

## Original Paper

**Cite this article:** Noppert GA, Aiello AE, O’Rand AM, Cohen HJ (2018). Investigating pathogen burden in relation to a cumulative deficits index in a representative sample of US adults. *Epidemiology and Infection* **146**, 1968–1976. <https://doi.org/10.1017/S095026881800153X>

Received: 29 December 2017

Revised: 19 April 2018

Accepted: 16 May 2018

First published online: 14 June 2018

**Key words:**

Epidemiology; herpesviruses; infectious disease epidemiology; virus infection; social determinants

**Author for correspondence:**

G. A. Noppert,  
E-mail: [gnop@email.unc.edu](mailto:gnop@email.unc.edu)

# Investigating pathogen burden in relation to a cumulative deficits index in a representative sample of US adults

G. A. Noppert<sup>1,2,3</sup>, A. E. Aiello<sup>3,4</sup>, A. M. O’Rand<sup>2</sup> and H. J. Cohen<sup>1,5</sup>

<sup>1</sup>Center for the Study of Aging and Human Development, Duke University, Durham, USA; <sup>2</sup>Duke University Population Research Institute, Duke University, Durham, USA; <sup>3</sup>Carolina Population Center, University of North Carolina, Chapel Hill, USA; <sup>4</sup>Department of Epidemiology Gillings School of Global Public Health, University of North Carolina, Chapel Hill, USA and <sup>5</sup>Claude D. Pepper Older Americans Independence Center, Durham, USA

**Abstract**

Pathogen burden is a construct developed to assess the cumulative effects of multiple, persistent pathogens on morbidity and mortality. Despite the likely biological wear and tear on multiple body systems caused by persistent infections, few studies have examined the impact of total pathogen burden on such outcomes and specifically on preclinical markers of dysfunction. Using data from two waves of the National Health and Nutrition Examination Survey, we compared three alternative methods for measuring pathogen burden, composed of mainly persistent viral infections, using a cumulative deficits index (CDI) as an outcome: single pathogen associations, a pathogen burden summary score and latent class analyses. We found significant heterogeneity in the distribution of the CDI by age, sex, race/ethnicity and education. There was an association between pathogen burden and the CDI by all three metrics. The latent class classification of pathogen burden showed particularly strong associations with the CDI; these associations remained after controlling for age, sex, body mass index, smoking, race/ethnicity and education. Our results suggest that pathogen burden may influence early clinical indicators of poor health as measured by the CDI. Our results are salient since we were able to detect these associations in a relatively young population. These findings suggest that reducing pathogen burden and the specific pathogens that drive the CDI may provide a target for preventing the early development of age-related physiological changes.

**Introduction**

A growing body of research has documented the impact of persistent viral and chronic bacterial infections on a host of health outcomes, including cardiovascular disease [1, 2] and cognitive impairment [3] as well as markers of cellular aging such as telomere length [4]. Persistent infections may operate through several biological mechanisms such as tissue damage, localised inflammation and/or cellular processes [5]. Of particular interest are the chronic effects of the reactivation of latent infections over the life course which cause the release of pro-inflammatory cytokines, such as interleukin-6 and acute-phase, C-reactive protein (CRP) both of which have been implicated in many of the hallmark diseases of aging. These infections likely result in biological dysfunction across multiple body systems, beginning early in the life course before clinical deficits are manifest by traditional biomarker assessment [6]. Thus, using a clinical laboratory value-based deficits index that detects those at the higher (or lower) end of the normal range may highlight individuals on a trajectory for the development of multiple, chronic conditions.

Given the increasing numbers of chronic diseases with an infectious etiology, understanding the biological consequences of persistent infection is of critical importance. Pathogen burden, or infection burden, is a construct developed to assess the cumulative effects of multiple, persistent pathogens on morbidity and mortality [7]. Pathogen burden has been measured in several different ways [1, 8–13]. Some studies have focused on single pathogen associations examining the impact of each individual infection on a distal health outcome. Other studies have examined pathogen burden as a summary index of all pathogens for which an individual is seropositive. In 2015, Simanek *et al.* conducted a study that demonstrated the need to go beyond the ‘black box’ of pathogen burden, by incorporating both the number of pathogens and the specific combinations of different pathogens that may influence health [8].

Few studies have explicitly examined the impact of the pathogen on a cumulative deficits index (CDI) among individuals in young and mid-adulthood. Using data from two waves of the National Health and Nutrition Examination Survey (NHANES), we compared three alternative methods for measuring pathogen burden using a laboratory-based CDI as an outcome [14]. The CDI offers a way to quantify multi-system biological dysfunction before

clinical disease is apparent [15]. We hypothesised that latent class methods that incorporate both the number and combination of pathogens would show the most robust relationship between pathogen burden and the CDI.

## Methods

### Study sample

We used two waves of the continuous NHANES collected by the National Center for Health Statistics. NHANES is a cross-sectional, nationally representative survey of the health and nutritional status of the US non-institutionalised, civilian population aged 2 months and older. Our study population was drawn from the 2003–2004 and 2009–2010 waves. Details of the sampling strategy for the continuous NHANES can be found at: <https://wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx>.

For each of the NHANES 2003–2004 and 2009–2010 waves, individuals provided information on a range of demographic characteristics based on the in-home interview. A subset of individuals also participated in the physical examinations and laboratory studies performed at the mobile examination center. We selected two waves so as to replicate the findings across samples and with slightly different combinations of pathogens. The 2003–2004 and 2009–2010 waves were preferable due to the quantity and types of pathogens tested in these waves. We did not use a combined study sample because we were also interested in testing the association of individual pathogens (not all of which were the same in each wave) with the CDI.

Our study sample included 20–49 years old who participated in the survey, physical examination and laboratory studies (Supplementary Fig. 1). The study sample included pregnant women, confirmed either by self-report or laboratory testing. There were 14 623 participants in the 2003–2004 wave and 10 846 in the 2009–2010 wave. Of those, 4980 participants were excluded because they did not participate in the interview, physical examination and laboratory studies in the 2003–2004 wave; 560 in the 2009–2010 wave. An additional 6934 individuals were excluded for whom there was not complete pathogen sample information available in the 2003–2004 wave; 8043 in the 2009–2010 wave. The final sample included only those with complete covariate information (2003–2004 Wave Final  $N=2168$ ; 2009–2010 Wave Final  $N=2546$ ).

This study was approved by the Campus Institutional Review Board of Duke University.

### Laboratory analyses

Laboratory information used to construct the CDI was based on the biomarkers assessed from blood serum collected as part of the standard biochemistry profile of the NHANES laboratory examination. Additional information on specimen collection can be found in the NHANES Laboratory/Medical Technologists Procedures Manual. The specific biomarkers included in the CDI are shown in Table 1. The pathogen measurement information is included in the Supplementary files.

### Measures

The pathogen burden summary score was created by summing the number of pathogens for which an individual was seropositive. The construction of the CDI followed a methodology utilised

by King *et al.* [15] and others examining laboratory indices of frailty [14, 16]. Twenty-eight biomarkers were incorporated in the CDI such as albumin levels, white blood cell count, calcium levels and cholesterol. Table 1 describes each of the 28 biomarkers utilised in the CDI along with their corresponding body system/process and the cut-points utilised. King *et al.* used cut-points based on clinical significance in older adults. Given the young age of our study sample, many are not likely showing abnormal values for the biomarkers assessed. Thus, we utilised an approach aimed at ascertaining those at the higher/lower end of the normal range depending on the specific biomarker. This determination was made for each biomarker depending on the direction indicating dysfunction. Each biomarker was split into quartiles and those in the highest (or lowest) quartile of the distribution received a score of 1. For some biomarkers, both high and low values can indicate dysfunction. For those markers, the distribution was split into eighths and both the top eighth and bottom eighth received a score of 1. This approach has been used in other biomarker studies [17]. The final CDI was constructed by summing the number of biomarkers for which an individual received a score of 1 divided by the total number of biomarkers available for that individual. We then examined the distribution of the CDI in the population as a whole and then by age, gender, race/ethnicity, education, smoking behavior and body mass index (BMI).

### Covariates

We included covariates in our regression analyses that are well established in the literature as relating to pathogen burden and/or the CDI. These covariates included basic demographic information as well as health characteristics.

Basic demographic variables included age, sex, race/ethnicity and education. Age was analysed as a continuous variable in years. Sex was binary coded as male or female. Race/ethnicity was categorised according to the NHANES guidelines as non-Hispanic White, non-Hispanic Black, Mexican American, other Hispanic and other race. Education was treated categorically according to NHANES classifications as less than a high school education, high school graduate, some college education and college graduate and above.

Health status characteristics included BMI and smoking status. BMI was examined categorically with those with a BMI less than 18.5 classified as underweight; 18.5–25 classified as normal; 25–30 classified as overweight and greater than 30 classified as obese. Smoking status was classified as having ever smoked *vs.* never smoked.

### Statistical analyses

#### Single pathogen associations

Linear regression models were constructed to investigate the association between each individual pathogen and the CDI. We estimated the mean difference and 95% confidence interval (CI) for those with- and without infection.

#### Pathogen burden summary score

We investigated the association between the pathogen burden summary score in two different ways. First, we examined the association between the continuous pathogen burden summary score and the CDI using linear regression models. The incremental increase in the CDI and 95% CI were calculated for each one-unit increase in the pathogen summary score. We then split the

**Table 1.** Biomarkers included in the laboratory-based cumulative deficits index based on the National Health and Nutrition Examination Survey

Biomarker	Body system/Process	Early dysfunction cutpoints <sup>a</sup>	% Early dysfunction
Albumin (µg/ml)	Liver, nutrition	≤4.0	28.1
AST (units/l)		≥27.0	27.7
Calcium (mg/dl)	Bone, nutrition	≤9.1; ≥9.9	30.5
Creatinine (mg/dl)	Kidney	≥1.0	27.5
Glucose (mg/dl)	Pancreas	≥94	25.9
Haemoglobin (g/dl)	Bone marrow	M: ≤14.9; F: ≤12.6	26.6
Alkaline phosphatase (units/l)	Liver, biliary tract	≥79.0	25.9
Phosphorus (mg/dl)		≤3.2; ≥4.5	26.1
Potassium (mmol/l)	Kidney, endocrine	≤3.6; ≥4.3	29.4
Total protein (g/dl)	Nutrition, inflammation & myeloma	≥7.6	25.5
Sodium (mmol/l)	Kidney, liver, heart	≤137; ≥141	22.1
BUN (mg/dl)	Kidney	≥13	27.2
White blood cell count (1000 cells/µl)	Inflammation	≥8.9	25.9
Platelet count (1000 cells/µl)	Inflammation	≤201; ≥345	25.4
Segmented neutrophils number		≥5.6	25.9
Cholesterol (mg/dl)	Liver, heart risk	≥225	25.1
HDL (mg/dl)	Heart risk	M: ≤40; F: ≤47	26.9
Uric acid (mg/dl)	Joints, kidney	M: ≤6.7; F: ≤5.0	27.7
Chloride (mmol/l)	Metabolism, kidney	≤101; ≥106	37.7
Lactic dehydrogenase (units/l)	Liver, malignancy	≥135	25.6
GGT (units/l)	Liver	≥29	25.4
Bilirubin (mg/dl)	Biliary tract, gall bladder	≥0.90	29.0
Triglycerides (mg/dl)	Metabolism, heart risk	≥155	25.1
Globulins (g/dl)	Inflammation	≥3.3	26.9
SGPT (units/l) (ALT)	Liver	≥31	25.1
Bicarbonate (milli-EQ/l)	Metabolism, liver, kidney	≤22; ≥27	20.6
C-Reactive Protein	Inflammation	≥0.49	25.3
Iron (µg/dl)		≤46; ≥129	25.7

M, male; F, female.

<sup>a</sup>Early dysfunction based on below or above the quartile cutpoint.

Results based on 2003–2004 wave; process was replicated in subsequent waves.

pathogen burden summary score into categories: 0–1 pathogens, 2 pathogens, 3 pathogens, and 4+ pathogens. Using linear regression models, we then examined the mean difference in the CDI using 0–1 pathogens as the referent.

#### Latent class analyses

To further characterise pathogen burden, we performed latent class analyses on each wave of data. This method assigns individuals with similar profiles to mutually exclusive classes using both manifest (measured) and latent (unmeasured) characteristics. We followed a three-step inclusive classify-analyse approach developed by Bray *et al.* [18]. This approach separates the classification of individuals into latent classes from the regression of the outcome on the latent classes.

In the first phase, we determined the appropriate number of latent classes for each wave by comparing a series of model fit

indices: the likelihood-ratio value (G<sub>2</sub>), Akaike information criteria (AIC), Bayesian information criterion (BIC), sample size-adjusted BIC, log-likelihood and entropy. Lower values for the AIC, BIC, adjusted AIC are preferred as they indicate better model fit. Higher values of entropy indicate the greater distinction between the classes. When deciding on the optimal number of classes, we prioritised both model fit and interpretability of the classes. Model estimation was repeated 1000 times with random starts to confirm model identification. Latent class models were estimated using PROC LCA SAS Version 9.4.

Once the appropriate number of latent classes was determined, we assigned individuals to a specific latent class based on their posterior probability of most likely class membership. Following the inclusive classify-analyse approach, the outcome variable, the CDI, was used as a covariate in the final model to reduce bias in estimation in the outcome model [18].

### Regression analyses

Using the three methods of classifying pathogen burden (single pathogen association, pathogen burden summary score and latent class analysis), we constructed a series of linear regression models to examine the association with each measure of pathogen burden and the CDI. For each method of pathogen burden, model 1 shows the association without any control variables (i.e. unadjusted model); model 2 shows the association with base controls (i.e. age, sex, BMI and smoking status); model 3 shows the association with base controls and race/ethnicity status; model 4 shows the association with base controls and education; and model 5 is the full model with base controls, race/ethnicity and education. These models were replicated using wave 1 and wave 2 study samples.

All regression analyses used appropriate sampling weights and adjustments to account for the complex survey design features. A two-side alpha of 0.05 was used to determine significance in all statistical analyses.

Statistical analyses were performed in SAS v.9.4 (Cary, NC).

## Results

### Sample characteristics

Wave 1 (2003–2004) consisted of 2168 individuals (Table 2). The mean age was 34.9 years and the sample was 69% non-Hispanic White. One-quarter of participants had a college education or above. Over half (52%) of the sample reported no smoking history. Moreover, a majority of the sample was classified as overweight or obese (64%). Wave 2 (2009–2010) consisted of 2546 individuals. The mean age was 34.7 years and the sample was 62% non-Hispanic White. In wave 2, 28% of participants had a college education or above. Again, over half the sample (58%) reported no smoking history and a majority were overweight or obese (66%).

In wave 1, we examined the proportion of the population that was seropositive to seven pathogens. Of the wave 1 sample, 60% were seropositive to HSV-1; 19% to HSV-2; 25% to HPV; 13% to toxoplasmosis; 52% to CMV; 0.44% to HIV; and 2% to syphilis. In wave 2, we examined the proportion of the study population seropositive to five pathogens. Of the wave 2 sample, 57% were seropositive to HSV-1; 15% to HSV-2; 34% to HPV; 12% to toxoplasmosis; and 0.37% to HIV. Pathogen correlations are given in Supplementary Table 1.

### Cumulative deficits index

In wave 1, the CDI ranged from 0.04 to 0.78 with a mean of 0.30 (standard error of the mean = 0.003) (Table 3, Supplementary Fig. 2a). Increasing age was associated with an increase of 0.002 in the CDI ( $P < 0.001$ ). There were also significant differences in the CDI by sex, race/ethnicity, education, smoking status and BMI.

In wave 2, the CDI ranged from 0 to 0.74 with a mean of 0.27 (standard error of the mean = 0.004) (Table 3, Supplementary Fig. 2b). Similar trends were observed in the wave 2 study population for race/ethnicity, education, smoking status and BMI.

### Latent class assignment of pathogen burden

Model fit statistics were assessed for latent class models with one to six latent classes. Based on both model fit indices and

**Table 2.** Population-weighted demographic and pathogen characteristics of the study population in the National Health and Nutrition Examination Survey, 2003–2004 and 2009–2010

	Wave 1 2003–2004 (N = 2168)	Wave 2 2009–2010 (N = 2546)
Age		
Mean age (s.d.)	34.9 (0.30)	34.7 (0.21)
	N (%)	N (%)
Sex		
Female	1130 (50)	1365 (52)
Race/Ethnicity		
Non-Hispanic White	1045 (69)	1127 (62)
Non-Hispanic Black	500 (12)	427 (12)
Mexican American	471 (10)	538 (12)
Other Hispanic	82 (4)	299 (7)
Other race	70 (4)	155 (7)
Education		
Less than high school	484 (15)	655 (18)
High school graduate	569 (27)	588 (22)
Some college	676 (34)	772 (32)
College graduate and above	439 (25)	531 (28)
Ever smoker		
Yes	975 (48)	1073 (42)
No	1193 (52)	1473 (58)
Body Mass Index (BMI)		
Under weight (<18.5)	41 (2)	43 (2)
Normal (≥18.5; <25)	703 (34)	772 (32)
Overweight (≥25; <30)	709 (33)	788 (31)
Obese (≥30)	715 (31)	943 (35)
Pathogens seropositivity		
Herpes simplex virus Type 1	1397 (60)	1637 (57)
Herpes simplex virus Type 2	472 (19)	566 (15)
Human papillomavirus	542 (25)	945 (34)
Toxoplasmosis	296 (13)	372 (12)
Human immunodeficiency virus	13 (0.44)	14 (0.37)
Cytomegalovirus	1315 (52)	NM
Syphilis	51 (2)	NM

NM, not measured.

interpretability, a three-class solution seemed best for both waves 1 and 2 as it optimised BIC, adjusted BIC and entropy, with near optimal levels of AIC (Supplementary Tables 2a and b).

Figures 1a and b show the conditional probabilities of pathogen positivity for each latent class. For each wave, one class of individuals had lower probabilities of being positive on all pathogens, the 'low burden' class. Similarly, for each wave one class of individuals had higher probabilities of being positive on several of the pathogens tested, the 'high burden' class. Finally, each wave had a significant

**Table 3.** The descriptive statistics of the cumulative deficits index and the distribution of the cumulative deficits index by key demographic characteristics in the National Health and Nutrition Examination Survey, 2003–2004 and 2009–2010.

	Wave 1		Wave 2	
	2003–2004		2009–2010	
	N = 2168		N = 2546	
Mean cumulative deficits index	0.30		0.27	
	Mean	<i>P</i>	Mean	<i>P</i>
Age (year)	0.0017	<0.001	0.0025	<0.001
Gender				
Male	Ref.		Ref.	
Female	–0.075	<0.001	–0.068	<0.001
Race/Ethnicity				
Non-Hispanic White	Ref.		Ref.	
Non-Hispanic Black	0.027	<0.001	0.035	0.001
Mexican American	0.04	<0.001	0.031	<0.05
Other Hispanic	0.045	0.01	0.024	0.06
Other race	0.025	0.05	0.0004	0.97
Education				
College grad and above	Ref.		Ref.	
Some college	0.021	<0.01	0.036	<0.001
High school graduate	0.035	<0.001	0.045	<0.001
Less than high school	0.048	<0.001	0.06	<0.001
Ever smoker				
No	Ref.		Ref.	
Yes	0.021	<0.01	0.013	<0.05
Body Mass Index (BMI)				
Under weight (less than 18.5)	–0.010	0.54	0.011	0.41
Normal ( $\geq 18.5$ ; <25)	Ref.		Ref.	
Overweight ( $\geq 25$ ; <30)	0.054	<0.001	0.045	<0.001
Obese ( $\geq 30$ )	0.1	<0.001	0.11	<0.001

Results based on linear regression models; *P* values reflect a two-sided test.

portion of individuals who were positive for common pathogens such as CMV and HSV-1, the ‘common pathogens’ class. For statistical modeling, the low burden class was used as the referent.

For wave 1, 50% of participants were assigned to the low burden class; 37% to the common pathogens class; and 13% to the high burden class. For the wave 2 sample, 37% of the participants were assigned to the low burden class; 46% to the common pathogens class; and 17% to the high burden class.

## Regression results

### Single pathogen associations

In wave 1 unadjusted single pathogen models, we observed a significant relationship between CMV, HSV-1, HPV, Toxoplasmosis, HIV and the CDI (Table 4). With the exception of HPV, infection with a pathogen resulted in a greater mean CDI compared with no infection. However, in the full model (model 5) only HIV remained significantly associated with the CDI. Individuals

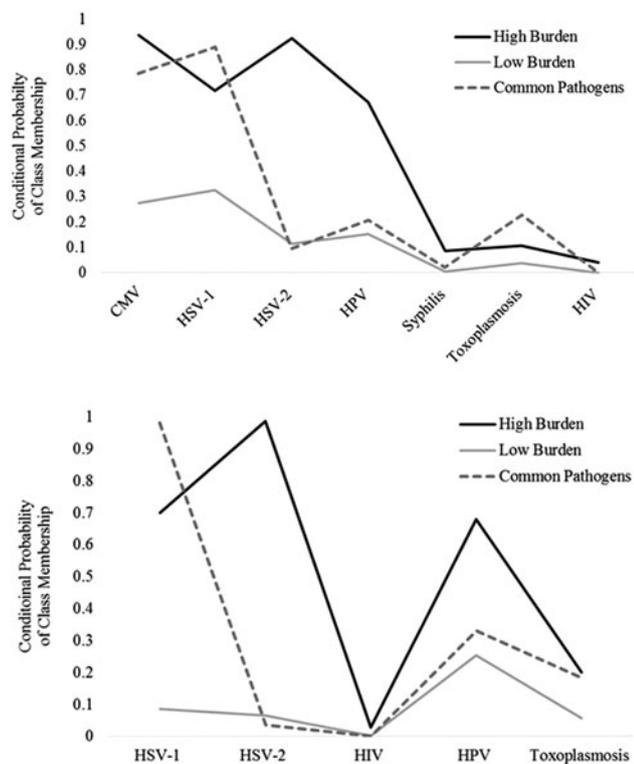
with HIV had a mean CDI 0.11 higher than those without HIV ( $P < 0.05$ ).

In wave 2, a similar pattern was observed with the exception of HPV. In the full model (model 5), both HPV and HIV were significantly associated with the CDI. Individuals with HPV had a mean CDI 0.01 higher than those without HPV ( $P < 0.05$ ). Those with HIV had a mean CDI 0.12 higher than those without HIV ( $P < 0.01$ ).

### Pathogen summary score

A significant relationship was observed between the pathogen summary score and the CDI in models 1–4 in wave 1. However, in the full model (model 5) this relationship was null ( $P = 0.11$ ). Similarly, when the pathogen summary score was split into categories, there were no significant relationships observed in the full model.

While a significant relationship between the pathogen summary score and the CDI appeared in models 1–4 in wave 2, in



**Fig. 1.** (a) The conditional probability of each infection given a three class solution for the National Health and Examination Survey, 2003–2004 wave ( $N = 2168$ ). (b) The conditional probability of each infection given a three class solution for the National Health and Examination Survey, 2009–2010 wave ( $N = 2546$ ).

the full model this relationship was null ( $P = 0.06$ ). However, when the pathogen summary score was split into categories, a significant relationship across models was observed between individuals having four or more pathogens compared with those having 0–1 pathogens. Those having four or more pathogens had a mean CDI 0.06 higher than those having only 0–1 pathogens ( $P < 0.001$ ).

#### Latent class analyses

A significant relationship between latent classes of pathogen burden and the CDI was observed across all models in wave 1. In the full model (model 5), those in the high burden class had a mean CDI 0.03 higher than those in the low burden class ( $P < 0.001$ ). Those in the common pathogens class had a mean CDI 0.04 higher than those in the low burden class ( $P < 0.001$ ).

Similarly, in wave 2 there was a significant relationship between those in the high burden class and those in the low burden class with the CDI across models. In model 5, those in the high burden class had a mean CDI 0.02 greater than those in the low burden class ( $P < 0.05$ ).

#### Sensitivity analyses

The finding in the single pathogen analysis in wave 1 that infection with HPV and syphilis both resulted in a lower mean CDI did not follow our hypothesised direction. We performed sensitivity analyses in which we removed HPV and syphilis from the pathogen summary score. We found modest changes in the association between the pathogen summary score and the CDI; however, the precision of the estimates did increase (Supplementary Table 3).

We also performed additional analyses excluding HIV from the pathogen burden summary score. While there was a modest attenuation in the effect size across models, the inferences regarding the associations were unchanged. In the full model (model 5), there remained a null relationship between the pathogen burden summary score and the CDI (results not shown).

We ran sensitivity analyses controlling for self-reported chronic conditions. Nearly a quarter (21%) of the study sample had one or more self-reported chronic health conditions. When we controlled for this in the analyses, we found no substantive changes in the effect estimates or the inferences. These results lend further support to the CDI as a preclinical marker of biological dysfunction. We also tested whether excluding pregnant women would change the association between the infections and the CDI. Excluding pregnant did not result in a substantial change in either the effect estimates or the inferences.

## Discussion

We used a cumulative deficits approach to examine the impact of total pathogen burden on biological functioning in a nationally representative sample of individuals aged 20–49 years. We found significant heterogeneity in the distribution of the CDI by age, sex, race/ethnicity and education suggesting that it may indeed be detecting underlying disparities in biological dysfunction, even in a middle-aged sample (aged 20–49 years). While other studies have demonstrated the impact of persistent infections on morbidity and mortality later in the life course, our findings demonstrate the impact of these infections on biological function before clinical disease typically manifests. We also compared several methods of assessing pathogen burden and their associations with the CDI. Our results suggest that the CDI may be a useful way to assess the impact of pathogen burden. Latent class analyses, in particular, may be advantageous for characterising pathogen burden as they can account for both the number and combinations of pathogen burden, as well as detect and adjust for unobserved heterogeneity in the population.

Detecting biological wear and tear before clinical dysfunction is critical to understanding patterns of morbidity and mortality throughout the life course and particularly how they relate to the aging process. Many of the hallmark diseases of aging begin with physiological changes earlier in the life course [19]. A growing body of work is examining biomarkers of aging [20] in younger cohorts as a way to pinpoint when in the life course these changes happen as well as to design interventions that may slow the aging process in younger individuals. The CDI we employed can be conceptualised as a manifestation of the health status of multiple body systems. Since our population spanned young adulthood to middle age, our aim in the use of the CDI was to detect individuals at pre-dysfunction levels, on a trajectory towards the development of multiple clinical deficits. The significant variation we observed by age, gender, race/ethnicity, education, smoking status and BMI suggest that CDI may be useful in detecting pre-dysfunction in a younger cohort. Additionally, the social disparities by race/ethnicity and education suggest that the CDI may be another method for examining the origins of health disparities. Current work is under way to further investigate these social disparities.

An increasing number of age-related processes and chronic diseases have been shown to have an infectious aetiology [21, 22]. Persistent infections are often subclinical and may not cause severe illness at the time of infection [23–25] but the

**Table 4.** Results of the regression analysis examining the association between measures of pathogen burden and the cumulative deficits index in the National Health and Nutrition Examination Survey, 2003–2004 and 2009–2010

Single pathogen association	Model 1: No controls		Model 2: Base controls (Age, Sex, BMI, and Smoking)		Model 3: Base + Race		Model 4: Base + Education		Model 5: Full	
	Coeff	P	Coeff	P	Coeff	P	Coeff	P	Coeff	P
<i>Wave 1: N = 2168</i>										
CMV	<b>0.01</b>	<b>&lt;0.05</b>	<b>0.02</b>	<b>&lt;0.001</b>	0.01	0.09	<b>0.02</b>	<b>&lt;0.01</b>	0.01	0.15
HSV-1	<b>0.02</b>	<b>&lt;0.01</b>	<b>0.01</b>	<b>0.05</b>	0.01	0.27	0.01	0.16	0.01	0.42
HSV-2	0.01	0.15	0.01	0.08	0.01	0.39	0.01	0.19	0.00	0.47
HPV	−0.02	<b>&lt;0.01</b>	−0.01	0.21	−0.01	0.25	−0.01	0.22	−0.01	0.24
Syphilis	−0.02	0.58	−0.01	0.63	−0.02	0.41	−0.02	0.48	−0.02	0.36
Toxoplasmosis	<b>0.02</b>	<b>&lt;0.05</b>	<b>0.01</b>	<b>0.05</b>	0.01	0.29	0.01	0.12	0.01	0.36
HIV	<b>0.12</b>	<b>&lt;0.01</b>	<b>0.12</b>	<b>&lt;0.01</b>	<b>0.11</b>	<b>&lt;0.05</b>	<b>0.11</b>	<b>&lt;0.01</b>	<b>0.11</b>	<b>&lt;0.05</b>
Pathogen summary score										
Continuous	<b>0.01</b>	<b>&lt;0.01</b>	<b>0.01</b>	<b>0.00</b>	<b>0.004</b>	<b>0.04</b>	<b>0.01</b>	<b>&lt;0.01</b>	0.003	0.11
Categories										
0–1 Pathogens	<b>Ref.</b>		Ref.		Ref.		Ref.		Ref.	
2 Pathogens	0.01	0.46	0.01	0.28	0.00	0.97	0.01	0.53	−0.001	0.89
3 Pathogens	0.01	0.28	<b>0.02</b>	<b>&lt;0.05</b>	0.01	0.44	0.01	0.11	0.004	0.58
4+ Pathogens	0.01	0.23	<b>0.02</b>	<b>&lt;0.05</b>	0.01	0.34	0.02	0.12	0.01	0.58
Latent class analysis										
'Low burden'										
'Common pathogens'	<b>0.05</b>	<b>&lt;0.001</b>	<b>0.05</b>	<b>&lt;0.001</b>	<b>0.04</b>	<b>&lt;0.001</b>	<b>0.04</b>	<b>&lt;0.001</b>	<b>0.04</b>	<b>&lt;0.001</b>
'High Burden' class	<b>0.04</b>	<b>&lt;0.001</b>	<b>0.04</b>	<b>&lt;0.001</b>	<b>0.03</b>	<b>&lt;0.001</b>	<b>0.04</b>	<b>&lt;0.001</b>	<b>0.03</b>	<b>&lt;0.001</b>
<i>Wave 2: N = 2246</i>										
HSV-1	<b>0.02</b>	<b>&lt;0.01</b>	<b>0.02</b>	<b>&lt;0.05</b>	0.01	0.07	0.01	0.17	0.01	0.29
HSV-2	<b>0.02</b>	<b>&lt;0.01</b>	<b>0.02</b>	<b>&lt;0.05</b>	0.01	0.20	0.01	0.19	0.004	0.55
HPV	−0.01	<b>&lt;0.05</b>	<b>0.01</b>	<b>&lt;0.01</b>	<b>0.01</b>	<b>&lt;0.05</b>	<b>0.01</b>	<b>&lt;0.05</b>	<b>0.01</b>	<b>&lt;0.05</b>
Toxoplasmosis	<b>0.03</b>	<b>&lt;0.01</b>	0.01	0.220	0.005	0.49	0.00	0.73	0.00	0.93
HIV	<b>0.16</b>	<b>&lt;0.001</b>	<b>0.13</b>	<b>&lt;0.01</b>	<b>0.12</b>	<b>&lt;0.01</b>	<b>0.13</b>	<b>&lt;0.01</b>	<b>0.12</b>	<b>&lt;0.01</b>
Pathogen summary score										
Continuous	<b>0.01</b>	<b>&lt;0.001</b>	<b>0.01</b>	<b>&lt;0.001</b>	<b>0.01</b>	<b>&lt;0.01</b>	<b>0.01</b>	<b>&lt;0.05</b>	0.01	0.06
Categories										
0–1 Pathogens	<b>Ref.</b>		Ref.		Ref.		Ref.		Ref.	
2 Pathogens	<b>0.01</b>	<b>&lt;0.05</b>	<b>0.02</b>	<b>&lt;0.05</b>	0.01	0.07	0.01	0.16	0.01	0.28
3 Pathogens	<b>0.02</b>	<b>&lt;0.05</b>	<b>0.02</b>	<b>&lt;0.05</b>	<b>0.02</b>	<b>&lt;0.05</b>	0.01	0.06	0.01	0.15
4+ Pathogens	<b>0.08</b>	<b>&lt;0.001</b>	<b>0.08</b>	<b>&lt;0.001</b>	<b>0.07</b>	<b>&lt;0.001</b>	<b>0.06</b>	<b>&lt;0.001</b>	<b>0.06</b>	<b>&lt;0.001</b>
Latent class analysis										
Low burden										
'Common pathogens'	<b>0.02</b>	<b>&lt;0.01</b>	<b>0.02</b>	<b>&lt;0.01</b>	<b>0.01</b>	<b>&lt;0.05</b>	0.01	0.110	0.01	0.18
'High burden' class	<b>0.04</b>	<b>&lt;0.001</b>	<b>0.03</b>	<b>&lt;0.001</b>	<b>0.03</b>	<b>&lt;0.01</b>	<b>0.02</b>	<b>&lt;0.01</b>	<b>0.02</b>	<b>&lt;0.05</b>

Coeff, coefficient; P, P-value.

Results are based on linear regression models using complex survey weights.

Significant results at  $\alpha = 0.05$  are bolded.

chronic reactivation of these infections across the life course can lead to a host of chronic diseases. Much of the seminal research in this area focused on understanding the link between persistent infections and cardiovascular disease, specifically atherosclerosis [7, 26–29]. Studies of CMV specifically have found links with not only cardiovascular mortality [30] but incident depression [31] and frailty in older adults [32]. Current work has extended this body of research to include understanding the underlying distribution of these infections along lines of social disadvantage [1, 12, 33, 34].

Recently research in this area has begun to examine the effects of pathogen burden on markers of biological aging. Dowd *et al.* reported that infection with several herpesvirus coinfections resulted in substantial declines in leukocyte telomere length over a 3-year period [4]. The magnitude of the association reported was suggestive of substantial accelerated biological aging resultant from pathogen burden. Previous studies using a cumulative deficits approach to examine frailty have reported consistent relationships with mortality [35, 36] and have proposed that such indices may proxy biological age [36]. While we could not quantify the direct impact of pathogen burden on other biomarkers of aging in this study, our results are consistent with the hypothesis that persistent infections are likely increasing the pace of decline, even in younger cohorts.

We also sought to compare several different methods of pathogen burden measurement using the CDI as an outcome. Capturing the impact of persistent, latent infections on the body system is complex; measurement techniques need to account for both the number of pathogens present as well as the interactions between pathogens.

Our results suggest latent class analyses may be advantageous for characterising pathogen burden. The particular utility of latent class methods is that they account for unobserved heterogeneity in the study population. In contrast to approaches driven by the distribution and covariance of different variables, latent class methods assume that individuals belong to different subpopulations [37], with differing combinations of manifest and latent traits. In this way, latent class methods can be considered person-centered.

A similar method was employed by Meier *et al.* to classify levels of cumulative immune response by examining pathogen IgG antibody levels [10]. In their study, three latent classes emerged: a low cumulative immune response class, a moderate immune response class and a high immune response class with elevated antibody levels for each of the four pathogens examined [10].

Our findings suggest that in the case of characterising pathogen burden the latent class method is advantageous for detecting and adjusting for unobserved heterogeneity in clusters of individuals with similar pathogen profiles. In this study, the latent construct may be best described as a means of sorting individuals based both on their likelihood of exposure to pathogens as well as on their susceptibility to infection once exposed to a pathogen. The high burden group may then be conceptualised as those most vulnerable to infection in the study population: more likely to be exposed to pathogens and more susceptible to infection once exposed.

The attenuation of the associations in the single pathogen associations and pathogen summary score when race/ethnicity and education were included in the model suggests the influence of a strong social component affecting the distribution of these pathogens. The results of the latent class analyses, however, were largely unchanged in models controlling for race/ethnicity and education. The underlying unmeasured social structuring of the distribution of pathogen burden may be a part of what the

latent class analysis is detecting. Future studies would benefit from a more explicit examination of socioeconomic status and other social/behavioral indicators and how they affect exposure and infection with pathogens.

Our study has several limitations. First, this is a cross-sectional analysis thus precluding any causal inferences regarding pathogen burden and the CDI. Future studies should investigate whether pathogen burden is predictive of changes in the CDI over time. Nonetheless, most of the infections of interest are transmitted in early childhood or adolescence while the CDI includes predominantly measures that emerge in later life, strengthening the likelihood that infection temporally precedes the CDI.

As discussed above, the results of this study suggest an underlying social structuring of the distribution of pathogens. We were limited by data availability on social variables in NHANES. Replication of this study with datasets that offer a more robust measurement of the social environment could shed further light on the social processes at work.

Our measurement of pathogen burden was limited by the pathogens tested in NHANES. There are likely other infections that may be important predictors of physiological dysfunction that we were not able to examine because they were not tested in NHANES. Moreover, the NHANES reports only seropositivity and not circulating antibody levels of infection. There is research to suggest that antibody levels may indicate the severity of infection [38, 39] and thus is a stronger measure of infection rather than simply seropositivity. Future studies should investigate whether antibody levels of infections are associated with early markers of dysfunction.

Finally, while the CDI is useful in detecting global dysfunction, it is not specific to a single physiological process. Follow-up studies should further interrogate specific pathophysiological pathways linking chronic infection to biological dysfunction.

In conclusion, our results suggest that pathogen burden may influence early clinical indicators of poor health as measured by the CDI. Our results are salient because we were able to detect associations between pathogen burden and CDI in a relatively young population, where the onset of the disease is rarer but represents a period in which interventions targeting pathogen burden may help offset early changes in CDI and ultimately disease risk. These findings suggest that reducing overall pathogen burden and particularly the specific pathogens that drive the CDI, may provide a target for preventing the early development of age-related physiological changes.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S095026881800153X>.

**Acknowledgements.** G.A. Noppert received support from the National Institute on Aging and the National Institute of Child Health and Human Development at the National Institutes of Health (grant numbers 5 T32-AG000029-41 and 1 T32 HD091058-01). This work was also partially supported by the Claude D. Pepper Older Americans Independence Center grant P30-AG028716. A.E. Aiello received support from the National Institute of Health Grants: P2C HD050924, R01 DK087864, R01 AG040115, T32 HD091058.

**Disclaimers.** None.

## References

1. Simanek AM, Dowd JB and Aiello AE (2008) Persistent pathogens linking socioeconomic position and cardiovascular disease in the US. *International Journal of Epidemiology* **38**, 775–787.

2. **Simanek AM *et al.*** (2011) Seropositivity to cytomegalovirus, inflammation, all-cause and cardiovascular disease-related mortality in the United States. *PLoS ONE* **6**, e16103.
3. **Tarter KD *et al.*** (2014) Persistent viral pathogens and cognitive impairment across the life course in the third national health and nutrition examination survey. *Journal of Infectious Disease* **209**, 837–844.
4. **Dowd JB *et al.*** (2017) Persistent herpesvirus infections and telomere attrition over 3 years in the Whitehall II Cohort. *Journal of Infectious Disease* **6**, e16103.
5. **Pawelec G, Goldeck D and Derhovanesian E** (2014) Inflammation, ageing and chronic disease. *Current Opinion in Immunology* **29**, 23–28.
6. **Aiello AE *et al.*** (2008) Persistent infection, inflammation, and functional impairment in older Latinos. *The Journals of Gerontology Series A Biological Sciences and Medical Sciences* **63**, 610–618.
7. **Zhu J** (2000) Effects of total pathogen burden on coronary artery disease risk and C-reactive protein levels. *American Journal of Cardiology* **85**, 140–146.
8. **Simanek A *et al.*** (2015) Unpacking the ‘black box’ of total pathogen burden: is number or type of pathogens most predictive of all-cause mortality in the United States? *Epidemiology and Infection* **143**, 2624–2634.
9. **Zajacova A, Dowd JB and Aiello AE** (2009) Socioeconomic and race/ethnic patterns in persistent infection burden among U.S. adults. *The Journals of Gerontology – Series A Biological Sciences and Medical Sciences* **64**, 272–279.
10. **Meier HC *et al.*** (2016) Early life socioeconomic position and immune response to persistent infections among elderly Latinos. *Social Science and Medicine* **166**, 77–85.
11. **Dowd JB *et al.*** (2007) Socioeconomic gradients in immune response to latent infection. *American Journal of Epidemiology* **167**, 112–120.
12. **Dowd JB, Zajacova A and Aiello A** (2009) Early origins of health disparities: burden of infection, health, and socioeconomic status in U.S. children. *Social Science and Medicine* **68**, 699–707.
13. **Katan M *et al.*** (2013) Infectious burden and cognitive function: the Northern Manhattan Study. *Neurology* **80**, 1209–1215.
14. **Howlett SE and Rockwood K** (2013) New horizons in frailty: ageing and the deficit-scaling problem. *Age and Ageing* **42**, 416–423.
15. **King KE, Fillenbaum GG and Cohen HJ** (2017) A cumulative deficit laboratory test-based frailty index: personal and neighborhood associations. *Journal of the American Geriatrics Society* **65**, 1981–1987.
16. **Rockwood K and Mitnitski A** (2007) Frailty in relation to the accumulation of deficits. *The Journals of Gerontology Series A Biological Sciences and Medical Sciences* **62**, 722–727.
17. **Cohen HJ, Harris T and Pieper CF** (2003) Coagulation and activation of inflammatory pathways in the development of functional decline and mortality in the elderly. *American Journal of Medicine* **114**, 180–187.
18. **Bray BC, Lanza ST and Tan X** (2015) Eliminating bias in classify-analyze approaches for latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal* **22**, 1–11.
19. **Barker D *et al.*** (2002) Fetal origins of adult disease: strength of effects and biological basis. *International Journal of Epidemiology* **31**, 1235–1239.
20. **Belsky DW *et al.*** (2015) Quantification of biological aging in young adults. *Proceedings of the National Academy of Sciences* **112**, E4104–E4110.
21. **Bjerke W** (2011) The impact of infectious disease on chronic disease: a review of contemporary findings. *Journal of Social, Behavioral, and Health Sciences* **5**, 45–57.
22. **O'Connor SM, Taylor CE and Hughes JM** (2006) Emerging infectious determinants of chronic diseases. *Emerging Infectious Diseases* **12**, 1051–1057.
23. **Smith J and Robinson N** (2002) Age-specific prevalence of infection with herpes simplex virus types 2 and 1: a global review. *Journal of Infectious Disease* **186**, S3–S28.
24. **McQuillan G *et al.*** (2004) Racial and ethnic differences in the seroprevalence of 6 infectious diseases in the United States: data from NHANES III, 1988–1994. *American Journal of Public Health* **94**, 1952–1958.
25. **Malaty H, El-Kasabany A and Graham D** (2002) Age at acquisition of *Helicobacter pylori* infection: a follow-up study from infancy to adulthood. *Lancet* **359**, 931–935.
26. **Epstein S *et al.*** (2000) Infection and atherosclerosis: potential roles of pathogen burden and molecular mimicry. *Arteriosclerosis, Thrombosis, and Vascular Biology* **20**, 1417–1420.
27. **Izadi M *et al.*** (2012) *Helicobacter* species in the atherosclerotic plaques of patients with coronary artery disease. *Cardiovascular Pathology* **21**, 307–311.
28. **Melnick JL *et al.*** (1994) Cytomegalovirus DNA in arterial walls of patients with atherosclerosis. *Journal of Medical Virology* **42**, 170–174.
29. **Epstein S** (2002) The multiple mechanisms by which infection may contribute to atherosclerosis development and course. *Circulation Research* **90**, 2–4.
30. **Roberts ET *et al.*** (2010) Cytomegalovirus antibody levels, inflammation, and mortality among elderly latinos over 9 years of follow-up. *American Journal of Epidemiology* **172**, 363–371.
31. **Simanek A *et al.*** (2014) Herpesviruses, inflammatory markers and incident depression in a longitudinal study of Detroit residents. *Psychoneuroendocrinology* **50**, 139–148.
32. **Wang GC *et al.*** (2010) Cytomegalovirus infection and the risk of mortality and frailty in older women: a prospective observational cohort study. *American Journal of Epidemiology* **171**, 1144–1152.
33. **Aiello AE *et al.*** (2009) Socioeconomic and psychosocial gradients in cardiovascular pathogen burden and immune response: the multi-ethnic study of atherosclerosis. *Brain, Behavior, and Immunity* **23**, 663–671.
34. **Zajacova A, Dowd J and Aiello A** (2009) Socioeconomic and race/ethnic patterns in persistent infection burden among US adults. *The Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* **64**, 272–279.
35. **Rockwood K *et al.*** (2006) Long-term risks of death and institutionalization of elderly people in relation to deficit accumulation at age 70. *Journal of American Geriatrics Society* **54**, 975–979.
36. **Mitnitski AB *et al.*** (2002) Frailty, fitness and late-life mortality in relation to chronological and biological age. *BMC Geriatrics* **2**, 1.
37. **Lubke GH and Muthén B** (2005) Investigating population heterogeneity with factor mixture models. *Psychological Methods* **10**, 21–39.
38. **Aiello AE *et al.*** (2008) Persistent infection, inflammation, and functional impairment in older Latinos. *The Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* **63A**, 610–618.
39. **Aiello AE, Haan MM, Blythe L *et al.*** (2006) The influence of latent viral infection on rate of cognitive decline over 4 years. *Journal of American Geriatrics Society* **54**, 1046–1054.