1.a. Full Title: Lifecourse socioeconomic determinants of infectious diseases
b. Abbreviated Title (Length 26 characters): Lifecourse and Infections

2. Writing Group (list individual with lead responsibility first):

   Lead: Allison E. Aiello
   Address: Allison E. Aiello, PhD
   Robert Wood Johnson Health and Society Scholar
   Center for Social Epidemiology & Population Health
   University of Michigan
   1214 S. University, 2nd Floor
   Ann Arbor, MI 48104-2548
   Phone: (734)615-9213       Fax: (734)998-0006
   Email: aielloa@umich.edu

   Writing group members:
   Ana Diez-Roux
   Javier Nieto
   Kathryn Rose

   This manuscript is part of the Lifecourse SES study, Gerardo Heiss, PI.

3. Timeline:

   Submit proposal to Publications Committee: September 2004
   Complete Analysis: November 2004
   Submit draft to Publications Committee: December 2004

4. Rationale:

   Prior infection with herpes simplex viruses (HSV-1), (HSV-2), cytomegalovirus (CMV),
   Chlamydia pneumoniae and Helicobacter pylori have been implicated as etiologic factors in
   chronic illnesses such as coronary artery disease (CAD).1-3 Within the lifecourse literature, early
   childhood SES has been shown to be associated with CAD even after controlling for adult SES.4
   However, consistent mediators of this relationship have not been established. Since several
   infections may be related to atherosclerosis and may also be related to socioeconomic trajectories
   over the lifecourse, infections could be one of the mediating factors linking early life SES
   exposures and SES trajectories to CAD.
   Studies have shown that seropositivity to HSV-1, HSV-2, CMV, and Helicobacter pylori
   is associated with socioeconomic status.5-12 For C. pneumoniae, in some studies infection
appears to be associated with socioeconomic indicators\textsuperscript{13-15} and in others it does not.\textsuperscript{16} Few studies have examined the relationship between socioeconomic status and co-infection with pathogens that have been implicated in CAD\textsuperscript{17} and none have examined whether lifecourse socioeconomic trajectories are associated with co-infection. Therefore, it is important to further characterize the relationship between lifecourse socioeconomic status and pathogen burden as well as overall mean antibody response to HSV-1, HSV-2, \textit{C. pneumoniae} and \textit{H. pylori}.

Generally HSV-1, CMV, \textit{C. pneumoniae}, and \textit{H. pylori} are acquired in early childhood.\textsuperscript{5, 6, 7, 12, 18} HSV-2 is usually acquired as a teen or young adult.\textsuperscript{6, 7} In contrast to lifelong latent infections such as HSV-1, HSV-2 and \textit{C. pneumoniae}, an individual with \textit{H. pylori} infection may clear and be re-infected throughout life.\textsuperscript{18} Therefore, immunological indicators of past infection, such as antibody response to these pathogens, are useful markers for early- and young-adult life acquisition and infection.

Low socioeconomic status in childhood may be associated with nutritional deficiencies, crowding, and hygiene factors which in-turn may provide a conduit for exposure and acquisition of infections.\textsuperscript{19} Over the lifecourse and during adult years, individuals with less favorable socioeconomic characteristics may be more likely to experience psychosocial stress and environmental insults leading to an increased likelihood of reactivation of latent infections. One proposed model underlying the relationship between latent infections and chronic illness is based on an accumulation of risk factors, including but not limited to total pathogen burden, immune response, and persistent reactivity and inflammation.\textsuperscript{1, 2, 20} Therefore, it is important to ascertain whether lifecourse trajectories characterized by low socioeconomic status are associated with infectious illnesses. The ARIC lifecourse SES study linked to the ARIC infection questionnaire and antibody data, provide a unique opportunity to examine whether socioeconomic factors over the lifecourse influences: 1) the type and number of infectious illnesses reported later in life, and 2) immunological parameters of latent infections such as antibody levels of HSV-1, HSV-2, CMV, \textit{C. pneumoniae}, and \textit{H. pylori} as an adult.

5. Main Hypothesis/Study Questions:

1) Childhood and adult socioeconomic factors will be associated with burden of infection (i.e. co-infection) with CMV, HSV-1, HSV-2, \textit{C. pneumoniae} and/or \textit{H. pylori} in middle age.
   a. Hypothesis: Lifecourse trajectories with low socioeconomic characteristics (i.e. low socioeconomic factors in childhood and adulthood) will be significantly associated with increasing number of co-infections among middle age subjects.

2) Childhood and adult socioeconomic factors will influence levels of pathogen specific antibodies among individuals infected with either CMV, HSV-1, HSV-2, \textit{C. pneumoniae} or \textit{H. pylori} in middle age.
   a. Hypothesis: Lifecourse trajectories with low socioeconomic characteristics will be significantly associated with higher levels of antibody response among middle age subjects seropositive for either CMV, HSV-1, HSV-2, \textit{C. pneumoniae} and \textit{H. pylori}. In addition, lifecourse low socioeconomic trajectories will be significantly associated with higher overall mean antibody response among subjects with co-infections.
3) Childhood and adult neighborhood crowding (as indicated by number of individuals living in the households or urban versus rural residence based on the census) will be associated with burden of infection in middle age.
   a. Hypothesis: Lifecourse trajectories characterized by high levels of crowding will be significantly associated with increasing number of co-infections among middle age subjects.

4) Childhood and adult neighborhood crowding will influence levels of pathogen specific antibodies among individuals infected with either CMV, HSV-1, HSV-2, C. pneumoniae or H. pylori in middle age.
   a. Hypothesis: Lifecourse trajectories characterized by high levels of crowding will be significantly associated with higher levels of antibody response among middle age subjects seropositive for either CMV, HSV-1, HSV-2, C. pneumoniae and H. pylori. In addition, lifecourse low socioeconomic trajectories will be significantly associated with higher overall mean antibody response among subjects with co-infections.

5) Childhood and adult socioeconomic factors will be associated with number of reported infectious illness episodes later in life.
   a. Hypothesis: Lifecourse trajectories characterized by low socioeconomic indicators will be significantly associated with increased number of reported diagnoses of several infectious illnesses, including hepatitis, tuberculosis, urinary tract infection, pneumonia, bronchitis, sinusitis, respiratory infection, shingles, and cold sores. In addition, lifecourse low socioeconomic trajectories will be significantly associated with at least one episode of infectious illness in the last 12 months as an adult.

6. Data (variables, time window, source, inclusions/exclusions):
   a) Outcomes:
      The seropositivity status and antibody titers for HSV-1, HSV-2, CMV, C.pneumoniae, and H.pylori are available using the combined data from 3 different ARIC case-cohort data sets. We will only utilize the random cohort sample in order to avoid potential biases that may arise when including incident cases of CAD. For aims one and three the outcome will be total number of infections with either HSV-1, HSV-2, CMV, or C. pneumoniae (as measured by a positive or negative immunoglobulin G (IgG) antibody response). For aims two and four the outcomes variables include: 1) mean IgG antibody levels for each pathogen (HSV-1, HSV-2, CMV, and C. pneumoniae), and 2) overall antibody level for all pathogens combined among those infected with more than one pathogen (this variable will be constructed by scaling the antibody titers for each pathogen and creating an overall summary score for all infections combined). For aim five the outcome variables include: 1) total number of reported diagnoses of infectious illnesses (hepatitis, tuberculosis, urinary tract infection, pneumonia, bronchitis, sinusitis, respiratory infection, shingles, and cold sores), 2) total number of reported infectious illnesses in which the subject had at least one episode of infection in the past 12 months.

   b) Independent variables:
      Individual socioeconomic indicators at childhood and adulthood will be used for the analysis. Initially, six separate dimensions will be investigated: education, occupation, home ownership, number of individuals in the household, population density, and area of residence.
Parental education (a marker of childhood socioeconomic position) is available from ARIC visit 4 and adult education is available from the ARIC baseline. Parental occupation when participant was a child and participant occupation at ages 30, 40, 50, is available from the Lifecourse Ancillary Study both in census categories and characterized based on the criteria developed by Wright. Participant occupational status during the baseline and subsequent exams is available from the ARIC Visit 1-4 datasets. Parental home ownership and participant home ownership at ages 30, 40, 50, and at the ARIC baseline are available from the Lifecourse Ancillary Study. Parental place of residence during childhood (at the county level) and participant place of residence at ages 30, 40, and 50 (at the tract level) is available from the Lifecourse Ancillary Study. The data for number of individuals in the household is available for childhood and ages 30, 40, 50, and for the most recent ARIC adult follow-up. Last, population density is available for childhood, ages 30, 40, 50, and the most recent ARIC adult follow-up (the data is available at the county level from 1930 to 1950 (childhood) and at the tract level from 1960 and beyond (ages 30-50).

Each SEP indicator for each time period will be classified into three categories (low, middle, high). Analyses for research questions 1-2 will be restricted to parental (childhood) and adult indicators collected at the latest available ARIC follow-up. Categories will be constructed based on cut-offs used in the literature. Alternatively, percentile-based categories (based on distribution observed at each time period in the sample) will be used. Analyses of trajectories (research question 3) will employ the full range of indicators available (childhood, ages 30, 40, 50, and the most recent ARIC adult measure).

c) Inclusions/Exclusions:

Participans with missing data on the outcomes of interest and with missing relevant lifecourse SES variables will be excluded from the analyses.

d) Covariates:

Age, sex, center, smoking, obesity, and race/ethnicity will be examined as covariates. Heterogeneity in associations will be investigated if sample size allows.

e) Statistical Analysis:

Power Analyses

For aims one through four we will only utilize the data from subjects in the cohort random sample (identified in data set UC250921 by the variable name INSAM2) that were derived from the ARIC incident CHD case-cohort studies. The total number of subjects that are available for analysis (i.e. subjects with both viral antibody data as well as lifecourse socioeconomic indicators) is N=357. For aim five we will utilize participants who attended the V4 exam (when the infection questionnaire was administered) and also participated in the AFU during which the LC-SES questionnaire was completed.

Since our sample size is limited by the availability of laboratory analyses for antibody titer for aims one through four, we conducted power analyses using one of the potential trajectory variables of interest (home ownership trajectory) and one outcome variables of interest (CMV antibody titer) using NCSS and Pass© (www.ncss.com):

1) **Logistic Regression:** A logistic regression of a binary outcome response Y (high infection antibody titer (Y=1) versus low antibody titer (Y=0)) on a binary independent variable X (subjects parents owned a home as a child and subject owns a home as an adult (X=1) versus either parents did not own a home when subject was a child or subject as an adult did not own a home (X=0)) with a sample size of 357 observations (of which 39% X=0 and 61% X=1)
achieves 80% power at a 0.05 significance level to detect an OR ≥ 2.05. Although there is no
literature directly examining CMV antibody titer and socioeconomic status among a population
based sample, this odds ratio is compatible with differences in the proportion of subjects co-
infected with HSV among those below versus above the poverty index in the NHANES sample
(i.e. 25% of subjects co-infected were below poverty index and 15% were above). 
This power
calculation also includes adjustment for multiple regression where the R-squared of the
independent variable on other variables in the model is approximately 0.10.

2) **Linear Regression:** A sample size of 357 achieves 80% power using an F-test to
detect a difference in antibody titer of 0.73 units under the alternative hypothesis when the
standard deviation of the independent variable (i.e. lifecourse trajectory of home ownership) is
0.49 (estimated from the available home ownership trajectory data for the 357 subjects), the
standard deviation of the outcome variable CMV antibody titer is 2.42 (estimate provided from
the ARIC case-cohort study), and the significance level is set at 0.05. Although there are little
data examining the influence of socioeconomic factors on mean antibody titer, the slope estimate
of 0.73 is less than the difference in mean antibody titer when comparing CHD cases versus
controls for CMV (1.1), but slightly above the mean differences between cases and controls for
HSV-1 (0.2) and HSV-2 (0.1) as reported in an earlier ARIC study. Thus, we will have
sufficient power to detect changes in mean antibody titer of ≥ 0.73 among subjects with low
socioeconomic lifecourse trajectories.

**Statistical Models**

We will first conduct descriptive analyses. Next, linear or logistic regression will be
used to estimate independent associations of childhood and adult SEP with antibody level and
pathogen burden after adjustment for age. Exploratory trajectory analysis (mixture models using
PROC TRAJ in SAS) will be used to explore trajectories in SEP present in the data. Using these
results (plus a priori hypotheses) participants will be categorized into trajectory types. These
trajectory types will be examined in relation to infectious outcomes in adulthood using linear and
logistic regression. Results will be shown before and after adjustment for age. Analyses will be
stratified by sex and race and also by pooled estimates where sample size is insufficient to
examine relationships within strata.

e) References

1. Epstein SE. The multiple mechanisms by which infection may contribute to

2. Epstein SE, Zhou YF, Zhu J. Infection and atherosclerosis: emerging mechanistic

factor for carotid intimal-medial thickening, a measure of subclinical atherosclerosis.

4. Galobardes B, Lynch JW, Davey Smith G. Childhood Socioeconomic Circumstances and
Cause-specific Mortality in Adulthood: Systematic Review and Interpretation. *Epidemiol

5. Becker TM, Lee F, Daling JR, Nahmias AJ. Seroprevalence of and risk factors for
antibodies to herpes simplex viruses, hepatitis B, and hepatitis C among southwestern

6. Fleming DT, McQuillan GM, Johnson RE, et al. Herpes simplex virus type 2 in the


19. West R. Commentary: adjustment for potential confounders may have been taken too far. *Bmj.* Jul 22 2000;321(7255):213.


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

b. If Yes, is the author aware that the file ICTDER02 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 ICTDER02 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 ICTDER02 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://bios.unc.edu/units/cscc/ARIC/stdy/studymem.html

  __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)?

11. 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.