

Supplementary Online Content

Leventhal AM, Bello MS, Galstyan E, Higgins ST, Barrington-Trimis JL. Association of cumulative socioeconomic and health-related disadvantage with disparities in smoking prevalence in the United States, 2008 to 2017. *JAMA Intern Med*. Published online April 22, 2019.
doi:10.1001/jamainternmed.2019.0192

eMethods. Supplemental Methods

eResults. Supplemental Results

eFigure 1. Percentage of Current Smokers Among Females by Heavy Drinking Status, 2008-2013

eFigure 2. Proportion of Current Smokers in Each Survey Year, by Each Form of Disadvantage, National Health Interview Survey, 2008-2017

eTable. Pseudo R² Values From Regression Models With Alternate Specifications of Cumulative Disadvantage

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods. Supplemental Methods

Description of smoking survey measure

Smoking Status. Have you smoked at least 100 cigarettes in your ENTIRE LIFE? (Yes; No; Refused; Not Ascertained; Don't know). If respondents answered "Yes" they were asked Do you NOW smoke cigarettes every day, some days, or not at all? (Every day; Some days; Not at all; Refused; Not ascertained; Don't know).

Respondents who refused to answer this question, whose data were not ascertained, or who reported "don't know" were not included in the analytic sample (N=1,511).

Description of disadvantage survey measures

Unemployment. Did you have a job or business at any time in the PAST 12 MONTHS? (Had job last week; No job last week, had job past 12 months; No job last week, no job past 12 months; Never worked; Refused; Not ascertained; Don't know). What is the main reason you did not work last week? Or, what is the main reason you did not have a job or business last week? (Response options included: Taking care of house or family; Going to school; Retired; On a planned vacation from work; On family or maternity leave; Temporarily unable to work for health reasons; Have job or contract and off-season; On layoff; Disabled; Other; Refused; Not ascertained; Don't know). Participants were considered unemployed if they reported not working in the past week or past year due to Temporarily unable to work for health reasons; Have job or contract and off-season; On layoff; Disabled. *N* missing data = 7,823 (2.8%).

Poverty. Poverty status for a given year was based on family size and household annual income relative to the preceding year's federal poverty line. *N* missing data = 20,292 (7.3%).

Low education. What is the highest level of school person has completed or highest degree person has received? (Never attended / kindergarten only; 1st – 11th grade [independent answers]; 12th grade, no diploma; GED or equivalent; High School Graduate; Some college, no degree; Associate degree: occupational, technical, or vocational program; Associate degree: academic program; Bachelor's degree [Example: BA, AB, MS, BBA]; Master's degree [Example: MA, MS, MEng, MEd, MBA]; Professional School degree [Example: MD, DDS, DVM, JD]; Doctoral degree [Example: PhD, EdD]; Child under 5, Refused, Not ascertained, Don't know). *N* missing data = 1,298 (0.5%).

Disability/Limitation. What condition or health problem causes you to have difficulty with walking a quarter mile; walking up 10 steps without resting; standing/being on feet for about 2 hours; sitting for about 2 hours; stooping/bending/kneeling; reaching up over head; using fingers to grasp/handle small objects; lifting/carrying 10 pounds; pushing/pulling large objects; going out for things (shopping/movies); participating in social activities; or relaxing at home (reading/sewing): 1) vision/problem seeing, 2) hearing problem, 3) arthritis/rheumatism, 4) back or neck problem, 5) fracture, bone/joint injury, 6) heart problem, 7) stroke problem, 8) hypertension/high blood pressure, 9) diabetes, 10) lung/breathing problem, 11) cancer, 12) birth defect, 13) intellectual disability, 14) other developmental problem (e.g., cerebral palsy), 15) senility, 16) depression/anxiety/emotional problem, 17) weight problem, 18) missing limbs/amputee, 19) kidney, bladder, or renal problem, 20) circulation problems, 21) benign tumors, cysts, 22) fibromyalgia, lupus, 23) osteoporosis, tendinitis, 24) epilepsy, seizures, 25) multiple sclerosis, muscular dystrophy, 26) polio, paralysis, para/quadruplegia, 27) Parkinson's disease, other tremors, 28) hernia, 29) ulcer, 30) varicose veins, hemorrhoids, 31) thyroid problems, Grave's disease, gout, 32) knee problems, 33) migraine headaches, 34) other injury, 35) other nerve damage, including carpal tunnel syndrome, 36) other impairment/problem. (Endorse all that apply; Refused; Don't know/Not sure). *N* missing data = 160 (0.06%).

Serious psychological distress. DURING THE PAST 30 DAYS, how often did you feel ... [that nothing could cheer you up? Nervous? Restless or fidgety? Hopeless? That everything was an effort? Worthless?] (4=ALL of the time; 3=MOST of the time; 2=SOME of the time; 1=A LITTLE of the time; 0=NONE of the time; Refused; Not ascertained; Don't know). Each response on the 0-4 rating scale to each of the six questions is summed in a composite score (range 0-24) with scores of 13+ indicating positive screen. *N* missing data = 6,245 (2.2%).

Heavy Drinking. In ANY ONE YEAR, have you had at least 12 drinks of any type of alcoholic beverage? (Yes; No; Refused; Not ascertained; Don't know). Participants who reported having had at least one alcoholic beverage in the past year were asked In the PAST YEAR, on how many DAYS did you have 5 or more drinks of any alcoholic beverage (continuous response 0-365 days; Refused; Not ascertained; Don't know). Beginning in 2014, the question was updated for females to ask In the PAST YEAR, on how many DAYS did you have 4 or more drinks of any alcoholic beverage (continuous response 0-365 days; Refused; Not ascertained; Don't know). *N* missing data = 4,865 (1.7%).

Observations for each disadvantage variable were missing for respondents who refused to answer the survey question, whose data was not ascertained, or who reported 'don't know.' Respondents with missing data for ≥ 1 disadvantage indicator were considered missing for the cumulative disadvantages composite (N missing data = 36,252; 13.0%).

Description of Race and Ethnicity Measures

Race: What race or races do you consider yourself to be? Please select 1 or more of these categories (White, Black/African American, Indian [American], Alaska Native, Native Hawaiian, Guamanian or Chamorro, Samoan, Other Pacific Islander, Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Other Asian, Some other race.)

Ethnicity: Do you consider yourself to be Hispanic or Latino? (Yes/No)

Influence of change in operationalization of heavy drinking for females during observation period on study findings.

The NHIS item assessing heavy drinking specifically asked both male and female respondents about the number of occasions on which 5 or more drinks were consumed in the past year from 2008-2013. In 2014, NHIS updated the question to correspond with sex-specific definitions and asked females to specifically report on frequency of episodes of 4 or more drinks; the 5 or more drinks threshold for males was consistent across all survey years.

To address whether the change in NHIS's operationalization of heavy drinking for females impacted the findings, we conducted additional sensitivity analyses to determine whether there were differences in trends of current smoking before and after the change of the heavy drinking measure in 2014 for female non-heavy drinkers and for female heavy drinkers. As illustrated in eFigure 1 in the online supplement, spline models found no evidence that the slope of current smoking over time differed between the 2008-2013 and 2014-2017 periods for both non-heavy drinking females (p-value for added spline term = .45) or for heavy drinking females (p=.76). Thus, it is unlikely that the change in NHIS's operationalization of heavy drinking for women substantively influenced associations of the heavy drinking variable and the cumulative disadvantage variable with smoking.

Influence of multicollinearity in the multivariable model involving all six disadvantages as independent simultaneous regressors.

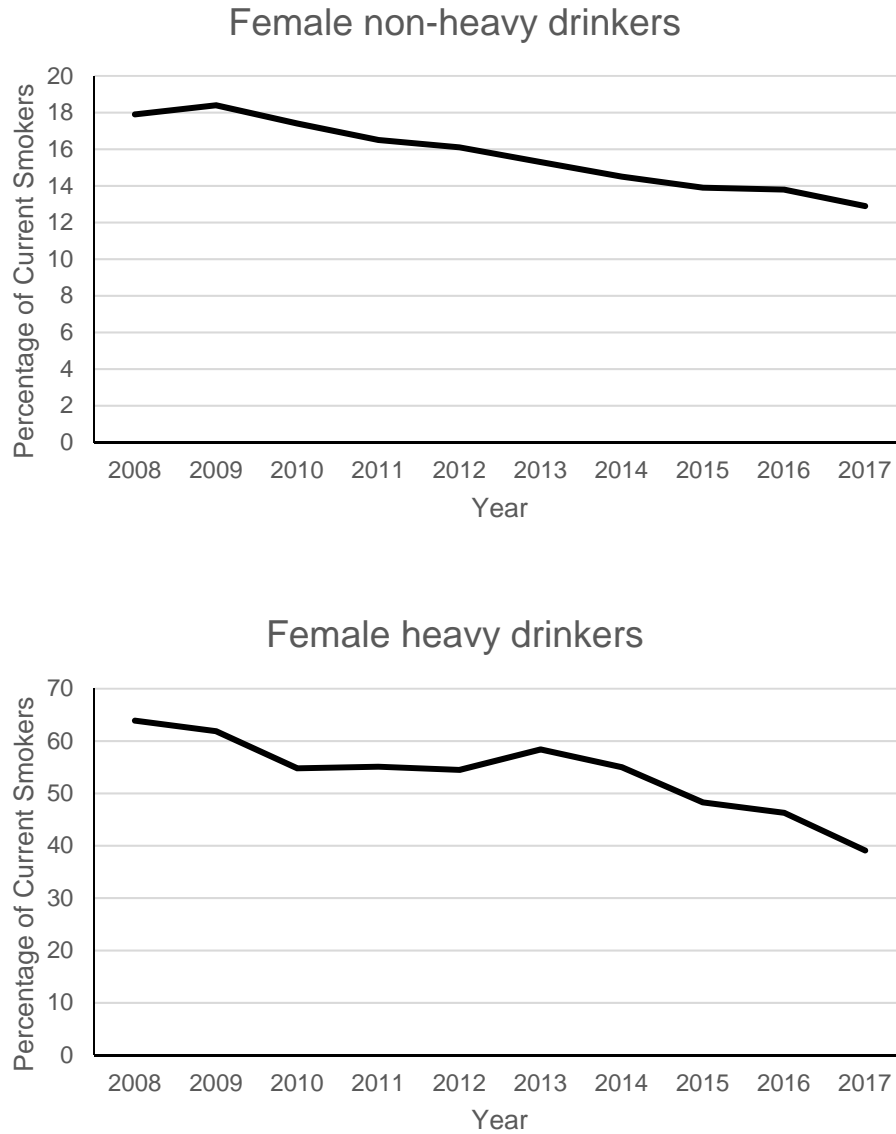
To evaluate the magnitude and implications of multicollinearity among the 6 disadvantage regressors, several aspects of the model results and additional tests were performed, which collectively indicate that the model is not subject to statistical problems introduced by multicollinearity. First, standard error estimates of the model did not appear to be inflated (a consequence of excessive multicollinearity), ranging from 0.019 to 0.038, which correspond to 2% to 5% of the size of the effect estimates, respectively. Second, the mean tetrachoric correlation was 0.32, and each correlation estimate was below 0.60, with one exception (a 0.85 correlation between unemployment and disability). To examine the influence of the collinearity of these two variables on study findings, we compared the OR estimates across three sets of models: 1) the full model with all covariates and all 6 disadvantages reported in the main manuscript in Table 4, 2) a reduced model with all covariates and 5 disadvantage variables that excluded unemployment, 3) a reduced model with all covariates and 5 disadvantage variables that excluded disability/limitation. Differences in effect estimates were negligible for poverty, low education, heavy drinking, and serious psychological distress across the three models (not reported). Effect estimates were merely ~25% higher for unemployment and disability/limitation when the other corresponding regressor was excluded from the model (OR[95%CI] for full model for current vs. never smoker [Table 4]: unemployment 1.83[1.73-1.93] and disability/limitation 1.42[1.37-1.47] vs. OR[95%CI] for reduced models for current vs. never smoker: unemployment 2.31[2.20-2.41] and disability/limitation 1.78[1.73-1.84]). Thus, the likelihood that effect estimates were artificially inflated, deflated, or biased due to excessive multicollinearity is low.

Comparison of the predictive capacity of different specifications of cumulative disadvantage. In the operationalization of the cumulative disadvantage presented in the primary analysis, each disadvantage provided equal weight to the index (i.e., a single point), which is straightforward but may provide insufficient accuracy in modeling the odds of smoking. It is possible that a more nuanced approach to quantifying cumulative disadvantage that allows differential weighting for each disadvantage may provide more precision for modeling the odds of smoking, despite being a more complex approach. Additionally, the primary analysis modeled the disadvantage index as categorical variable, but it is possible that by modeling the 6-level cumulative disadvantage index as continuous with either linear or quadratic trend could augment the predictive power.

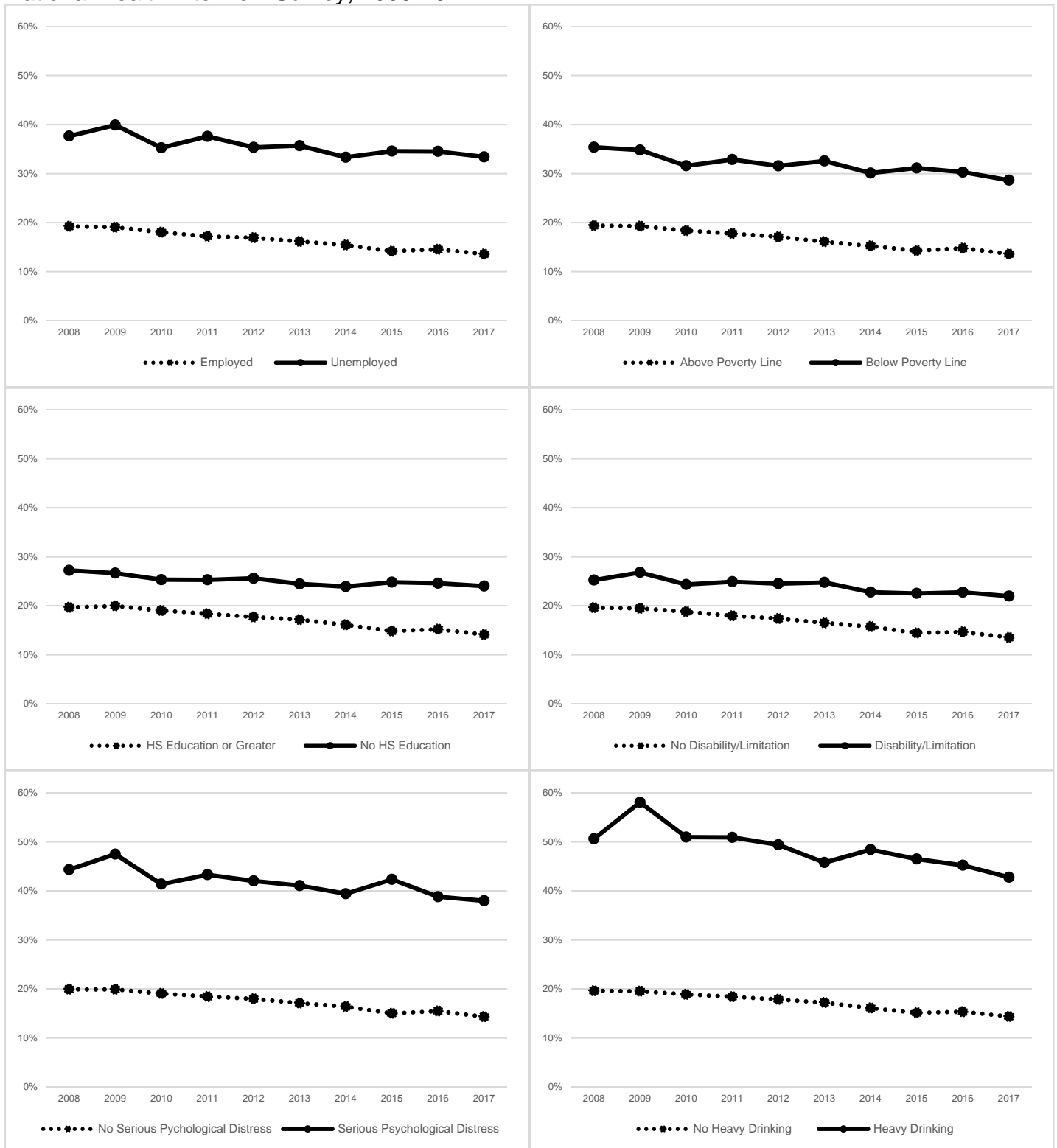
To address these questions, the models were retested with alternate specifications of the cumulative disadvantage variable varying in parsimony-complexity. One alternate operationalization included linear and quadratic specifications of the original 6-level index in which each disadvantage provided equal weighting. Another alternative operationalization of cumulative disadvantage was tested by generating a factor-analytic derived composite score, which reflected an empirically-based latent continuous dimension derived from patterns of covariance across the 6 disadvantage variables used to assess a linear relation with smoking status. An iterated principal factors method for factor analysis with orthogonal varimax rotation was used,¹ which found a single factor solution with one eigenvalue (variance) of 1.16 (explaining 61% of the variance of the factors) and the other eigenvalues <0.59. The factor loadings for each indicator were unemployment (.72), low education (.09), poverty (.21), serious psychological distress (.32), hazardous drinking (.004), and disability/limitation (.70). Finally, the multivariable model depicted in Table 4 of the main manuscript including each disadvantage as a separate variable permitting six separate empirically-derived regression weights was considered, which was the conceptualization involving the highest number of estimated parameters. For each operationalization of cumulative disadvantage, the main models without and with the cumulative index × disadvantage interaction terms were tested.

The predictive accuracy of each specification was estimated by the Pseudo R² of the respective models.² For descriptive purposes, the Pseudo R² of the covariates-only model omitting any cumulative disadvantage variable was also tested. Pseudo R² values for all alternate operationalization models are presented in the eTable in the online supplement. As detailed in the eTable, estimates of the predictive capacity of models did not differ substantially across the various specifications (Pseudo R² range: .071 - .076), except for the factor-analytically-derived composite which provided less predictive power (Pseudo R² = .061). The categorical specification of the cumulative disadvantage indicator appeared to be the most parsimonious operationalization without sacrificing model precision (Pseudo R² = .072).

eFigure 1. Percentage of Current Smokers Among Females by Heavy Drinking Status, 2008-2013



eFigure 2. Proportion of Current Smokers in Each Survey Year, by Each Form of Disadvantage, National Health Interview Survey, 2008-2017^a



^aWeighted to US population. See article’s methods section of manuscript for operational definition of each individual form of disadvantage. See eMethods for description of disability/limitation. Total N = 278,048; varies due to missing data for each disadvantage.

eTable. Pseudo R² Values From Regression Models With Alternate Specifications of Cumulative Disadvantage

Operationalization of cumulative disadvantage	Pseudo R ² for Model ^d	
	Excluding Interaction	Including Year × Disadvantage Interaction
Base models excluding cumulative disadvantage variable (covariates only) ^a	0.0486	--
Each disadvantage indicator as separate regressors ^b	0.0758	0.0760
6-level cumulative disadvantage index with equal weighting to each disadvantage ^c		
Categorical variable	0.0715	0.0717
Continuous variable (linear trend only)	0.0708	0.0709
Continuous variable (linear and quadratic trends)	0.0713	0.0715
Factor-analytically derived continuous latent dimensions (linear trend) ^c	0.0606	0.0607

Note. ^aMultivariable polytomous regression model of odds of current vs. former vs. never smoking including sex, race/ethnicity, age, survey year, and region. ^bMultivariable polytomous regression model of odds of current vs. former vs. never smoking including unemployment, poverty, education, disability/limitation, serious psychological distress, and heavy drinking as simultaneous regressors adjusted for sex, race/ethnicity, age, survey year, and region weighted to US population. ^cMultivariable polytomous regression model of odds of current vs. former vs. never smoking including the respective disadvantage variable and with adjustment for age, sex, race/ethnicity, survey year, and region weighted to US population. ^dYear treated as a 10-level continuous variable (2008-2017). Models are unweighted and fitted using missing indicators to maintain sample size for comparison.

eReferences

1. Kaiser HF. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*. 1958;23(3):187-200.
2. Scott LJ, Freese J. Regression models for categorical dependent variables using Stata. *College Station, Tex: StataCorp LP*. 2006.