1.a. Full Title: Aging and psychosocial causes of cancer disparities in understudied subpopulations – residents of lower population/density areas, those of low SES, and the elderly, including those who are African-American

b. Abbreviated Title (Length 26 characters): Cancer disparities causes

2. Writing Group:
Writing group members: Joshu, Barber, Browner, Deal, Gross, Heiss, Jones, McDoom, Mosley, Windham, Platz

Others are always welcome to join.

We have also invited Dr. Ana Diez Roux; we are awaiting her response.

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

First author: Elizabeth Platz (submitter)
Address: Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD

Phone: (410) 614-9674 Fax: (410) 614-2632
E-mail: eplatz1@jhu.edu

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: Corinne Joshu
Address: Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD

Phone: 443-287-3821 Fax: 410-614-2632
E-mail: cjoshu1@jhu.edu

3. Timeline: Manuscript will be submitted by September 2018

4. Rationale: Given that nonmetropolitan/rural residents, those of low SES, and the elderly including elderly African-Americans are disproportionately burdened by cancer and its consequences, it is remarkable that they are so understudied in epidemiology (1). Thus, we
propose to investigate – in these subpopulations low population/density residents, those of low SES, and the elderly – aging and psychosocial factors in relation to and as consequences of cancer. These factors have been studied for mortality and chronic diseases (2) but have not been systematically investigated for their cancer-related influence in these subpopulations. These subpopulations may be less able to compensate for negative influences of aging and psychosocial factors because they are less “resourced” at the individual, social, and physical environment levels. ARIC is a cohort in which we have the ability – with existing data and extended expertise in aging and social determinants – to contribute data to help build the evidence base for subpopulation-targeted cancer-related interventions and policies as called for by Martin et al. (1), for example to better target resources to persons and geographic areas for cancer screening and supportive care.

Main Hypothesis/Study Questions:

Our overall hypothesis is that members of understudied subpopulations, specifically residents of less populated/dense places, those of low socioeconomic status (SES), and the elderly including elderly African-Americans are disproportionately burdened by cancer and its consequences, in part, because aging and adverse psychosocial factors are more influential in these subpopulations. These subpopulations may be less able to compensate for the negative influences of aging and psychosocial factors because they are less “resourced” at the individual, social, and physical environment levels.

Aim 1: Disparities in the links of aging and psychosocial factors with cancer incidence and mortality

Among participants in ARIC who attended Visit 2 and never had a cancer diagnosis by that time (N=13,143), to evaluate whether:

a. Lower cognitive function, lower social support, and greater vital exhaustion (i.e., feeling burned out) are associated with a greater risk of cancer and cancer death overall and by major cancer site.
   
   b. Residence in less populated/dense places, low lifecourse SES, and elderly age accentuate the associations of lower cognitive function, lower social support, and greater vital exhaustion with greater risk of cancer and cancer death overall and by major cancer site.

Aim 2: Disparities in the links of aging and psychosocial factors with diagnosis stage and case-fatality

Among participants in ARIC who attended Visit 2 and never had a cancer diagnosis by that time but subsequently were diagnosed with cancer (N=3,561), to evaluate whether:

a. Lower cognitive function, lower social support, and greater vital exhaustion are associated with a later stage at cancer diagnosis and case-fatality overall and by major cancer site.

b. Residence in less populated/dense places, low lifecourse SES, and elderly age accentuate the associations of lower cognitive function, lower social support, and greater vital exhaustion with later stage at cancer diagnosis and case-fatality overall and by major cancer site.

Aim 3: Disparities in the aging and psychosocial consequences of cancer
Among participants in ARIC diagnosed with cancer after Visit 2 and who survived to and attended Visit 5 (N=1,150; 66% are breast, prostate, colorectal, or endometrial cancers), to evaluate whether:

a. Residents of less populated/dense places, those of low lifecourse SES, and the elderly are more likely to experience accelerated aging – measured by low cognitive function and low physical function, and to score higher on a depression scale (Visit 5) after diagnosis than other cancer survivors.

b. Higher cognitive function (Visits 2) and higher social support (Visit 2) prior to diagnosis attenuate the consequences of residence in less populated/dense places, low lifecourse SES, and elderly age.

c. Co-morbidities (CVD, stroke, hypertension, diabetes, COPD) prior to diagnosis accentuate consequences of residence in less populated/dense places, low lifecourse SES, and elderly age.

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).

Understudied subpopulations: We will classify each ARIC participant with respect to:

• Residence in less populated or less densely populated places per US Census in 1990 (3), the one closest in time to the ARIC recruitment period, and for Aims 2 and 3, also residence at diagnosis based on the closest Census, including for those who moved out of ARIC sites. Home addresses are already geocoded (4) and linked to Census tract and block group data for the 1990 and 2000 Censuses (5). Dr. Jones has linked the 2010 Census (as part of a separate ARIC manuscript proposal). Using the geocoded data, we will classify participants with respect to population size or density based on the distribution at baseline (e.g., below 25th percentile, at or above the 25th percentile).

• Low SES in early childhood, early and mature adulthood, and lifecourse. SES was ascertained at Visit 4 by interview inquiring about several educational and economic attributes, which were scored and the scores were summed. The scores can be divided into low and high SES for each life period and cumulatively (or used continuously or divided into quantiles; details in (6, 7). 50% participants have low cumulative SES (6).

• Elderly will be defined as age 65+ (8). At Visit 2 (1990-92) and without prevalent cancer, 12.3% were 65+ yrs, and of African-American participants, 11.2% were 65+. At Visit 4 1996-98, 37.8% were 65+ yrs, and of African-Americans 31.4% were 65+. At end of follow-up in 2012, all participants are 68+ years old. Mean age at cancer diagnosis is 68 yrs and cancer death is 71 yrs; in African-Americans these means are 67 and 69 yrs. In subanalyses, we will explore Aims 1-3 for cancer overall using age cutpoints older than 65 yrs.

Aim 1: Disparities in the links of aging and psychosocial factors with cancer incidence and mortality

Study design and analytic population. We will use the prospective cohort study design and include 13,143 participants who attended Visit 2 (first assessment of the aging and psychosocial factors) and did not have a cancer diagnosis by then. Follow-up will end at the date of an outcome, end of 2012 (most recent diagnoses from cancer registry linkage), death from another cause, or withdrawal from ARIC, whichever comes first.
Exposures. a) cognitive function by Delayed Word Recall (memory), Digit Symbol Substitution Test (executive attention), and Word Fluency (language) collected at Visit 2; b) social support by an Interpersonal Support Evaluation list and the Lubben Social Network scale collected at Visit 2; and c) vital exhaustion by the Maastricht Vital Exhaustion Questionnaire collected at Visit 2. We will classify individuals with respect to cognitive function (9), social support (10), and vital exhaustion (10) as before in ARIC.

Outcomes. In the ARIC Cancer U01, we have ascertained cancers from 4 state cancer registries supplemented with case finding by cohort active surveillance. Participants self-report all hospitalizations during annual and semi-annual follow-up phone calls and we review discharge summaries from hospitals in ARIC catchment areas; we obtain medical records for these and review to verify cases not ascertained through cancer registry linkage. Deaths from cancer as the underlying cause are obtained from death certificates. We will evaluate first, incident primary cancers, cancer deaths, total and major cancer sites – lung, prostate, (female) breast, colorectal, and other sites such as endometrial and bladder cancer as sample size allows – diagnosed after Visit 2.

Covariates. Covariates will be derived from Visit 2, including demographics (age, sex, race, study site, education) and family history of cancer, and at Visit 2 and subsequent visits before cancer diagnosis (e.g., Visits 3, 4) including anthropometrics (measured weight, height, waist), reproductive factors (menopause status, age at menopause, HRT, age at menarche, parity, oral contraceptives), lifestyle (smoking, pack-years, inactivity, alcohol), access and use of care. Ever diagnosis of comorbidities – hypertension, high cholesterol, CVD and stroke, diabetes, COPD – will be cumulative across visits before cancer diagnosis.

Statistical analysis. We will use Cox proportional hazards regression to estimate the hazard ratio (RR) and 95% confidence interval (CI) of total cancer incidence and mortality associated with a) lower cognitive function (quartiles (9)), b) lower social support (low ≤25 [8%], moderate 26-30 [14%], high ≥31 points [78%] (10)), or c) greater vital exhaustion (<14 [70%], ≥14 points [30%] (10)). Cutpoints are as previously used in ARIC studies; other parameterizations will be considered. We will adjust for sex, race*study site, education (<high school, high school/equivalent, some college or higher), and family history of cancer (model 1) as well as for modifiable risk factors including smoking (never, former current) and pack-years smoked (updated, continuous), obesity (BMI [updated, continuous], waist [updated, continuous]), physical inactivity (quintiles), alcohol (0, >0-<1, 1, >1 drink/day), and hormone replacement therapy (female-HRT and female-noHRT both versus male; updated) (model 2) in the total cancer incidence and mortality analyses. In model 3, we will additionally adjust for measures of access (insurance: yes/no) and use of healthcare (check up at least once a year, at least once every 5 years, less than once every 5 years or never). We will repeat these models mutually adjusting the aging/psychosocial factors (model 4) given expected correlations. By major cancer site, we will adjust for site-specific risk factors (e.g., reproductive for breast cancer including HRT [vs no use]). To determine whether population size or density of residence (e.g., lower [25%], higher [75%]), low SES (binary for childhood, early adulthood, mature adulthood, and cumulative; yes [50%], no [50%]), and elderly age (yes [65+ 40% of PY], no [60% of PY]; will consider older age cutpoints and restrict to elderly African-Americans) accentuate these associations, we will stratify models 1- 4 by the subpopulations. To test for subpopulation (i.e.,
stratum-specific) RR differences, we will enter into the model main effects terms for the aging/psychosocial factor and subpopulation along with a term for their cross product and test that model against the model without the cross-product term by the likelihood ratio test. Given our goal, we will report these RRs in the understudied populations even when the p-interactions are not statistically significant. We will repeat the analyses for total cancer incidence and mortality in women and in men. We will use SAS 9.4. Tests will be 2-sided.

Minimum detectable associations. Estimates are for 2-sided tests with $\alpha=0.05$ and power=80%. For exposures with >2 levels, minimum detectable RRs are for top (riskiest) versus bottom category assuming a linear trend on the ln scale (11). Aim 1a detectable RR range: 1.15 (incidence: 3,561 cases in 13,143 participants; vital exhaustion: Y=30%, N=70%) to 1.35 (mortality: 1,372 deaths in 13,143 participants; lower social support: low=8%, middle=14%, high=78%) or larger. Detectable RRs for cancer sites with 500+ incident cases (breast, prostate) or 125+ deaths (pancreatic, colorectal; 500+ in lung) in 13,143 participants (6,572 for sex-specific site) post Visit 2, minimum detectable RRs range from 1.38 (breast, prostate incidence, vital exhaustion) to 2.65 (pancreatic, colorectal, mortality, social support; 1.62 for lung) or larger. Aim 1b detectable RR range: 1.21 (low SES - incidence: 1,781 cases in 6,572 participants; vital exhaustion) to 1.73 (lower population/density – mortality: 434 cases in 3,286 participants; lower social support) or larger. Detectable interaction sizes range from 1.3 (incidence) to 1.6 (mortality) or larger (NCI’s Power (12)). For sufficient power for site-specific cancers within the subpopulations we will group by organ systems (e.g., female reproductive, male GU, GI) and by associated causes (e.g., smoking, obesity, [female] hormone). If RRs of lower cognitive function (low vs high: CVD RRadj=1.6 (13)) and low social support (low vs high: heart failure RRadj=1.2 (10)) are of the magnitude observed in ARIC previously, then our total cancer incidence and mortality analyses are appropriately powered. We hypothesize that RRs are accentuated in residents in lower population/density areas, those of low SES, and elderly, and thus expect sufficient power for total cancer incidence and mortality in the subpopulation analyses.

**Aim 2: Disparities in the links of aging and psychosocial factors with diagnosis stage and case-fatality**

*Study design and analytic population.* For case-fatality, we will use the prospective cohort study design and include 3,561 participants who attended Visit 2, never had a cancer diagnosis by that time, but were subsequently diagnosed. We will begin follow-up at diagnosis date, use time since diagnosis as the time scale, and end follow-up at date of death from that cancer, death from another cause, withdrawal from ARIC, or end of 2012, whichever comes first. For stage at cancer diagnosis, the design will be cross-sectional.

*Exposures.* We will use the same aging and psychosocial factors as in Aim 1. We will classify the cancer cases respect to their prediagnosis (Visit 2) cognitive function, social support, and vital exhaustion.

*Outcomes.* In ARIC Cancer, we ascertain cancer deaths using the ongoing collection of vital status, cause and date of death by the parent study. For case-fatality, the underlying cause of death must be from the same cancer as diagnosed after Visit 2. 1,372 such deaths have occurred
in 3,561 cases diagnosed post Visit 2. In the U01, we collect stage at diagnosis from 4 cancer registries supplemented with medical record abstraction.

Covariates. We will use the same covariates as in Aim 1.

Statistical analysis. For case-fatality, we will use the same general approach as in Aim 1 including evaluating associations in the 3 understudied groups, but will start follow-up at date of cancer diagnosis and will adjust for time between aging/psychosocial factor assessment and diagnosis date (indicator variables). Separately, we will stratify by time between assessment and diagnosis to determine whether more distant prediagnostic cognitive function, social support, and vital exhaustion are more or less relevant to case-fatality that recent. We will repeat these analyses for solid tumors adjusting for stage at diagnosis (indicator variables) to assess how these factors influence case-fatality beyond effects on stage. In an exploratory analysis, we will adjust for first course of treatment (surgery, radiation, chemotherapy, hormones). We are aware that treatment decision-making may depend on subpopulation (lower population density–access, elderly–chemotherapy tolerance) or aging/psychosocial factors (low cognition/low social support–less aggressive treatment), and could be a mediator; we will interpret cautiously. Prediagnostic access to/use of care are likely highly correlated with lower population/density of the residence and low SES and could influence health status; we will perform a subanalysis restricting to participants living in lower population/density areas or low SES participants with the same level of access/use of care. For diagnosis stage, we will calculate advanced stage prevalence by category of cognitive function, social support, and vital exhaustion overall and stratified by subpopulation. We will estimate the association between advanced stage and aging/psychosocial factors overall and within the subpopulations by logistic regression (or binomial regression, if converges) to adjust for covariates from Aim 1.

Minimum detectable RRs. Aim 2a: For most (vital exhaustion) and least (social support) powered case-fatality analyses, minimum detectable RRs comparing top vs lowest category are 1.27 and 1.44 or larger. Aim 2b: For the most (low SES/vital exhaustion) and least (low population/density of residence/social support) powered subpopulation/exposure, these RRs comparing top vs lowest category are 1.40 and 2.13 or larger. These sized RRs could be of public health import for interventions and policy. We will group by organ system or cause to achieve sufficient power.

Aim 3: Disparities in the aging and psychosocial consequences of cancer
Study design and analytic population. We will use the prospective, closed cohort study design and include 1,150 participants who attended Visit 2, never had a cancer diagnosis by that time, were subsequently diagnosed, and survived to Visit 5 (not time-to-event by design).

Exposures, Outcomes, Covariates. “Exposures” are the subpopulations: residence in a lower population/density area, low SES, elderly (and African-American elderly). Outcomes (Visit 5) are: a) accelerated aging, assessed by low cognitive function (bottom quartile of the Aim 1 variable) and low physical function (bottom quartile; Short Physical Performance Battery); and depression (CES-D score 16+ or top quartile) after diagnosis. Covariates are as in Aim 1.
**Statistical analysis.** For 3a, we will estimate ORs of low cognitive function, low physical function, and high depression scale score in relation to residence in a lower population/density area, low SES, and elderly (and African-American elderly) adjusting for Aim 1 covariates using logistic regression. For cognitive function, we will repeat a) adjusting for Visit 2 cognitive function (quartiles), and b) excluding those with Visit 2 cognitive function in the bottom quartile. For 3b, we will stratify 3a analyses by cognitive function (Visit 2; cutpoint at median) and higher social support (low/medium, high [78%]). For 3c, we will stratify 3a analyses by comorbidities classified as a) present, absent, and b) number (cutpoint at median). We will assess interaction by the likelihood ratio test.

**Minimum detectable ORs:** 3a: For the most (low SES) and least (lower population/density area) powered groups, the minimum detectable ORs are 1.52 and 1.65 or larger. 3b and 3c: The minimum detectable size of interactions range from 2.0 (low SES/number of comorbidities (cutpoint at median) to 2.6 (lower population density residence/social support).

**Limitations:**
We will not be able to study rural residents because few ARIC participants live in truly rural areas (e.g., farming). We will use lower population size or density of residence based on Census tract to attempt to capture this understudied population. For example, we will define lower population/density based on the actual distribution in ARIC, and draw the cutpoint at the 25th percentile. Whether being below the 25th percentile of population/density of the residence will capture what we intend, is unclear.

The sample size may be too small to detect the more modest interactions of exposure*modifier.

We will not be able to fully disentangle location from race in these analyses.

7.a. Will the data be used for non-CVD analysis in this manuscript? __X__ 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? __X__ 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 __X__ 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

__X__ 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)?

Many manuscripts have investigated these “exposures” in association with CVD and relate outcomes, including:
#454
#757
# 880
# 1580
With respect to cancer, the closest is:
#1798

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

11.b. If yes, is the proposal
__X__  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.cscc.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.

13. Per Data Use Agreement Addendum, approved manuscripts using CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes __X__ No.
References