1.a. Full Title: Adulthood BMI latent class trajectory modelling (LCTM) and obesity-related cancers in the ARIC cohort study

[Part of the Adulthood BMI And Cancer using clusters (ABACus) project. PROSPERO CRD42017079621]

b. Abbreviated Title (Length 26 characters):

2. Writing Group:
   Writing group members:
   Charlotte Watson, Matt Sperrin, Hannah Lennon, Corinne Joshu, Elizabeth Platz, Andrew Renehan and other interested ARIC investigators.

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. ___CW___ [please confirm with your initials electronically or in writing]

First author: Charlotte Watson
   Address: Vaughan House, Portsmouth St, University of Manchester, M13 9GB

   Phone: +44161 3067925   Fax:
   E-mail: charlotte.watson-3@postgrad.manchester.ac.uk

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).

   Name: Professor Elizabeth Platz
   Address: Department of Epidemiology
   Johns Hopkins Bloomberg School of Public Health
   615 N. Wolfe St., Room E6132
   Baltimore, MD 21205

   Phone: 410.614.9674   Fax: 410.614.2632
   E-mail: eplatz1@jhu.ac.uk

3. Timeline:

4. Rationale:
Excess adiposity, commonly approximated as body mass index (BMI) over 25 kg/m², is an established risk factor for several incident cancers, and in many western populations, is the second commonest cause after smoking.

The 2016 IARC report (1) on weight and cancer concluded that elevated BMI is associated with 13 cancer types – endometrial, oesophageal adenocarcinoma, gastric cardia, kidney, liver, multiple myeloma, meningioma, pancreas, colorectal, gallbladder, postmenopausal breast, ovary, thyroid - referred to as obesity-related cancers.

The IARC report noted that the BMI-cancer association may be modified by other risk factors, such as smoking.

The above epidemiology is based on a once-only BMI measurement, typically at cohort entry. However, at a biological level, the life-course exposure of adiposity is likely to be mechanistically more relevant (2). Relatively simple measures, such as weight change across adulthood, are associated with increased cancer risk (3), in patterns mirroring those for baseline ‘once-only’ BMI.

A more recent, and statistically more advanced approach to estimate life-course exposure is considering clusters of weight changes, known as latent class trajectory modelling (LCTM). These methods have been used in the social medicine and psychology literature (4) but are now gaining ground in mainstream epidemiology. In the context of repeated BMI, there is the added value of identifying adverse body weight trajectories with potential to serve as an early alert system to intervene with lifestyle changes. To-date, for cancer incidence (5, 6) and mortality (7), there have been reports with BMI LCTM, but numbers of classes and findings have been inconsistent.

The statistical application of LCTM is not trivial. The present collaboration has written an eight-step framework on how to optimally select models taking into account the selection of number of classes; the non-linearity of trajectories; and the variance within classes (fixed- versus random-effect models). This manuscript is under revision with *BMJ Open* (8).

Other methods include linear mixed models which can be used to link trajectories to other variables. For example, Windham et al (2017) demonstrated that maintaining a normal BMI in mid-late life may help preserve later life mobility (9).

**LCTM of repeated BMI measures in ARIC**

We acknowledge that latent classes have been previously developed in the ARIC cohort to explore the changes in blood pressure (10).

Derived BMI (from directly measured weight and height) is available in the ARIC cohort at baseline and repeated weights through follow-up: measured 3 years later (1990-1992), 6 years later (1993-1995) and 9 years later (1996-1999). This data collection pattern is mirrored for height data excluding follow-up 1 (11).

We found no report or registration of intent to develop LCTM from BMI data in ARIC and related these to cancer risk.

5. **Main Hypothesis/Study Questions:**

   **Overall aim:**
Using algorithms detailed in our *BMJ Open* manuscript, we seek to extend our methodological framework across several cohorts with repeated BMI measures, in the ABACus consortium (PROSPERO CRD42017079621), and specifically test whether the variations in results noted to-date are due to true population differences or differences in model selection.

**Specific objectives:**
1. Characterize baseline and repeated BMI measures through follow-up, assessing for general patterns of BMI changes and missingness of data;
2. Build LCTM of repeated BMI measurements from baseline entry through follow-up in ARIC participants, using an increasing hierarchical (fixed to fully random-effects) approach; test for various optimal models; and optimise class separation;
3. Determine associations between latent classes and obesity-related cancers (rather than individual cancer types), and latent classes and non-obesity-related cancers (as negative controls);
4. There will be unique opportunities in ARIC due to the richness of data to explore for different patterns of LCTM between races in both sexes.

The strength of latent class trajectory modelling is that BMI trajectories are not just an exposure, but a novel indicator i.e. identification of adverse trajectories early offers opportunities for early intervention for weight control e.g. rapid weight gain in first five years after menopause.

6. **Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).**

This proposal sits within a larger project, Adulthood BMI And Cancer using clusters (ABACus) project. PROSPERO CRD42017079621, which is collecting data, centralizing in Manchester, United Kingdom, from other global cohorts with repeated BMI and running similar LCTMs.

Importantly, this is not a pooled analysis across cohorts – as it is not known at this time how one would ‘pool’ aggregates from LCTM. Instead, ABACus aims to run equivalent models across different populations and evaluate for differences attributable to population differences rather than model structure differences.

**Data analysis strategy:**
We envisage minimum need for data clean-up and/or harmonisation.

We will initially characterise baseline and repeated BMI measures through follow-up, assessing for general patterns of BMI changes and missingness of data, in the ARIC.

We will assess for missingness completely at random versus missing at random using Little’s test (though we note that there is currently no intention to impute missing data, as LCTMs handle missing data directly, under a missing at random assumption). The length of follow-up will be informative as this describes the time to event outcome. We will explore both a joint modelling framework and partly conditional (landmarking) approach to account for this (12, 13).

We will then develop LCTMs and related to cancer incidence in a two stage process.
Stage 1

We will use latent class linear mixed models, which allows for individual variation within groups, and allows for uncertainty in assignment of individuals to latent classes. The method simplifies a heterogeneous population into relatively homogeneous latent classes whilst incorporating random effects for each individual and accounting for individual level data. Models with differing number of classes will be fitted to the data and the optimal model will be chosen by considering both the Bayesian Information Criterion (BIC) as well as the clinical relevance of the composition of the latent classes. These classes will encompass the obesity exposure of each participant during the course of ARIC data collection and make up our longitudinal trajectories.

We have reported an eight-step approach to systematically build a ‘preferred’ or ‘core’ LCTM (now accepted by BMJ Open). To summarise; a scoping model which provisionally selects the plausible number of classes from the literature - typically four to six. This is then refined to determine the optimal number of classes, tested separately on men and women. The preferred number of classes is used to determine optimal model structure using a variety of models and multiple statistic packages. A number of model adequacy assessments are performed and graphical presentation approaches used to confirm model choice. This is then assessed for clinical plausibility and characterisation to show that these latent classes provide information above and beyond standard characterisation.

Finally sensitivity analyses are conducted using various numbers of BMI values to demonstrate that our model can cope with more or less data and still provide the same results.

Stage 2

From the classes derived in stage 1, we will undertake standard association modelling and for the following outcomes: (i) overall cancer incidence; (ii) obesity related cancer incidence; (iii) non-obesity-related cancer incidence (these serve as a ‘negative’ control i.e. we expect no association here).

In the first instance, multivariable Cox proportional hazard regression will be used to relate class membership with cancer incidence, with age as the time scale, left truncating at recruitment date.

We will adjust for common cancer risk factors, typical list below:

Typical required variables

1. Study ID
2. Age at study entry
3. Smoking status (current, former, never)
4. Postmenopausal Hormone therapy status (current, former, ever)
5. Race/Ethnicity
6. Alcohol units consumed per week (g/d)
7. Highest obtained educational level
8. Aspirin use
9. Recreational Physical activity
10. Cholesterol levels (from baseline questionnaire)
11. Cancer family history (main sites; site-specific)
12. Screened detected cancer (obesity related; any cancer)
We will also extract other variables relevant to characterizing the derived latent classes e.g. waist circumference and body fatness percentage.

7.a. Will the data be used for non-CVD analysis in this manuscript? __✓__ Yes    ____ No

    b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = “CVD Research” for non-DNA analysis, and for DNA analysis RES_DNA = “CVD Research” would be used? _____ Yes    _____ No

    (This file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? _____ Yes    __✓__ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = “No use/storage DNA”? _____ Yes    _____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status. ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.cscc.unc.edu/ARIC/search.php

    __✓__ Yes    ________ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

    #1792 (The influence of obesity, diabetes, and associated metabolic perturbations on cancer risk, Joshu CE)

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? __✓__ Yes    ____ No

11.b. If yes, is the proposal

    __✓__ A. primarily the result of an ancillary study (list number* 2011.07 & 1995.04)
    ____ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* __________ __________ __________)

*ancillary studies are listed by number at http://www.cscc.unc.edu/aric/forms/
12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is your responsibility to upload manuscripts to PubMed Central whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscce.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

References: