**Supplemental Materials**

**Supplemental Material:** **Measurement of Pathogens**

For both the 2003-2004 and 2009-2010 waves seropositivity for herpes simplex virus type 1 (HSV-1) and herpes simplex virus type 2 (HSV-2) was determined based on immunodot assay using the glycoprotein gG-1 antigen. For both the 2003-2004 and 2009-2010 waves, the assessment of human papillomavirus (HPV) was based on multiplex luminex assays of antibodies to neutralizing epitopes for HPV types 6, 11, 16, and 18 L1-Virus-Like Particles in blood serum. In 2009-2010 oral HPV was assessed based on purified DNA extracted from the oral rinse specimens. Each specimen was analyzed for 37 types of HPV through a polymerase-chain-reaction (PCR) assay targeted at the viral genome. Participants in the 2003-2004 wave were classified as being seropositive for any of the 4 sub-types (=1) versus not being seropositive for any sub-type (=0). Participants in the 2009-2010 wave were classified as being positive for any HPV, either blood serum or oral (=1) or not positive for any HPV (=0). For the 2003-2004 wave, toxoplasmosis was determined based on the presence and quantity of IgG antibodies to *Toxoplasma gondii* by performing an enzyme immunoassay with *toxoplasma* antigen. The data were released as International Units (IU). Values ≥ 10 IU were coded as positive, values between 0 and 9 IU were coded as negative. For the 2009-2010 wave, toxoplasmosis was determined based on the presence of toxoplasma gondii antibodies measured with Toxoplasma IgG enzyme immunoassay. Optical density units were determined and normalized as International Units (IU/mL). Samples with results ≥ 33 IU/mL were coded as positive, samples with results < 27 IU/mL were coded as negative, and samples with results ≥ 27 IU/mL and < 33 IU/mL were considered equivocal. Infection with the Human Immunodeficiency Virus (HIV) was determined based on blood or urine antibodies to human immunodeficiciecy virus type 1 or type 2 or both (type 1 only in urine). For 2009-2010, only blood was tested. For the 2003-2004 wave, cytomegalovirus (CMV) positivity was determined based on blood serum CMV specific IgG antibodies measured with an ELISA. Optical density units were reported and classified as positive, negative, or equivocal. For the 2003-2004 wave, syphilis infection was determined based on an enzyme immunoassay (EIA) to detect IgG antibodies to *Treponema pallidum.* Results where then classified as positive, negative, or equivocal. For all analyses of pathogens, equivocal results were then coded as missing. Notably, while HIV has a very different natural history than the other infections with more substantial biological consequences, we included it in the CDI because of both its chronic nature and its detrimental effect across body systems.

**Supplementary Table 1a.** Correlation matrix depicting the tetrachoric correlations between each of the pathogen variables in NHANES Wave 1.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **CMV**  | **HSV-1** | **HSV-2** | **HPV** | **Syphilis**  | **Toxoplasmosis**  | **HIV**  |
| **CMV**  |  | 1 | 0.44\*\*\* | 0.33\*\*\* | 0.24\*\*\* | 0.36\*\*\* | 0.24\*\*\* | 0.99\*\*\*\* |
| **HSV-1** |  | 0.44\*\*\* | 1 | 0.04 | 0.13\*\*\* | 0.20\* | 0.31\*\*\* | 0.22 |
| **HSV-2**  |  | 0.33\*\*\* | 0.04 | 1 | 0.47\*\*\* | 0.36\*\*\* | -0.03 | 0.58\*\*\* |
| **HPV**  |  | 0.24\*\*\* | 0.13\*\*\* | 0.47\*\*\* | 1 | 0.1 | -0.07 | 0.40\*\*\* |
| **Syphilis**  |  | 0.36\*\*\* | 0.20\* | 0.36\*\*\* | 0.1 | 1 | 0.16 | 0.36\* |
| **Toxoplasmosis**  | 0.24\*\*\* | 0.31\*\*\* | -0.03 | -0.07 | 0.16 | 1 | 0.03 |
| **HIV**  |   | 0.99\*\*\*\* | 0.22 | 0.58\*\*\* | 0.40\*\*\* | 0.36\* | 0.03 | 1 |
| \*<0.05, \*\*<0.01, \*\*\*<0.001, \*\*\*\*perfect correlation |  |  |  |

**Supplementary Table 1b.** Correlation matrix depicting the tetrachoric correlations between each of the pathogen variables in NHANES Wave 2.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **HSV-1** | **HSV-2** | **HPV** | **Toxoplasmosis**  | **HIV**  |
| **HSV-1** |  | 1 | 0.04 | 0.12\*\*\* | 0.33\*\*\* | 0.07 |
| **HSV-2**  |  | 0.04 | 1 | 0.45\*\*\* | 0.12\*\* | 0.66\*\*\* |
| **HPV**  |  | 0.12\*\*\* | 0.45\*\*\* | 1 | -0.03 | 0.24\* |
| **Toxoplasmosis**  | 0.33\*\*\* | 0.12\*\* | -0.03 | 1 | 0.17 |
| **HIV**  |   | 0.07 | 0.66\*\*\* | 0.24\* | 0.17 | 1 |
| \*<0.05, \*\*<0.01, \*\*\*<0.001, \*\*\*\*perfect correlation |  |  |  |
|  |  |  |

**Supplementary Table 2a.** Latent class model fit statistics comparing differing class solutions for the National Health and Nutrition Examination Survey, 2003-2004 wave (N= 2,168).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No. of classes | G2 | df | AIC | BIC | Log-Likelihood | Adjusted BIC | Entropy |
|  |  |  |  |  |  |  |  |  |
| 2 |  | 243.32 | 112 | 273.32 | 358.55 | -6214.04 | 310.89 | 0.48 |
| **3** |  | **60.34** | **104** | **106.34** | **237.01** | **-6122.55** | **163.94** | **0.56** |
| 4 |  | 43.34 | 96 | 105.34 | 281.47 | -6114.05 | 182.98 | 0.54 |
| 5 |  | 33.39 | 88 | 111.39 | 332.97 | -6109.08 | 209.07 | 0.53 |
| 6 |  | 26.78 | 80 | 120.78 | 387.81 | -6105.77 | 238.49 | 0.55 |
| G2 refers to the likelihood ratio value; df= degrees of freedom; AIC = Akaike’s Information Criteria; BIC = Bayesian Information Criterion |

**Supplementary Table 2b.** Latent class model fit statistics comparing differing class solutions for the National Health and Examination Survey, 2009-2010 wave (N= 2,546).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No. of classes | G2 | df | AIC | BIC | Log-Likelihood | Adjusted BIC | Entropy |
|  |  |  |  |  |  |  |  |  |
| 2 |  | 99.53 | 20 | 121.53 | 185.79 | -5721.42 | 150.84 | 0.79 |
| **3** |  | **21.81** | **14** | **55.81** | **155.13** | **-5682.57** | **101.12** | **0.79** |
| 4 |  | 6.95 | 8 | 52.95 | 187.32 | -5675.14 | 114.25 | 0.53 |
| 5 |  | 5.08 | 2 | 63.08 | 232.51 | -5674.2 | 140.37 | 0.57 |
| G2 refers to the likelihood ratio value; df= degrees of freedom; AIC = Akaike’s Information Criteria; BIC = Bayesian Information Criterion\*note: a 6 class model did not converge  |

**Supplementary Table 3.** Sensitivity analyses comparing a pathogen summary score with and without HPV and syphilis in NHANES wave 1, 2001-2002.

|  |  |  |
| --- | --- | --- |
|  | Model 1: no controls | Model 2: full controls\* |
|  | *Coeff* | *P* |  | *Coeff* | *P* |
| Pathogen summary score (CMV, HSV-1, HSV-2, HPV, Syphilis, Toxoplasmosis, and HIV) | 0.010 | 0.01 |  | 0.003 | 0.11 |
|  |  |  |  |  |  |
| Pathogen summary score (CMV, HSV-1, HSV-2, Toxoplasmosis, and HIV)  | 0.012 | <0.0001 |  | 0.006 | 0.03 |
| \*Model 2 includes age, sex, BMI, smoking status, race/ethnicity, and education  |

**Supplementary Figure 1.** Description of the composition of the study sample for the National Health and Nutrition Examination Survey, 2003-2004. These same exclusions were applied to the 2009-2010 wave.

All participants aged 20-49 years in the 2003-2004 wave of the NHANES: 14,623

Excluded due to missing data on one of the pathogens tested: 6,934

Final sample size: 2,168
including only those individuals with complete covariate information (age, sex, BMI, smoking, race/ethnicity, and education).

Individuals with complete pathogen information: 2,709

Excluded cases due to missing complete covariate information: 541

Excluded due to missing either the survey, physical examination, and/or laboratory components: 4,980

Individuals who participated in the survey, physical examination, and laboratory studies: 9,643

**Supplementary Figure 2.** Descriptive statistics and histogram of the cumulative deficits index for the National Health and Nutrition Examination Survey, A) 2003-2004 wave (N= 2,168) and B) 2009-2010 wave (N= 2,546).

Panel A

Range: 0.04-0.78

Mean: 0.30

Median: 0.28

Q1: 0.19

Q3: 0.38

Panel B

Range: 0-0.74

Mean: 0.27

Median: 0.25

Q1: 0.16

Q3: 0.33