

# Unpacking the ‘black box’ of total pathogen burden: is number or type of pathogens most predictive of all-cause mortality in the United States?

A. M. SIMANEK<sup>1</sup>, J. B. DOWD<sup>2</sup>, A. ZAJACOVA<sup>3</sup> AND A. E. AIELLO<sup>4\*</sup>

<sup>1</sup>Joseph J. Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

<sup>2</sup>CUNY School of Public Health, Hunter College, City University of New York, NY, USA

<sup>3</sup>Department of Sociology, College of Arts and Sciences, University of Wyoming, Laramie, WY, USA

<sup>4</sup>Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA

Received 19 May 2014; Final revision 15 October 2014; Accepted 4 November 2014;  
first published online 18 December 2014

## SUMMARY

A ‘black box’ paradigm has prevailed in which researchers have focused on the association between the total number of pathogens for which individuals are seropositive (i.e. total pathogen burden) and various chronic diseases, while largely ignoring the role that seropositivity for specific combinations of pathogens may play in the aetiology of such outcomes and consequently mortality. We examined the association between total pathogen burden as well as specific pathogen combinations and all-cause mortality in the United States. Data were from individuals aged  $\geq 25$  years tested for cytomegalovirus (CMV), herpes simplex virus (HSV)-1, HSV-2 and *Helicobacter pylori*, with mortality follow-up to 31 December 2006 in the National Health and Nutrition Examination Survey (NHANES) III ( $N = 6522$ ). We did not observe a statistically significant graded relationship between total pathogen burden level and all-cause mortality. Furthermore, compared to those seronegative for all four pathogens, the greatest statistically significant rate of all-cause mortality was for those CMV+/HSV-2+ (hazard ratio 1.95, 95% confidence interval 1.13–3.35) adjusting for age, gender, race/ethnicity, education level, body mass index ( $\text{kg}/\text{m}^2$ ) and smoking status. Interventions targeting prevention or treatment of particular pathogens may be more effective for reducing mortality than those focused solely on reducing overall pathogen burden.

**Key words:** Cytomegalovirus, *Helicobacter pylori*, herpes simplex virus, mortality, pathogen burden.

## INTRODUCTION

Over the past few decades, persistent viral pathogens such as herpesviruses and chronic bacterial infections

such as *Helicobacter pylori* have been hypothesized to be involved in the aetiology of cardiovascular disease (CVD) as well as other chronic diseases of ageing because of the ability of such pathogens to cause damage to various tissues and organs of the body via both direct and indirect pathways [1]. For example, herpesviruses as well as *H. pylori* have been detected in tissues of the vascular [2, 3], central nervous [4], and gastrointestinal systems [5] where these pathogens may cause tissue damage, induce localized inflammation and/or directly influence cellular processes that

\* Author for correspondence: Dr A. E. Aiello, Professor of Epidemiology, Social Epidemiology Program Leader, Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, 2101C McGavran-Greenberg Hall, 135 Dauer Drive, Campus Box 7435, Chapel Hill, NC 27599-7435, USA.  
(Email: aaiello@unc.edu)

contribute to the aetiology of CVD [6], cognitive impairment [7, 8], and cancer [9, 10]. However, even in the absence of direct tissue invasion, these pathogens may cause damage to tissues and organs of the body indirectly via the process of molecular mimicry whereby antibodies targeted against such pathogens may cross-react with and attack host tissues expressing proteins homologous to those contained by the pathogen [11]. Furthermore, infection with and reactivation of persistent pathogens may trigger the release of pro-inflammatory cytokines such as interleukin-6 (IL-6) and the acute phase protein C-reactive protein (CRP) [12, 13], elevated levels of which have been shown to be predictive of CVD as well as other chronic diseases with inflammation-related aetiology [14].

Findings from studies examining the role of individual pathogens in the aetiology of outcomes such as CVD [15, 16], physical and cognitive impairment [8, 17, 18], and metabolic disorders [19, 20] have, however, been mixed. Researchers have hypothesized that more important than the contribution of any single pathogen to the aetiology of various chronic diseases may be the cumulative damage resulting from exposure to multiple pathogens over one's lifetime [11]. Indeed, in recent years numerous studies have aimed to examine whether there is a dose-response relationship between the total number of pathogens for which individuals are seropositive (i.e. total pathogen burden level) and a wide array of chronic disease outcomes [21–24]. However, despite these hypothesized relationships few studies have examined the association between total pathogen burden level and all-cause mortality and those that exist have been conducted in selected older populations such as those with coronary artery disease (CAD) and/or from specific geographical regions [25, 26]. In addition, in the majority of previous studies it has been assumed that there is a monotonic relationship between the total number of pathogens for which individuals are seropositive and mortality, irrespective of pathogen type. For this reason, a 'black box' paradigm has prevailed in which the role that seropositivity for specific combinations of pathogens may play in the aetiology of chronic disease and mortality has been largely ignored. Overall, whether a graded relationship between total pathogen burden and all-cause mortality exists and the role that seropositivity for specific combinations of pathogens may play in predicting all-cause mortality in the general US population remains unknown.

The objectives of this study are therefore: (1) to examine the association between total pathogen burden with four persistent pathogens [cytomegalovirus (CMV), herpes simplex virus (HSV)-1, HSV-2 and *H. pylori*] and all-cause mortality, (2) to examine the association between seropositivity for each specific combination of these pathogens and all-cause mortality and (3) to assess whether the total number or specific combination of pathogens for which individuals are seropositive better predict all-cause mortality in a US representative population of individuals aged  $\geq 25$  years.

## METHODS

### Study population

Data were obtained from the National Health and Nutrition Examination Survey (NHANES) III (1988–1994), a population-based, multistage stratified probability survey which collected information on the health and nutrition of the US civilian non-institutionalized population. A total of 33 994 subjects were interviewed in NHANES III. Our study sample was limited to subjects that were aged  $\geq 25$  years at the time of examination ( $N = 15\,242$ ), were tested for serostatus for CMV, HSV-1, HSV-2 and *H. pylori* ( $N = 6525$ ), and had mortality follow-up to 31 December 2006 ( $N = 6522$ ). All participants in NHANES III provided informed consent at the time of enrolment.

### Laboratory analyses

CMV-specific immunoglobulin G (IgG) was measured by a commercially available enzyme linked immunosorbent assay (ELISA) (Quest International Inc., USA) with sera values near the ELISA cut-off confirmed with a second ELISA assay (bioMérieux Inc., USA). If there was disagreement between the results from the first two tests, an Immunofluorescence Assay (Bion International Inc., USA) was performed and results from this test were considered the final result [27]. HSV-1 and HSV-2 seropositivity was assessed by solid-phase enzymatic immunodot assays using purified glycoprotein gG-1 of HSV-1 and gG-2 of HSV-2 as the antigen [28]. Subjects aged  $\geq 20$  years from phase 1 of NHANES III (1988–1991) were tested for *H. pylori* antibody using *H. pylori* IgG ELISA (Wampole Laboratories, USA) [29].

## Measures

Total pathogen burden was constructed as a five-level categorical variable in which individuals were categorized according to the total number of the four pathogens (CMV, HSV-1, HSV-1 and/or *H. pylori*) for which they were seropositive (range 0–4, referent category 0). In addition, a 16-level categorical variable was constructed in which individuals were classified according to which of the 16 possible combinations of pathogens they were seropositive for: CMV+, HSV-1+, HSV-2+, *H. pylori*+, CMV+/HSV-1+, CMV+/HSV-2+, CMV+/*H. pylori*+, HSV-1+/HSV-2+, HSV-1+/*H. pylori*+, HSV-2+/*H. pylori*+, CMV+/*H. pylori*+/HSV-2+, CMV+/*H. pylori*+/HSV-1+, CMV+/*H. pylori*+/HSV-2+/*H. pylori*+, HSV-1+/*H. pylori*+/HSV-2+/*H. pylori*+, CMV+/*H. pylori*+/HSV-1+/*H. pylori*+/HSV-2+/*H. pylori* or seronegative to all four pathogens (referent category). Mortality status as of 31 December 2006, of participants was obtained by NCHS primarily from the National Death Index; however, other sources of mortality status included indication of deceased status from the Social Security Administration, the Centers for Medicare and Medicaid Services, or death certificate review [30]. Cause of death was coded using the International Classification of Diseases (ICD), Ninth Revision up to 1998 and ICD, Tenth Revision for 1999–2006. All deaths before 1999 were re-coded by the NCHS into comparable ICD, Tenth Revision codes [30]. Follow-up time for each person was calculated as the difference in months between the NHANES III examination date and the last known date alive or censored [30]. Persons who survived the entire follow-up period were administratively censored on 31 December 2006.

## Covariates

Covariates of interest included age, gender, race/ethnicity, education level, body mass index (BMI; kg/m<sup>2</sup>) and smoking status. Age in years at examination was self-reported and treated as a continuous variable. Gender was dichotomized as female and male. Race/ethnicity was self-reported as non-Hispanic white, non-Hispanic black, Mexican-American, or Other and individuals were categorized as non-Hispanic white, non-Hispanic black or Other (Mexican-American or Other were combined into Other due to small sample sizes). Years of education was self-reported and treated as continuous. BMI was computed by NCHS from weight and standing

height, measured at examination, and treated as continuous. Smoking status was self-reported and categorized into never/former smoker (did not smoke  $\geq 100$  cigarettes in one's lifetime or smoked  $\geq 100$  cigarettes in one's lifetime but do not currently smoke) and current smoker (smoked  $\geq 100$  cigarettes in one's lifetime and currently smoke).

## Statistical analyses

Statistical analyses were performed using SAS v. 9.4, with SAS-callable SUDAAN, v. 10.0.1 (SAS Institute Inc., USA) [31]. All analyses used appropriate weights and adjustments for strata and clustering used in the complex study design in NHANES III. We first estimated the weighted population proportion of individuals seropositive for each combination of pathogens within each total pathogen burden level. Next, bivariate associations between total pathogen burden level, all-cause mortality and covariates of interest were assessed using Pearson's  $\chi^2$  tests of independence for difference in proportions and Cochran–Mantel–Haenszel tests for linear trend for categorical variables and independent sample *t* tests for difference in means across mortality status or analysis of variance for comparison of means across total pathogen burden levels, for continuous variables. Covariates associated with total pathogen burden level and mortality in bivariate analyses and/or hypothesized *a priori* to be relevant confounders in the association between total pathogen burden and mortality were included in the confounder-adjusted models.

Next, we used Cox proportional hazard models to estimate the crude and confounder-adjusted hazard ratio (HR) and 95% confidence interval (CI) for the association between total pathogen burden level and all-cause mortality treating seronegativity to all four pathogens as the referent category. Models were first unadjusted, then adjusted for sociodemographic characteristics (age, gender, race/ethnicity, education level) and last for clinical factors (BMI, smoking status). These analyses were repeated with models in which total pathogen burden level was treated as a continuous ordinal variable and the adjusted Wald *F* statistic was estimated to test whether there was a statistically significant linear trend for all-cause mortality with increasing total pathogen burden level. Last, we estimated the HR and 95% CI for all-cause mortality using Cox proportional hazard models in which the 16-level pathogen burden variable was included (referent category was seronegativity for all four pathogens)

to estimate the independent and joint effects of seropositivity for each combination of pathogens and all-cause mortality. In addition, we conducted sensitivity analyses in which we used PROC MULTTEST in SAS v. 9.4 to adjust for multiple comparisons between each pathogen combination and mortality, using the false discovery methods described by Benjamini & Hochberg [32].

## RESULTS

The number and population-weighted proportion of individuals seropositive for each combination of pathogens is shown in Table 1. Nearly 22% of individuals were seropositive for one pathogen; 29.5% for two pathogens; 29.1% for three pathogens; and 8.7% for all four pathogens (see Table 1). Overall, the greatest weighted proportion of the US population were CMV+/HSV-1+/H. pylori+, CMV+/HSV-1+ or HSV-1+ only (18.8%, 18.5% and 11.5%, respectively) (see Table 1). Table 2 shows population-weighted estimates of demographic and clinical characteristics of those included in our analyses and the unadjusted bivariate associations between covariates of interest and total pathogen burden level. The mean [ $\pm$ standard error (S.E.)] age of those included in our analyses was 47.7 ( $\pm$ 0.4) years, 52.4% were female and the majority were non-Hispanic white (79.6%). Individuals had a mean ( $\pm$ S.E.) of 12.3 ( $\pm$ 0.1) years of education, a mean ( $\pm$ S.E.) BMI of 26.5 ( $\pm$ 0.1) kg/m<sup>2</sup>, and ~30% were current smokers. Older age, female gender, non-Hispanic black and Other race/ethnicity, lower education level and higher BMI were statistically significantly associated with increasing total pathogen burden level (see Table 2). The population-weighted proportion of individuals who died during the mean 15.0 years of follow-up was 21% (see Table 3). Those who died were more likely to be of older age, lower education level, higher BMI, less likely to be never smokers and less likely to be categorized as Other race/ethnicity (see Table 3).

Table 4 shows the HR (95% CI) from crude- and covariate-adjusted Cox proportional hazard models examining the association between total pathogen burden level and all-cause mortality, adjusting first for age, gender, race/ethnicity, and education level (see model 2) and second additionally controlling for BMI and smoking status (see model 3). In the crude model, we observed a statistically significantly graded relationship between increasing total pathogen burden level and all-cause mortality ( $P < 0.001$  for linear

trend). However, the rate of mortality associated with each pathogen burden level compared to those seronegative for all four pathogens was attenuated and no longer statistically significant in the fully adjusted model (see Table 4). In addition, there was no overall association ( $P = 0.528$ ) nor a statistically significant linear trend ( $P = 0.733$ ) for the association between total pathogen burden level and all-cause mortality (see model 3, Table 4).

The independent and joint effects of seropositivity for each possible combination of pathogens on all-cause mortality, compared to those seronegative for all four pathogens (referent category) is shown in Table 5. In the crude model, there were statistically significant independent effects of CMV (HR 3.63, 95% CI 2.17–6.09), HSV-1 (HR 2.02, 95% CI 1.22–3.35) and *H. pylori* (HR 3.45, 95% CI 1.45–8.20) seropositivity on all-cause mortality and combined seropositivity for all four pathogens was associated with the highest rate of all-cause mortality (HR 7.85, 95% CI 5.27–11.70). In the fully adjusted model, however, the independent effects of each individual pathogen on all-cause mortality were no longer statistically significant (see model 3, Table 5). Moreover, seropositivity for all four pathogens no longer yielded the highest rate of all-cause mortality (HR 1.38, 95% CI 0.92–2.06). By contrast, after covariate adjustment, combined seropositivity for CMV and HSV-2 (HR 1.95, 95% CI 1.13–3.35) and for HSV-2 and *H. pylori* (HR 2.07, 95% CI 0.59–7.21) were associated with greater all-cause mortality rates than that yielded for all other combinations of pathogens (see model 3, Table 5), although the latter association was not statistically significant. After adjustment for multiple comparisons, the joint effect of CMV and HSV-2 on all-cause mortality was no longer statistically significant ( $P = 0.270$ ).

## DISCUSSION

To the best of our knowledge, this study is the first to examine the association between both the total number and specific combinations of pathogens for which individuals are seropositive and all-cause mortality in a large, nationally representative sample of US adults. After adjustment for age, gender, race/ethnicity, education level, BMI and smoking status, we did not observe a statistically significantly graded relationship between increased total pathogen burden level with CMV, HSV-1, HSV-2 and/or *H. pylori* and all-cause mortality. By contrast, we found that

Table 1. Population-weighted proportion of individuals seropositive for each combination of pathogens by total pathogen burden level in subjects aged  $\geq 25$  years in the National Health and Nutrition Examination Survey III (1988–1994)

	Total pathogen burden level				
	0	1	2	3	4
Pathogen combinations, <i>N</i> (%)	349 (11.0)	857 (21.7)	1706 (29.5)	2472 (29.1)	1138 (8.7)
Seronegative for all pathogens	349 (11.0)				
HSV-1+		449 (11.5)			
CMV+		277 (6.7)			
HSV-2+		72 (2.3)			
<i>H. pylori</i> +		59 (1.2)			
CMV+/HSV-1+			1087 (18.5)		
HSV-1+/ <i>H. pylori</i> +			214 (3.6)		
HSV-1+/HSV-2+			107 (2.9)		
CMV+/HSV-2+			142 (2.2)		
CMV+/ <i>H. pylori</i> +			143 (2.1)		
HSV-2+/ <i>H. pylori</i> +			13 (0.2)		
CMV+/HSV-1+/ <i>H. pylori</i> +				1693 (18.8)	
CMV+/HSV-1+/HSV-2+				592 (8.0)	
CMV+/HSV-2+/ <i>H. pylori</i> +				114 (1.1)	
HSV-1+/HSV-2+/ <i>H. pylori</i> +				73 (1.1)	
CMV+/HSV-1+/HSV-2+/ <i>H. pylori</i> +					1138 (8.7)

CMV, Cytomegalovirus; HSV-1, herpes simplex virus-1; HSV-2, herpes simplex virus-2.

compared to those seronegative for all four pathogens, individuals who were CMV+/HSV-2+ and HSV-2/*H. pylori*+ had the greatest rates of all-cause mortality. Our results suggest that future studies should think outside the ‘black box’ paradigm in which the number but not type of pathogens for which individuals are seropositive has been the focus, in order to identify the specific combinations of pathogens for which seropositivity is most predictive of chronic disease and consequently mortality in the United States.

We are only aware of two previous studies that have examined the association between total pathogen burden and all-cause mortality [25, 26]. Zhu *et al.* examined the association between total pathogen burden with CMV, hepatitis A virus, HSV-1, and/or HSV-2 and all-cause mortality in a mostly male cohort of Utah residents with existing CAD (mean age, 65.3 years) [25]. The authors identified a statistically significant trend for all-cause mortality with increasing total pathogen burden level ( $P = 0.020$ ); however, consistent with our findings, there were no statistically significant associations between each level of total pathogen burden and all-cause mortality in models adjusting for age, gender, and various CVD risk factors [25]. Importantly, Zhu *et al.* did not control for socioeconomic status and the study population was

older, from a single region of the United States and limited to individuals with a history of CAD [25]. For these reasons, it is possible that the statistically significant linear trend observed by Zhu *et al.* may have resulted from residual confounding by socioeconomic factors and/or may reflect demographic or clinical differences between those included in this previous study and the general US population [25].

In a study by Elkind *et al.*, instead of categorizing individuals according to the total number of pathogens for which they were seropositive, the authors constructed an infectious burden (IB) index based on the strength of the association between the pathogens for which individuals were seropositive and rate of stroke in 1625 individuals (mean age 68.4 years) in Northern Manhattan [26]. Specifically, the authors estimated the hazard rate of stroke associated with seropositivity for each of five pathogens (*Chlamydia pneumoniae*, CMV, HSV-1, HSV-2 and/or *H. pylori*) and then summed the  $\beta$  coefficients corresponding to the log hazard rate yielded for each pathogen for which individuals were seropositive [26]. Elkind *et al.* then estimated the HR (95% CI) for all-cause mortality per 1 s.d. (0.33) increase in the IB index, finding that after adjusting for demographic and cardiovascular risk factors, there was no

Table 2. Population-weighted demographic and clinical characteristics by total pathogen burden level in subjects aged  $\geq 25$  years assessed for mortality status in the National Health and Nutrition Examination Survey III (1988–1994)

Covariates	Total pathogen burden level						P value†
	Total (N = 6522)	0 (N = 349)	1 (N = 857)	2 (N = 1706)	3 (N = 2 472)	4 (N = 1 138)	
Age, years, mean (s.e.)	47.7 (0.4)	38.0 (0.7)	42.6 (0.4)	48.3 (0.5)	52.7 (0.9)	54.0 (1.0)	<0.001*
Gender, %							
Female (N = 3253)	52.4	41.9	45.3	55.7	55.9	60.8	<0.001*
Male (N = 3269)	47.6	58.1	54.7	44.3	44.1	39.2	
Race/ethnicity, %							
Non-Hispanic white (N = 3084)	79.6	95.8	91.8	83.8	69.7	47.8	<0.001*
Non-Hispanic black (N = 1571)	9.7	2.2	3.4	7.5	12.6	32.0	
Other (N = 1867)	10.7	2.0	4.8	8.7	17.6	20.2	
Education, years, mean (s.e.)‡	12.3 (0.1)	14.0 (0.2)	13.6 (0.2)	12.5 (0.1)	11.1 (0.2)	10.3 (0.2)	<0.001*
Body mass index, kg/m <sup>2</sup> , mean (s.e.)§	26.5 (0.1)	25.5 (0.3)	26.4 (0.2)	26.3 (0.2)	26.9 (0.2)	27.4 (0.3)	<0.001*
Smoking status, %							
Never (N = 2891)	42.1	47.9	43.1	39.4	41.4	44.0	0.064
Former (N = 1858)	28.1	26.2	28.5	29.3	28.8	22.8	
Current (N = 1773)	29.8	25.9	28.4	31.3	29.8	33.2	

s.e., Standard error.

† Two-sided P values for Cochran–Mantel–Haenszel tests for trend for categorical variables and analysis of variance for tests of linear trend for continuous variables.

‡ N = 6480 due to 42 subjects missing data on education level.

§ N = 6499 due to 23 subjects missing data on body mass index.

\* P < 0.05.

Table 3. Population-weighted demographic and clinical characteristics by mortality status in individuals aged  $\geq 25$  years tested for pathogens in the National Health and Nutrition Examination Survey III (1988–1994)

Covariates	Mortality status, <i>N</i> (%)		<i>P</i> value†
	Survived 4553 (79.0)	Died 1969 (21.0)	
Age, years, mean (S.E.)	42.9 (0.3)	65.8 (0.7)	<0.001*
Gender, %			
Female ( <i>N</i> = 3253)	52.7	51.3	0.381
Male ( <i>N</i> = 3269)	47.3	48.7	
Race/ethnicity, %			
Non-Hispanic white ( <i>N</i> = 3084)	78.2	84.8	0.002*
Non-Hispanic black ( <i>N</i> = 1571)	9.6	9.9	
Other ( <i>N</i> = 1867)	12.2	5.3	
Education, years, mean (S.E.)‡	12.6 (0.1)	11.0 (0.2)	<0.001*
Body mass index, kg/m <sup>2</sup> , mean (S.E.)§	26.4 (0.2)	27.0 (0.2)	0.044*
Smoking status, %			
Never ( <i>N</i> = 2891)	44.0	35.1	0.032*
Former ( <i>N</i> = 1858)	25.7	36.9	
Current ( <i>N</i> = 1773)	30.3	27.9	

S.E., Standard error.

† Two-sided *P* values for Pearson's  $\chi^2$  tests for difference in proportions or Cochran–Mantel–Haenszel tests for trend for categorical variables and *t* tests for difference in means for continuous variables.

‡ *N* = 6480 due to 42 subjects missing data on education level.

§ *N* = 6499 due to 23 subjects missing data on body mass index.

\* *P* < 0.05.

statistically significant association between increasing IB index and all-cause mortality (HR 1.12, 95% CI 0.99–1.26) [26].

The results from the study by Elkind *et al.* suggest, similarly to our findings, that there is no statistically significant linear association between increasing total pathogen burden and all-cause mortality [26]. Importantly, however, because a 1 s.d. increase in the IB index is not necessarily equivalent to seropositivity for an additional pathogen, it is difficult to directly compare the results from Elkind *et al.* [26] to those observed in our study as well as those from the study by Zhu *et al.* [25]. Moreover, given that the IB index constructed by Elkind *et al.* was based on the strength of the association between seropositivity for individual pathogens and rate of stroke, the associations observed by the authors may more closely reflect the rate of death associated specifically with increased risk of infection-related stroke than between infectious burden and mortality, *per se* [26]. Last, although Elkind *et al.* constructed the IB index based upon the pathogens for which individuals were seropositive, because the authors did not report the rates of all-cause mortality associated with specific IB index scores, whether seropositivity for particular

combinations of pathogens were most predictive of mortality in this previous study, cannot be ascertained [26].

When we examined the independent and joint effects of each combination of pathogens on all-cause mortality, compared to those seronegative for all four pathogens, we found that there was not a strictly monotonic relationship between increasing total pathogen burden level and all-cause mortality, regardless of pathogen combination. For example, in the fully adjusted model, the combined effects of seropositivity for CMV and HSV-2 as well as HSV-2 and *H. pylori* (although not statistically significant) yielded higher rates of all-cause mortality than that yielded for all combinations of three pathogens as well as for seropositivity for all four pathogens. While CMV, HSV-2 and *H. pylori* have all been linked to numerous chronic disease outcomes including CVD [15, 33, 34] cancer [9, 10, 35] and cognitive impairment [8, 36], there is a paucity of studies examining the biological interaction between these pathogens. Indeed, we are unaware of any studies which have examined the interactive effects of HSV-2 and *H. pylori* in the aetiology of chronic disease outcomes or mortality. *In vitro* studies suggest, however, that

Table 4. Association between total pathogen burden level and all-cause mortality in subjects aged  $\geq 25$  years in the National Health and Nutrition Examination Survey III (1988–1994)

	Hazard ratio (95% confidence interval)					
	Model 1†	<i>P</i> value	Model 2‡	<i>P</i> value	Model 3§	<i>P</i> value
Total pathogen burden level		<0.001*		0.260		0.528
0	1.0	<0.001*	1.00	0.185	1.00	0.733
1	2.48 (1.58–3.89)*		1.35 (0.92–1.99)		1.28 (0.88–1.88)	
2	4.53 (2.90–7.08)*		1.55 (1.05–2.29)*		1.41 (0.96–2.09)	
3	6.09 (4.11–9.01)*		1.42 (1.03–1.96)*		1.24 (0.90–1.71)	
4	7.85 (5.27–11.69)*		1.55 (1.04–2.33)*		1.37 (0.92–2.04)	
Age, years	—		1.09 (1.08–1.10)*		1.10 (1.09–1.11)*	
Gender						
Female	—		1.0		1.0	
Male	—		1.47 (1.32–1.64)*		1.37 (1.24–1.51)*	
Race/ethnicity						
Non-Hispanic white	—		1.0		1.0	
Non-Hispanic black	—		1.24 (1.07–1.44)*		1.22 (1.05–1.41)*	
Other	—		0.68 (0.56–0.84)*		0.72 (0.60–0.87)*	
Educational level, years	—		0.96 (0.94–0.98)*		0.96 (0.94–0.98)*	
Body mass index, kg/m <sup>2</sup>	—		—		1.01 (1.00–1.03)	
Smoking status						
Never	—		—		1.0	
Former	—		—		1.29 (1.11–1.49)*	
Current	—		—		2.22 (1.87–2.64)*	

† Model 1 unadjusted.

‡ Model 2 adjusted for age, gender, race/ethnicity, and education level ( $N = 6480$ ).

§ Model 3 additionally adjusted for body mass index and smoking status ( $N = 6457$ ).

|| Two-tailed *P* value for test for linear trend estimated from model with total pathogen burden level treated as a continuous ordinal variable.

\*  $P < 0.05$ .

HSV-2 replication may be induced by CMV [37, 38], thus it is possible that HSV-2 and CMV may indeed biologically interact to increase rates of chronic disease and thus mortality, beyond that of other pathogens included in our study. For this reason, although the combined effect of CMV and HSV-2 seropositivity was no longer statistically significant after adjustment for multiple comparisons, further examination of the effect of seropositivity for this combination of pathogens on mortality may be warranted. Our findings also suggest that there may be negative biological interaction between and/or competing antagonism among pathogens acting through similar biological pathways such that there will not always be a strictly monotonic relationship between increasing total pathogen burden level and mortality. Importantly, violations of this assumption will go undetected in models which only consider the total number and not types of pathogens for which individuals are seropositive. For these reasons, future studies should consider thinking outside the total

pathogen burden ‘black box’ to also consider the role that seropositivity for and/or biological interactions between specific combinations of pathogens may play in the aetiology of various chronic diseases and consequently mortality.

There are a few limitations to our study. First, we were unable to include other persistent pathogens that have also been associated with chronic diseases such as CVD as well as mortality including *C. pneumoniae* [39] in our measure of total pathogen burden, because only selected pathogens were serologically tested for in NHANES III. For this reason, we cannot rule out that the associations between specific combinations of pathogens and all-cause mortality observed in our study are due to unmeasured co-infection with other persistent pathogens not measured in our study. Second, the number of individuals seropositive for certain combinations of pathogens was small (i.e.  $N = 13$  for those seropositive for HSV-2+ and *H. pylori*+ combined), thus power to detect associations between all combinations of pathogens and mortality may

Table 5. Association between each pathogen combination and all-cause mortality in subjects aged  $\geq 25$  years in the National Health and Nutrition Examination Survey III (1988–1994)

	Hazard ratio (95% confidence interval)		
	Model 1†	Model 2‡	Model 3§
Pathogen combinations			
Seronegative for all pathogens	1.0	1.0	1.0
CMV+	3.63 (2.17–6.09)*	1.57 (0.98–2.52)	1.50 (0.91–2.45)
HSV-1+	2.02 (1.22–3.35)*	1.12 (0.72–1.76)	1.07 (0.70–1.65)
HSV-2+	1.05 (0.49–2.23)	1.37 (0.77–2.42)	1.27 (0.73–2.22)
<i>H. pylori</i> +	3.45 (1.45–8.20)*	1.92 (0.85–4.34)	1.81 (0.85–3.86)
CMV+/HSV-1+	4.89 (3.14–7.62)*	1.45 (0.99–2.14)	1.34 (0.92–1.95)
CMV+/HSV-2+	3.91 (2.06–7.43)*	2.10 (1.21–3.65)*	1.95 (1.13–3.35)*
CMV+/ <i>H. pylori</i> +	4.73 (2.94–7.62)*	1.42 (0.83–2.42)	1.22 (0.71–2.08)
HSV-1+/HSV-2+	1.77 (0.85–3.68)*	1.45 (0.71–2.95)	1.40 (0.68–2.91)
HSV-1+/ <i>H. pylori</i> +	5.49 (2.84–10.61)*	2.01 (1.10–3.67)*	1.75 (0.95–3.22)
HSV-2+/ <i>H. pylori</i> +	2.72 (0.56–13.15)	2.08 (0.60–7.19)	2.07 (0.59–7.21)
CMV+/HSV-1+/HSV-2+	4.90 (2.93–8.19)*	1.66 (1.11–2.47)*	1.44 (0.98–2.12)
CMV+/HSV-1+/ <i>H. pylori</i> +	6.90 (4.77–9.99)*	1.34 (0.97–1.85)	1.18 (0.86–1.62)
CMV+/HSV-2+/ <i>H. pylori</i> +	4.59 (2.09–10.07)*	1.73 (1.00–3.00)*	1.34 (0.72–2.51)
HSV-1+/HSV-2+/ <i>H. pylori</i> +	3.23 (1.26–8.25)*	1.31 (0.56–3.04)	1.11 (0.46–2.66)
CMV+/HSV-1+/HSV-2+/ <i>H. pylori</i> +	7.85 (5.27–11.70)*	1.56 (1.04–2.35)*	1.38 (0.92–2.06)
Age, years	—	1.09 (1.09–1.10)*	1.10 (1.09–1.11)*
Gender			
Female	—	1.0	1.0
Male	—	1.48 (1.32–1.66)*	1.38 (1.24–1.53)*
Race/ethnicity			
Non-Hispanic white	—	1.0	1.0
Non-Hispanic black	—	1.18 (1.01–1.39)	1.16 (1.00–1.36)**
Other	—	0.67 (0.54–0.81)*	0.70 (0.58–0.85)*
Education level, years	—	0.96 (0.94–0.98)*	0.96 (0.94–0.98)*
Body mass index, kg/m <sup>2</sup>	—	—	1.01 (1.00–1.03)
Smoking status			
Never	—	—	1.0
Former	—	—	1.28 (1.10–1.49)*
Current	—	—	2.22 (1.87–2.63)*

CMV, Cytomegalovirus; HSV-1, herpes simplex virus-1; HSV-2, herpes simplex virus-2.

† Model 1 unadjusted ( $N = 6522$ ).

‡ Model 2 adjusted for age, gender, race/ethnicity and education level ( $N = 6480$ ).

§ Model 3 additionally adjusted for body mass index and smoking status ( $N = 6457$ ).

\*  $P < 0.05$ .

have been limited. For these reasons, future studies which include a comprehensive array of pathogens, and possibly even microbiome data, are needed to fully understand which pathogens, alone or in combination, are most strongly associated with all-cause mortality in the general US population. Nonetheless, NHANES III provided us with the largest and only population-based study available for addressing questions related to total pathogen burden and all-cause mortality in the general US population.

Overall, our findings suggest that interventions targeting the prevention or treatment of particular

combinations of persistent pathogens may be more effective for reducing all-cause mortality than those focused solely on reducing overall pathogen burden. For this reason, further investigations into the total pathogen burden 'black box' in which researchers focus on the number but not type of pathogens for which individuals are seropositive, may ultimately misinform preventive measures aimed at reducing infection-related death. Future studies should therefore aim to examine the role of specific combinations of pathogens in the aetiology of chronic disease and to identify which pathogens, in addition to those identified in our

study, individually or in combination, are most predictive of mortality in the general US population.

## ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health (1-R21-NR-011181-01) (to J.B.D. and A.E.A.) and National Institutes of Aging, and National Institutes of Health (1-R01-AG-040115) (to J.B.D. and A.E.A.).

## DECLARATION OF INTEREST

None.

## REFERENCES

- O'Connor SM, Taylor CE, Hughes JM. Emerging infectious determinants of chronic diseases. *Emerging Infectious Diseases* 2006; **12**: 1051–1057.
- Izadi M, et al. Helicobacter species in the atherosclerotic plaques of patients with coronary artery disease. *Cardiovascular Pathology* 2012; **21**: 307–311.
- Melnick JL, et al. Cytomegalovirus DNA in arterial walls of patients with atherosclerosis. *Journal of Medical Virology* 1994; **42**: 170–174.
- Fraser NW, et al. Herpes simplex type 1 DNA in human brain tissue. *Proceedings of the National Academy of Sciences USA* 1981; **78**: 6461–6465.
- Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* 1984; **1**: 1311–1315.
- Epstein SE. The multiple mechanisms by which infection may contribute to atherosclerosis development and course. *Circulation Research* 2002; **90**: 2–4.
- Kountouras J, et al. Association between Helicobacter pylori infection and mild cognitive impairment. *European Journal of Neurology* 2007; **14**: 976–982.
- Tarter KD, et al. Persistent viral pathogens and cognitive impairment across the lifecourse in the Third National Health and Nutrition Examination Survey. *Journal of Infectious Diseases* 2014; **209**: 837–844.
- Ferreira AC, et al. Helicobacter and gastric malignancies. *Helicobacter* 2008; **13** (Suppl. 1): 28–34.
- Soroceanu L, Cobbs CS. Is HCMV a tumor promoter? *Virus Research* 2011; **157**: 193–203.
- Epstein SE, et al. Infection and atherosclerosis: potential roles of pathogen burden and molecular mimicry. *Arteriosclerosis, Thrombosis, and Vascular Biology* 2000; **20**: 1417–1420.
- Nazmi A, et al. The influence of persistent pathogens on circulating levels of inflammatory markers: a cross-sectional analysis from the Multi-Ethnic Study of Atherosclerosis. *BMC Public Health* 2010; **10**: 706.
- Jackson L, et al. A population-based epidemiologic study of Helicobacter pylori infection and its association with systemic inflammation. *Helicobacter* 2009; **14**: 108–113.
- Emerging Risk Factors Collaboration C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: an individual participant meta-analysis. *Lancet* 2010; **375**: 132–140.
- Simanek AM, Dowd JB, Aiello AE. Persistent pathogens linking socioeconomic position and cardiovascular disease in the US. *International Journal of Epidemiology* 2009; **38**: 775–787.
- Sealy-Jefferson S, et al. Antibody levels to persistent pathogens and incident stroke in Mexican Americans. *PLoS ONE* 2013; **8**: e65959.
- Aiello AE, et al. The influence of latent viral infection on rate of cognitive decline over 4 years. *Journal of the American Geriatric Society* 2006; **54**: 1046–1054.
- Aiello AE, et al. Persistent infection, inflammation, and functional impairment in older latinos. *Journals of Gerontology, Series A: Biological Sciences and Medical Sciences* 2008; **63**: 610–618.
- Jeon CY, et al. Helicobacter pylori infection is associated with an increased rate of diabetes. *Diabetes Care* 2012; **35**: 520–525.
- Xia HH, et al. Helicobacter pylori infection is not associated with diabetes mellitus, nor with upper gastrointestinal symptoms in diabetes mellitus. *American Journal of Gastroenterology* 2001; **96**: 1039–1046.
- Katan M, et al. Infectious burden and cognitive function: the Northern Manhattan Study. *Neurology* 2013; **80**: 1209–1215.
- Miller GE, et al. Relation of depressive symptoms to C-reactive protein and pathogen burden (cytomegalovirus, herpes simplex virus, Epstein-Barr virus) in patients with earlier acute coronary syndromes. *American Journal of Cardiology* 2005; **95**: 317–321.
- Fernandez-Real J-M, et al. Burden of Infection and insulin resistance in healthy middle-aged men. *Diabetes Care* 2006; **29**: 1058–1064.
- Zhu J, et al. Effects of total pathogen burden on coronary artery disease risk and C-reactive protein levels. *American Journal of Cardiology* 2000; **85**: 140–146.
- Zhu J, et al. Prospective study of pathogen burden and risk of myocardial infarction or death. *Circulation* 2001; **103**: 45–51.
- Elkind MS, et al. Infectious burden and risk of stroke: the northern Manhattan study. *Archives of Neurology* 2010; **67**: 33–38.
- Staras SA, et al. Seroprevalence of cytomegalovirus infection in the United States, 1988–1994. *Clinical Infectious Diseases* 2006; **43**: 1143–1151.
- Schillinger JA, et al. National seroprevalence and trends in herpes simplex virus type 1 in the United States, 1976–1994. *Sexually Transmitted Diseases* 2004; **31**: 753–760.
- U.S. Department of Health and Human Services (DHHS). National Center for Health Statistics. Third National Health and Nutrition Examination Survey (NHANES III: 1988–1994. NHANES III Second Laboratory File Documentation Series 11, No. 2A.

- Hyattsville, Maryland: Centers for Disease Control and Prevention, 1998.
30. **Data Linkage Team.** Comparative analysis of the NHANES III public-use and restricted-use linked mortality files: 2010 public-use data release. National Center for Health Statistics. March 2010. Hyattsville, Maryland ([http://www.cdc.gov/nchs/data/datalinkage/nhanes3\\_mort\\_compare\\_2010\\_final.pdf](http://www.cdc.gov/nchs/data/datalinkage/nhanes3_mort_compare_2010_final.pdf)).
  31. **SUDAAN.** Example Manual, Release 10-0-1. Research Triangle Park, NC.: Research Triangle Institute, 2010.
  32. **Benjamini Y, Hochberg Y.** Controlling the false discovery rate – a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B: Methodological* 1995; **57**: 289–300.
  33. **Sun Y, et al.** Herpes simplex virus type 2 infection is a risk factor for hypertension. *Hypertension Research* 2004; **27**: 541–544.
  34. **de Luis DA, et al.** Association of Helicobacter pylori infection with cardiovascular and cerebrovascular disease in diabetic patients. *Diabetes Care* 1998; **21**: 1129–1132.
  35. **Thomas F, et al.** Herpes simplex virus type 2 and cancer: a medical geography approach. *Infection, Genetics and Evolution* 2011; **11**: 1239–1242.
  36. **Beydoun MA, et al.** Helicobacter pylori seropositivity and cognitive performance among US adults: evidence from a large national survey. *Psychosomatic Medicine* 2013; **75**: 486–496.
  37. **Colberg-Poley AM, Isom HC, Rapp F.** Reactivation of herpes simplex virus type 2 from a quiescent state by human cytomegalovirus. *Proceedings of the National Academy of Sciences USA* 1979; **76**: 5948–5951.
  38. **Colberg-Poley AM, Isom HC, Rapp F.** Involvement of an early human cytomegalovirus function in reactivation of quiescent herpes simplex virus type 2. *Journal of Virology* 1981; **37**: 1051–1059.
  39. **Strachan DP, et al.** Relation of Chlamydia pneumoniae serology to mortality and incidence of ischaemic heart disease over 13 years in the caerphilly prospective heart disease study. *British Medical Journal (Clinical Research Edition)* 1999; **318**: 1035–1039.