parts of the “Strengths and limitations” section.

First, we have now noted a limitation of our FFQ (i.e., the C-DHQ II), which like all FFQs, has a finite food list that limits its ability to capture all foods consumed by participants. However, it was designed using national consumption data available at the time to identify the most commonly consumed items. We have included the following in the second paragraph of the "Strengths and limitations" section:

> Like all FFQs, the C-DHQ II has a finite food list that limits the ability to capture all foods consumed by participants. However, by using the most recent national intake data at the time[10] along with consultation with researchers and Health Canada collaborators with access to market data, the most commonly consumed foods, including items newly available up to 2010, were included on the C-DHQ II.[8]

Although it is also possible that new foods and dietary patterns have emerged since 2010, comprehensive FFQs such as ours are intended to capture the intake of broad groups of foods, rather than specific brands and products. As such, the C-DHQ II design optimizes the inclusion of many new foods within existing questions which limits the need for frequent updates. Nonetheless, annual 24-hour recall data will be collected going forward, which will update and enrich the diet data described in this cohort profile. We have now provided more detail on the 24-hour recall in the "Collaboration section":

> Of note, starting in 2024, annual diet assessments using two 24-hr dietary recalls over a 30-day period will be administered in collaboration with NutriQuébec (Université Laval), further expanding on the strengths of the cohort as a resource for dietary studies.

We also note the completeness of the data in the CARTaGENE cohort, as indicated in the Reviewer's comment #1. Specifically, in the first paragraph of the "Strengths and limitations" section, we have added:

> The dietary data along with other questionnaire data, physical measures and biospecimens makes the CARTaGENE Cohort Nutrition Study one of the richest resources available for dietary studies in Canada. Indeed, these dietary data have already been used in reports on subsets of the CARTaGENE population.[17-20] The present profile provides the first descriptive analysis of the dietary profile for all CARTaGENE participants with complete dietary data.

We also further note in the "Collaboration" section that only one other cohort participating in the larger Canadian Partnership for Tomorrow's Health Study has collected diet data as comprehensive as that collected in the CARTaGENE Cohort Nutrition Study:

> Only one other regional cohort of the Canadian Partnership for Tomorrow's Health study, the Alberta Tomorrow Project[23-25], has also collected comprehensive diet data.

These baseline data are key to maintaining the correct temporal order between dietary exposures of interest and the majority of subsequent outcomes that have already occurred in this cohort and will be studied in etiologic studies.