A sociodemographic index identifies sex-related effects on insomnia in the Hispanic Community Health Study/Study of Latinos
Supplementary Information

Natali Sorajja[1,2,†], Joon Chung[2,†], Carmela Alcántara[3], Sylvia Wassertheil-Smoller[4], Frank J Penedo[5], Alberto R Ramos[6], Krista M Perreira[7], Martha L Daviglus[8], Shakira F Suglia[9], Linda C Gallo[10], Peter Y Liu[11], Susan Redline[2], Carmen R Isasi[4], Tamar Sofer[1,2,12*]

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
[2]Division of Sleep and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA
[3]School of Social Work, Columbia University, New York, New York, USA
[4]Department of Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, New York, USA
[5]Department of Psychology, University of Miami, Miami, Florida, USA
[6]Department of Neurology, University of Miami Miller School of Medicine, Miami, Florida, USA
[7]Department of Social Medicine, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA
[8]Institute for Minority Health Research, University of Illinois at Chicago, Chicago, Illinois, USA
[9]Department of Epidemiology, Rollins School of Public Health, Atlanta, Georgia, USA
[10]Department of Psychology, San Diego State University, San Diego, California, USA
[11]Division of Genetics, Lundquist Institute at Harbor-UCLA Medical Center, Torrance, CA 90502, USA
[12]CardioVascular Institute (CVI), Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

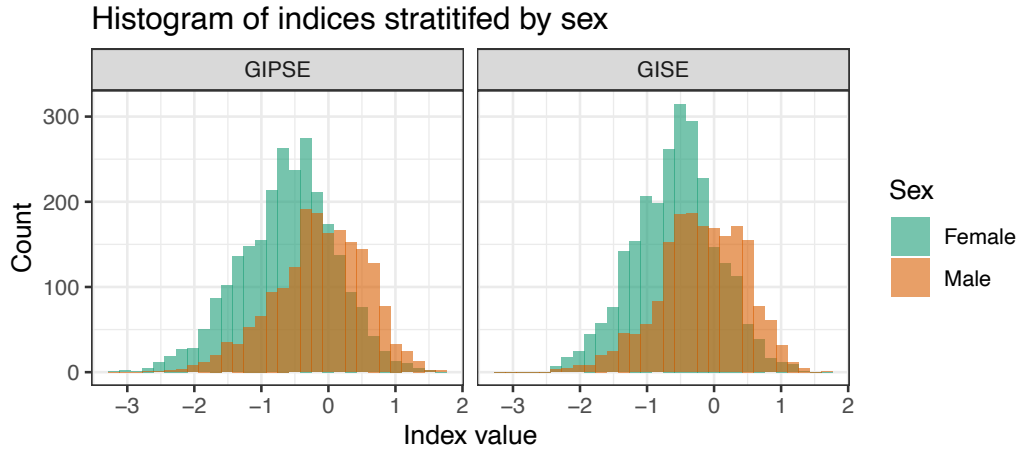
[†]Equally contributing authors

*Correspondence:
Tamar Sofer
Center for Life Sciences CLS-934
3 Blackfan St
Boston, MA 02115
tsofer@bidmc.harvard.edu

Supplementary Figures....3
Supplementary Figure 1: Distribution of gendered indices in independent test dataset....3
Supplementary Figure 2: Missingness patterns....4
Supplementary Figure 3: Distribution of gendered indices in imputed data....5
Supplementary Figure 4: Principal component analysis of demographic, acculturation, and psychological variables demonstrate similar relationship between variables in males and females....6
Supplementary Tables....7
Supplementary Table 1: Sex stratified associations of sociodemographic variables with insomnia....7
Supplementary Table 2: Sex stratified associations of sociodemographic variables with WHIIRS....8
Supplementary Table 3: Psuedo R^2 from association analyses with insomnia....9
Supplementary Table 4: Association analysis of sex and gendered indices with insomnia using imputed data....10
Supplementary Table 5: Association analysis of sex and gendered indices with WHIIRS (complete data)....11
Supplementary Table 6: Association analysis of sex and gendered indices with WHIIRS using imputed data....12
Supplementary Table 7: Interaction analysis of sex and gendered indices with insomnia WHIIRS (complete data)....13
Supplementary Table 8: Interaction analysis of sex and gendered indices with insomnia WHIIRS (imputed data)....13

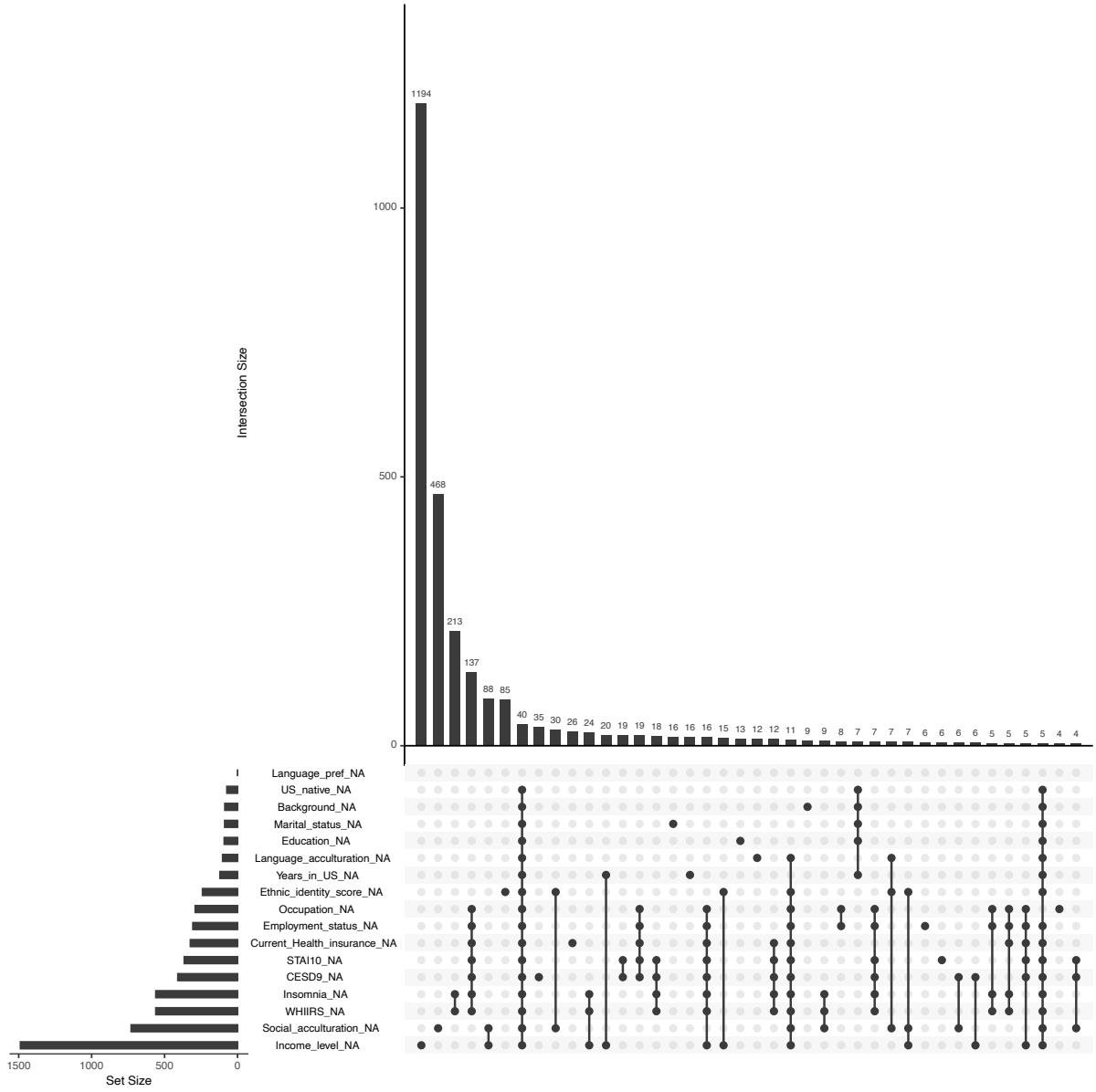
Supplementary Figures

Supplementary Figure 1: Distribution of gendered indices in independent test dataset



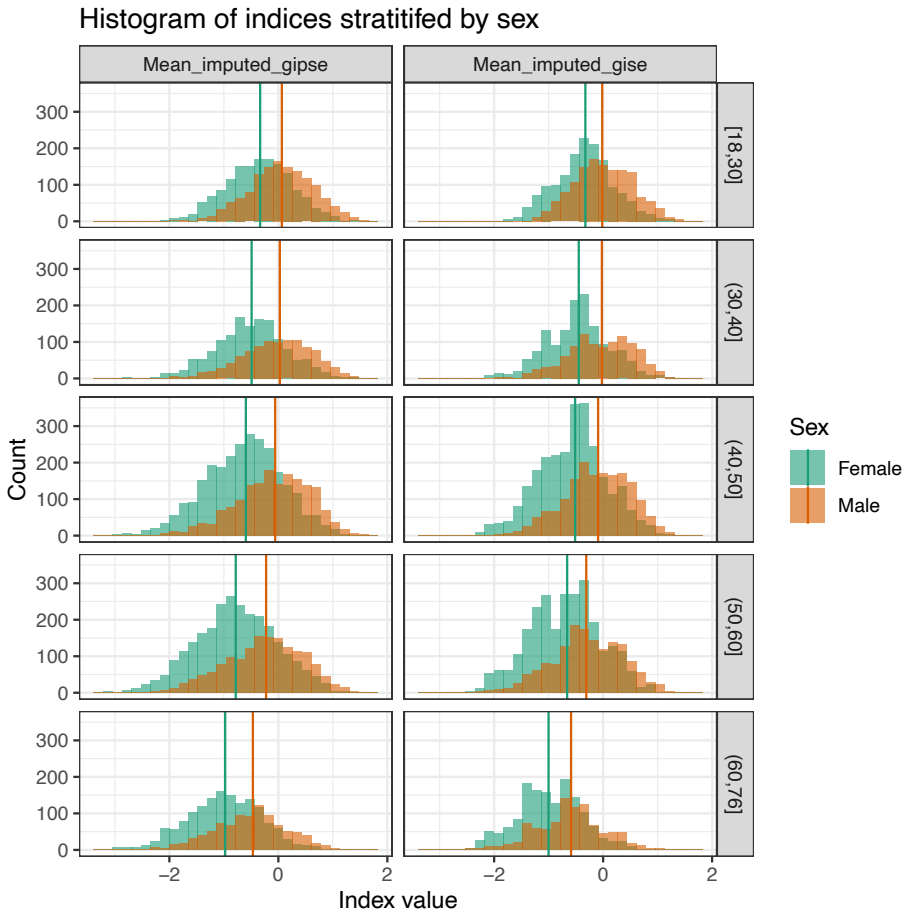
Histograms of the GISE and GIPSE, stratified by age group, in a dataset composed of 4,070 individuals who are all from different primary sampling units of the 9,596 individuals who were used to train the indices.
GISE: gendered index of sociodemographic environment; GIPSE: gendered index of psychological and sociodemographic environment.

Supplementary Figure 2: Missingness patterns



Missingness patterns across variables used in the analysis as potentially associated with socio-demographic sex-related patterns. The left panel display the various variables and cumulative number of observations with missing values for these variables. The top panel provides the number of missing values for each "pattern" of missing values, where a pattern is defined by combinations variables having missing values. The bottom right, connected point panel, describes missingness patterns. The figure only visualizes patterns observed in up to 4 observations.

Supplementary Figure 3: Distribution of gendered indices in imputed data



Histograms of GISE and GIPSE computed based on the imputed dataset. The lines correspond to median values of the indices within male and female participants in the relevant age group.

GISE: gendered index of sociodemographic environment; GIPSE: gendered index of psychological and sociodemographic environment.

Supplementary Figure 4: Principal component analysis of demographic, acculturation, and psychological variables demonstrate similar relationship between variables in males and females

Loadings of the first principal components of sociodemographic variables used, computed separately in males and in females only.

Supplementary Tables

Supplementary Table 1: Sex stratified associations of sociodemographic variables with insomnia

	Estimated association in female stratum			Estimated association in male stratum			Sex interaction
Exposure	**Estimated OR**	**95% CI**	**p-value**	**Estimated OR**	**95% CI**	**p-value**	**p-value**
Marital status: Married or living with a partner	0.89	(0.74,1.08)	0.250	0.74	(0.58,0.93)	0.009**	0.146
Marital status: Separated, divorced, or widow(er)	1.1	(0.88,1.37)	0.415	0.84	(0.63,1.12)	0.229	0.084
Occupation: Service Worker	1.14	(0.87,1.49)	0.340	1.01	(0.74,1.39)	0.945	0.665
Occupation: Skilled Worker	0.97	(0.75,1.26)	0.833	1.14	(0.92,1.43)	0.226	0.437
Occupation: Professional/technical, administrative/executive, or office staff	1.04	(0.82,1.32)	0.752	0.93	(0.67,1.28)	0.639	0.516
Occupation: Other occupation	1.08	(0.84,1.38)	0.558	1.13	(0.87,1.46)	0.355	0.753
Income level: $10,001-$20,000	0.99	(0.82,1.2)	0.897	0.7	(0.54,0.92)	0.011*	0.052
Income level: $20,001-$40,000	0.85	(0.68,1.06)	0.145	0.61	(0.46,0.8)	<0.001***	0.068
Income level: $40,001-$75,000	0.73	(0.53,0.99)	0.041*	0.65	(0.47,0.9)	0.011*	0.651
Income level: More than $75,000	0.51	(0.33,0.78)	0.002**	0.5	(0.28,0.89)	0.019*	0.923
Employment status: Employed part-time (<=35 hours/week)	0.79	(0.65,0.96)	0.019*	0.77	(0.59,1.01)	0.059	0.945
Employment status: Employed full-time (>35 hours/week)	0.71	(0.6,0.84)	<0.001***	0.66	(0.54,0.81)	<0.001***	0.761
Education: At most high school diploma/GED	0.94	(0.75,1.18)	0.588	1.1	(0.87,1.39)	0.426	0.262
Education: Greater than high school/GED	**0.8**	**(0.67,0.96)**	**0.016***	**1.07**	**(0.86,1.35)**	**0.531**	**0.050**
Years in US: 10 Years or More	1.13	(0.94,1.36)	0.190	1.12	(0.89,1.42)	0.328	0.657
Years in US: US born	1.47	(1.13,1.91)	0.004**	1.54	(1.13,2.11)	0.007**	0.761
Language acculturation	1.12	(1.02,1.21)	0.012*	1.15	(1.04,1.27)	0.008**	0.375
Social acculturation	1.05	(0.91,1.22)	0.494	0.97	(0.82,1.16)	0.772	0.591
Ethnic identity score	0.98	(0.86,1.12)	0.797	0.97	(0.83,1.14)	0.730	0.869
STAI10	**1.12**	**(1.1,1.13)**	**<0.001*** **	**1.15**	**(1.13,1.17)**	**<0.001*** **	**0.007****
CESD9	1.13	(1.11,1.15)	<0.001***	1.13	(1.1,1.15)	<0.001***	0.705
Language preference	1.21	(0.97,1.5)	0.086	1.43	(1.11,1.84)	0.005**	0.188
Current health insurance	0.84	(0.7,1)	0.047*	0.9	(0.74,1.09)	0.274	0.909

Association analyses were performed separately within each sex stratum, using complete data. Interaction test was performed in a combined-sex analysis, with a multiplicative interaction term. Each sociodemographic variable was assessed in a separate analysis, but variables with multiple levels were included in the same regression model. Analyses were adjusted for age, study center, self-reported Hispanic/Latino background (and combined-sex model to sex as well). We used survey logistic regression with quasibinomial link function.

Supplementary Table 2: Sex stratified associations of sociodemographic variables with WHIIRS

	Estimated association in female stratum			Estimated association in male stratum			Sex interaction
Exposure	**Estimated effect**	**95% CI**	**p-value**	**Estimated effect**	**95% CI**	**p-value**	**p-value**
Marital status: Married or living with a partner	0.67	(0.43,1.04)	0.073	0.5	(0.31,0.81)	0.005**	0.479
Marital status: Separated, divorced, or widow(er)	1.48	(0.85,2.58)	0.165	0.84	(0.44,1.6)	0.603	0.086
Occupation: Service Worker	1.12	(0.61,2.04)	0.714	1.61	(0.82,3.15)	0.167	0.418
Occupation: Skilled Worker	0.98	(0.58,1.65)	0.934	1.62	(1.05,2.51)	0.030*	0.256
Occupation: Professional/technical, administrative/executive, or office staff	1.01	(0.59,1.7)	0.985	1.06	(0.57,1.97)	0.842	0.895
Occupation: Other occupation	1.21	(0.67,2.18)	0.526	1.16	(0.71,1.91)	0.551	0.849
Income level: $10,001-$20,000	0.62	(0.38,1)	0.048*	0.33	(0.16,0.66)	0.002**	0.183
Income level: $20,001-$40,000	0.38	(0.22,0.65)	<0.001***	0.22	(0.11,0.45)	<0.001***	0.382
Income level: $40,001-$75,000	0.31	(0.17,0.59)	<0.001***	0.27	(0.12,0.57)	<0.001***	0.941
Income level: More than $75,000	0.16	(0.06,0.41)	<0.001***	0.18	(0.07,0.46)	<0.001***	0.636
Employment status: Employed part-time (<=35 hours/week)	0.54	(0.34,0.84)	0.007**	0.41	(0.24,0.69)	<0.001***	0.786
Employment status: Employed full-time (>35 hours/week)	0.38	(0.25,0.57)	<0.001***	0.3	(0.2,0.45)	<0.001***	0.728
Education: At most high school diploma/GED	0.76	(0.46,1.26)	0.290	0.97	(0.59,1.6)	0.912	0.271
Education: Greater than high school/GED	**0.49**	**(0.31,0.75)**	0.001**	**1.01**	**(0.63,1.62)**	0.972	0.012*
Years in US: 10 Years or More	1.27	(0.85,1.92)	0.248	1.3	(0.83,2.04)	0.258	0.498
Years in US: US born	2.77	(1.57,4.86)	<0.001***	4.21	(2.36,7.48)	<0.001***	0.338
Language acculturation	1.43	(1.19,1.71)	<0.001***	1.6	(1.33,1.93)	<0.001***	0.300
Social acculturation	1.11	(0.78,1.57)	0.579	1.28	(0.94,1.74)	0.114	0.551
Ethnic identity score	0.9	(0.66,1.24)	0.524	0.92	(0.68,1.24)	0.564	0.935
STAI10	**1.38**	**(1.34,1.41)**	<0.001***	**1.44**	**(1.39,1.49)**	<0.001***	0.042*
CESD9	1.43	(1.39,1.48)	<0.001***	1.39	(1.34,1.45)	<0.001***	0.135
Language preference	1.78	(1.13,2.82)	0.013*	2.84	(1.74,4.64)	<0.001***	0.207
Current health insurance	0.71	(0.48,1.04)	0.078	0.83	(0.57,1.22)	0.353	0.680

Association analyses were performed separately within each sex stratum, using complete data. Interaction test was performed in a combined-sex analysis, with a multiplicative interaction term. Each sociodemographic variable was assessed in a separate analysis, but variables with multiple levels were included in the same regression model. Analyses were adjusted for age, study center, self-reported Hispanic/Latino background (and combined-sex model to sex as well). We used survey linear regression with gaussian link function.

Supplementary Table 3: Psuedo R^2 from association analyses with insomnia

model	Sex combined analysis	Male stratum	Female stratum
Model 0: adjusting for covariates without sex adjustment	4.36	3.55	5.3
Model 1: adjusting for covariates.	5.56	-	-
Model 2: adjusting for covariates, GISE	5.64	3.61	5.42
Model 3: adjusting for covariates, GIPSE	8.09	5.46	8.53
Model 4: adjusting for covariates, components of GISE	6.89	5.21	6.82
Model 5: adjusting for covariates, components of GIPSE	14.62	12.64	15.27

Pseudo R^2 are reported on percentage scale for interpretability.
Sex-combined n=13,666; male n=5,593; female n=8,083.
Adjusting covariates were age, Hispanic/Latino background, and study center.
GISE: gendered index of sociodemographic environment; GIPSE: gendered index of psychological and sociodemographic environment.

Supplementary Table 4: Association analysis of sex and gendered indices with insomnia using imputed data

Model	Estimated male sex effect			Estimated gendered index effect		
	Estimated OR	95% CI	p-value	Estimated OR	95% CI	p-value
Model 1: adjusting for baseline covariates.	0.61	(0.55,0.67)	<0.001***	NA	NA	NA
Model 2: adjusting for baseline covariates, GISE	0.63	(0.57,0.70)	<0.001***	0.93	(0.88,0.99)	0.016*
Model 3: adjusting for baseline covariates, GIPSE	0.79	(0.71,0.88)	<0.001***	0.65	(0.61,0.69)	<0.001***
Model 4: adjusting for baseline covariates, components of GISE	0.63	(0.56,0.70)	<0.001***	NA	NA	NA
Model 5: adjusting for baseline covariates, components of GIPSE	0.75	(0.66,0.84)	<0.001***	NA	NA	NA
Analysis in male stratum						
Model adjusting for baseline covariates, GISE	NA	NA	NA	0.95	(0.86,1.03)	0.194
Model adjusting for baseline covariates, GIPSE	NA	NA	NA	0.68	(0.62,0.75)	<0.001***
Analysis in Female stratum						
Model adjusting for baseline covariates, GISE	NA	NA	NA	0.92	(0.85,0.99)	0.030*
Model adjusting for baseline covariates, GIPSE	NA	NA	NA	0.63	(0.58,0.68)	<0.001***

Supplementary Table 5: Association analysis of sex and gendered indices with WHIIRS (complete data)

Model	Estimated male sex effect				Estimated gendered index effect	
	Estimated effect	95% CI	p-value	Estimated effects	95% CI	p-value
Model 1: adjusting for baseline covariates.	-1.25	(-1.51,-0.99)	<0.001***	NA	NA	NA
Model 2: adjusting for baseline covariates, GISE	-1.11	(-1.38,-0.84)	<0.001***	-0.23	(-0.38,-0.09)	0.002**
Model 3: adjusting for baseline covariates, GIPSE	-0.47	(-0.75,-0.2)	<0.001***	-1.18	(-1.32,-1.03)	<0.001***
Model 4: adjusting for baseline covariates, components of GISE	-1.12	(-1.39,0.85)	<0.001***	NA	NA	NA
Model 5: adjusting for baseline covariates, components of GIPSE	-0.6	(-0.85,0.35)	<0.001***	NA	NA	NA
Analysis in male stratum						
Model adjusting for baseline covariates, GISE	NA	NA	NA	-0.26	(-0.45,-0.06)	0.010**
Model adjusting for baseline covariates, GIPSE	NA	NA	NA	-1.05	(-1.26,-0.85)	<0.001***
Analysis in Female stratum						
Model adjusting for baseline covariates, GISE	NA	NA	NA	-0.22	(-0.44,-0.01)	0.042*
Model adjusting for baseline covariates, GIPSE	NA	NA	NA	-1.29	(-1.5,-1.09)	<0.001***

Supplementary Table 6: Association analysis of sex and gendered indices with WHIIRS using imputed data

Model	Estimated male sex effect			Estimated gendered index effect		
	Estimated effect	95% CI	p-value	Estimated effects	95% CI	p-value
Model 1: adjusting for baseline covariates.	-1.26	(-1.49,-1.03)	<0.001***	NA	NA	NA
Model 2: adjusting for baseline covariates, GISE	-1.14	(-1.37,-0.9)	<0.001***	-0.21	(-0.34,-0.07)	0.003**
Model 3: adjusting for baseline covariates, GIPSE	-0.49	(-0.73,-0.24)	<0.001***	-1.19	(-1.32,-1.06)	<0.001***
Model 4: adjusting for baseline covariates, components of GISE	-1.14	(-1.38,-0.91)	<0.001***	NA	NA	NA
Model 5: adjusting for baseline covariates, components of GIPSE	-0.6	(-0.82,-0.38)	<0.001***	NA	NA	NA
Analysis in male stratum						
Model adjusting for baseline covariates, GISE	NA	NA	NA	-0.24	(-0.42,-0.05)	0.013*
Model adjusting for baseline covariates, GIPSE	NA	NA	NA	-1.06	(-1.25,-0.86)	<0.001***
Analysis in Female stratum						
Model adjusting for baseline covariates, GISE	NA	NA	NA	-0.19	(-0.39,0.01)	0.059
Model adjusting for baseline covariates, GIPSE	NA	NA	NA	-1.3	(-1.48,-1.12)	<0.001***

Supplementary Table 7: Interaction analysis of sex and gendered indices with insomnia WHIIRS (complete data)

		Male sex effect			Index effect			Index-male sex multiplicative effect		
Outcome	**Index**	**Estimated effect**	**95% CI**	**p-value**	**Estimated effect**	**95% CI**	**p-value**	**Estimated effect**	**95% CI**	**p-value**
Insomnia	**GISE**	0.63	(0.56,0.71)	<0.001***	0.91	(0.84,0.98)	0.015*	1.04	(0.93,1.17)	0.458
Insomnia	**GIPSE**	0.79	(0.69,0.89)	<0.001***	0.63	(0.58,0.68)	<0.001***	1.1	(0.97,1.24)	0.140
WHIIRS	**GISE**	-1.11	(-1.38,-0.84)	<0.001***	-0.27	(-0.48,-0.07)	0.009**	0.08	(-0.18,0.35)	0.537
WHIIRS	**GIPSE**	-0.48	(-0.75,-0.2)	<0.001***	-1.33	(-1.53,-1.13)	<0.001***	0.32	(0.04,0.6)	0.024*

Association analyses used complete data (n=13,666 individuals) and used survey regression, using the R survey package, to account for study design (stratification, sampling, non-response, etc.). When insomnia was the outcome, we used survey logistic regression with quasibinomial link function. When WHIIRS was the outcome, we used survey linear regression with gaussian link function. Association models were adjusted for age, study center, and self-reported Hispanic/Latino background.

Supplementary Table 8: Interaction analysis of sex and gendered indices with insomnia WHIIRS (imputed data)

		Male sex effect			Index effect			Index-male sex multiplicative effect		
Outcome	**Index**	**Estimated effect**	**95% CI**	**p-value**	**Estimated effect**	**95% CI**	**p-value**	**Estimated effect**	**95% CI**	**p-value**
Insomnia	**GISE**	0.63	(0.57,0.71)	<0.001***	0.91	(0.85,0.98)	0.012*	1.05	(0.94,1.17)	0.361
Insomnia	**GIPSE**	0.8	(0.72,0.9)	<0.001***	0.62	(0.58,0.67)	<0.001***	1.1	(0.99,1.23)	0.089
WHIIRS	**GISE**	-1.14	(-1.37,-0.9)	<0.001***	-0.25	(-0.44,-0.06)	0.011*	0.08	(-0.17,0.34)	0.532
WHIIRS	**GIPSE**	-0.5	(-0.74,-0.25)	<0.001***	-1.33	(-1.51,-1.16)	<0.001***	0.32	(0.06,0.57)	0.014*

Association analyses used multiply-imputed data (n=16,415 individuals) and used survey regression, using the R survey package, to account for study design (stratification, sampling, non-response, etc.). When insomnia was the outcome, we used survey logistic regression with quasibinomial link function. When WHIIRS was the outcome, we used survey linear regression with gaussian link function. Association models were adjusted for age, study center, and self-reported Hispanic/Latino background. Results from the 5 imputed datasets were combined using Rubin's rule.