These Supplemental Digital Contents are provided by the authors of the following paper, to provide additional detail for interested readers:

Mahendran M, Lizotte D, Bauer GR. Describing intersectional health outcomes: An evaluation of data analysis methods. *Epidemiology*.

Supplemental Digital Content

eAppendix 1: Additional simulation processes

1a. Comparing Bayesian and Frequentist MAIHDA models

100 simulations were conducted for each of the three scenarios. Sample sizes of 10,000 were used for each model. The Bayesian multilevel models were calculated using the R brms package. Bayesian (B) multilevel models were performed each with 1000 burn ins, 2000 total. Frequentist (F) multilevel models were created with package lme4, using R version 3.5.3. Presented below are 0.025 and 0.975 percentiles of estimates from the 100 simulations).

Scenario 1: y = x1 + x2 + x3 + x4 + x5 + x1*x2

P(x1=1) = **50%;** P(x2=1) = **50%;** P(x3=1) = 50%; P(x4=1) = 50%; P(x5=1) = 50%;

eTable 1. 0.025 and 0.975 percentiles of Scenario 1 from 100 simulations

	OLS	OLS with interaction	MLM (B)	MLM (F)	MLM (B) with interaction	MLM (F) with interaction
Intercept	(-0.298 , -0.197)	(-0.056 , 0.052)	(-0.299 , -0.201)	(-0.297, -0.200)	(-0.056 , 0.052)	(-0.056, 0.052)
x1	(1.466 , 1.538)	(0.936 , 1.056)	(1.459 , 1.534)	(1.463, 1.532)	(0.936 , 1.056)	(0.935, 1.056)
x2	(1.454 , 1.540)	(0.948 , 1.062)	(1.458 , 1.539)	(1.456, 1.538)	(0.948 , 1.062)	(0.948, 1.062)
x3	(0.970 , 1.037)	(0.970 , 1.040)	(0.969 , 1.039)	(0.969, 1.041)	(0.970 , 1.041)	(0.970, 1.041)
x4	(0.966 , 1.038)	(0.968 , 1.036)	(0.969 , 1.036)	(0.969, 1.036)	(0.968 , 1.036)	(0.968, 1.036)
x5	(0.959 , 1.042)	(0.961 , 1.038)	(0.960 , 1.039)	(0.961, 1.039)	(0.961 , 1.039)	(0.961, 1.038)

x1:x2	-	(0.924 , 1.071)	-	-	(0.923 , 1.072)	(0.924, 1.071)

Scenario 2: y = x1 + x2 + x3 + x4 + x5 + x1*x2

P(x1=1) = **70%;** P(x2=1) = **70%;** P(x3=1) = 50%; P(x4=1) = 50%; P(x5=1) = 50%;

eTable 2. 0.025 and 0.975 percentiles of Scenario 2 from 100 simulations

	OLS	OLS with interaction	MLM (B)	MLM (F)	MLM with interaction (B)	MLM with interaction (F)
Intercept	(-0.552 , -0.441)	(-0.078 , 0.064)	(-0.322 , -0.208)	(-0.326, -0.211)	(-0.078 , 0.063)	(-0.077, 0.064)
x1	(1.663 , 1.736)	(0.925 , 1.082)	(1.465 , 1.555)	(1.461, 1.554)	(0.925 , 1.082)	(0.925, 1.081)
x2	(1.648 , 1.749)	(0.927 , 1.084)	(1.461 , 1.558)	(1.461, 1.559)	(0.927 , 1.083)	(0.927, 1.084)
x3	(0.968 , 1.041)	(0.97 , 1.04)	(0.963 , 1.039)	(0.963, 1.042)	(0.969 , 1.039)	(0.969, 1.039)
x4	(0.967 , 1.037)	(0.968 , 1.036)	(0.959 , 1.044)	(0.961, 1.042)	(0.969 , 1.037)	(0.967, 1.036)
x5	(0.962 , 1.045)	(0.961 , 1.039)	(0.96 , 1.044)	(0.957, 1.045)	(0.959 , 1.039)	(0.958, 1.040)
x1:x2	-	(0.889 , 1.097)	-	-	(0.889 , 1.097)	(0.889, 1.097)

Scenario 3: y = x1 + x2 + x3 + x4 + x5 – 2(x1*x2)

P(x1=1) = **20%;** P(x2=1) = **20%;** P(x3=1) = 50%; P(x4=1) = 50%; P(x5=1) = 50%;

eTable 3. 0.025 and 0.975 percentiles of Scenario 3 from 100 simulations

	OLS	OLS with interaction	MLM (B)	MLM (F)	MLM (B) with interaction	MLM (F) with interaction
Intercept	(0.040 , 0.118)	(-0.039 , 0.033)	(0.432 , 0.537)	(0.432, 0.535)	(-0.039 , 0.033)	(-0.038, 0.033)
x1	(0.546 , 0.653)	(0.949 , 1.053)	(-0.048 , 0.077)	(-0.045, 0.078)	(0.949 , 1.054)	(0.949, 1.053)
x2	(0.544 , 0.657)	(0.949 , 1.056)	(-0.051 , 0.081)	(-0.051, 0.085)	(0.950 , 1.056)	(0.949, 1.056)
x3	(0.964 , 1.040)	(0.970 , 1.040)	(0.947 , 1.060)	(0.946, 1.060)	(0.967 , 1.039)	(0.966, 1.040)
x4	(0.968 , 1.038)	(0.968 , 1.036)	(0.935 , 1.057)	(0.936, 1.057)	(0.968 , 1.037)	(0.968, 1.037)
x5	(0.957 , 1.036)	(0.961 , 1.038)	(0.941 , 1.062)	(0.943, 1.058)	(0.958 , 1.039)	(0.959, 1.038)
x1:x2	-	(-2.123 , -1.889)	-	-	(-2.122 , -1.888)	(-2.123, -1.889)

1b. Description of power calculation for beta coefficients

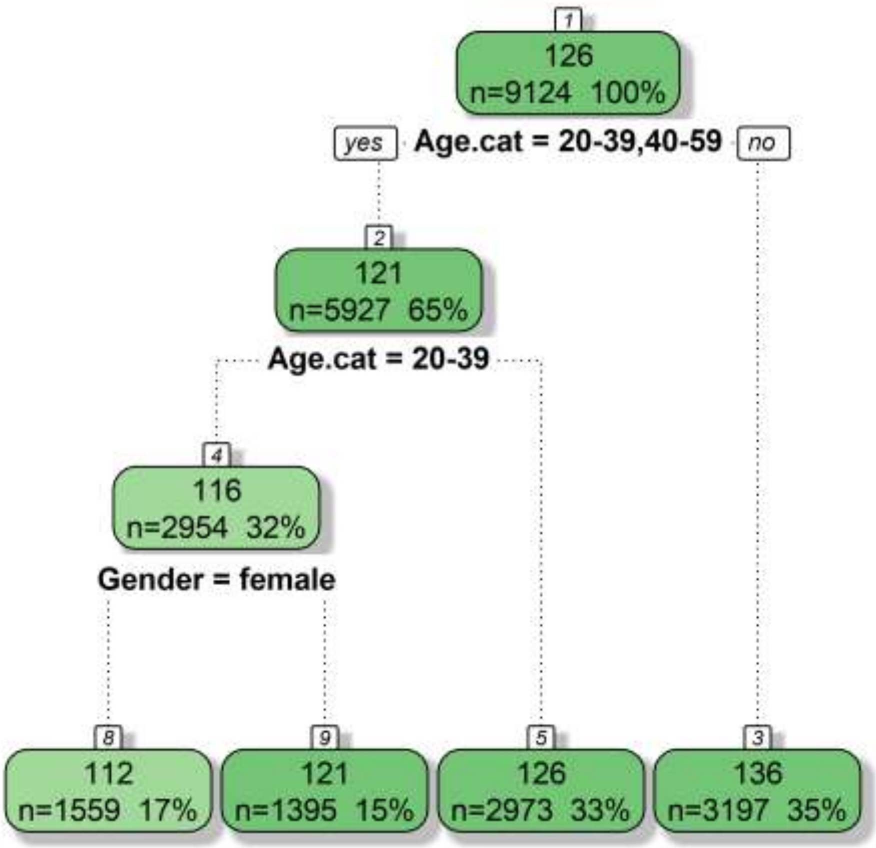
Minimum effect size was determined by a power analysis using a main effects regression at sample size 25000. The minimum effect size was when variables X1 to X5 were detected as significant 80% of the time at $p<0.05$. 100 iterations were used to determine 80% power.

The input variables for the power calculation were X1 to X6, where all variables were either binary or categorical, based on the predictor combination shown in Table 1 of the main text. Two changes were made to the categorical inputs model that differed from what is shown in Table1, that were justified based on the aim to remain relevant to intersectionality research. Firstly, the models for the power calculations were created and evaluated with main effects only, even though the models in the actual simulations include interaction terms. Presumably if an effect size is significant for an "additive effects" model (additive effects by the intersectionality definition, meaning no interaction), then it is still an important enough size for the detection of interaction terms. Secondly, the input variables did not have the same distribution as in the actual simulation models. In the power calculation models, variables X1 to X3, and X5 and X6 were split in equally sized categories. Only X4 was not equally distributed, due to the mediation relationship between X3 and X4. The justification is that in ideal circumstances, calculating outcomes for each intersectional group would not be affected by intersection size, especially when those experiencing marginalization may belong to groups with smaller cell sizes.

The sampling of positive and negative beta coefficients was centred around 1 and -1. Based on the power calculation the minimum effect size was 0.06/ Positive coefficients were selected from a truncated normal distribution with a minimum of 0.06 and a maximum of (2 – 0.06 =1.94). The negative coefficients were selected from a truncated normal distribution with a minimum of (-2 + 0.06 = -1.94) and a maximum of -0.06.

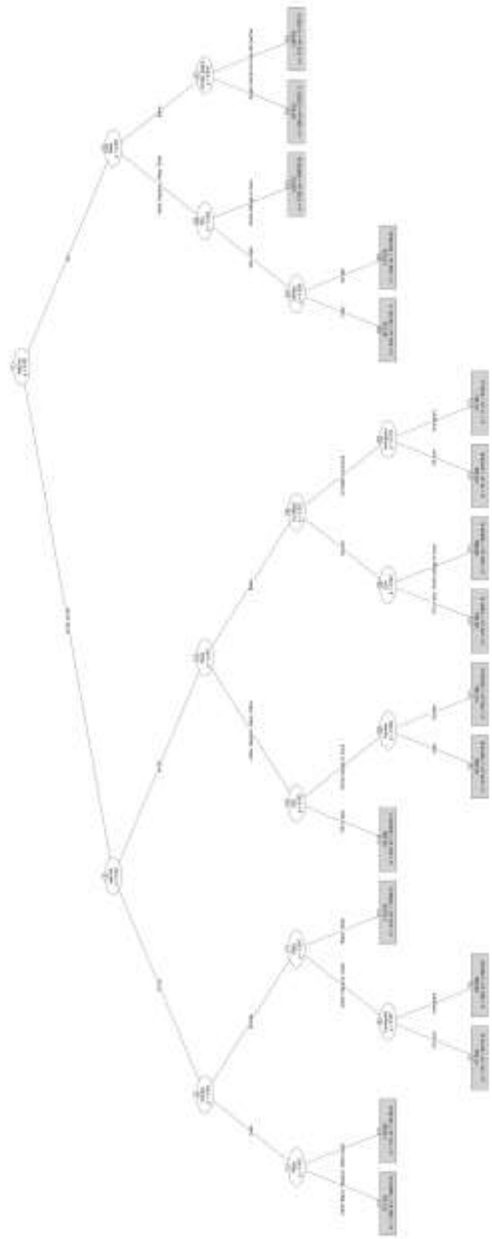
eAppendix 2: Method-specific NHANES results

2a. Variable selection results (single decision trees)



eFigure 1. NHANES variable selection example CART model

The CTree image below can be zoomed in to view the individual splitting patterns.



eFigure 2. NHANES variable selection example CTree model

2b. Cross-classification results

eTable 4. Cross-classification results

Intersection	mean SBP (mm Hg)	SE	n
female Asian 20-39 HS or less	106.79	1.07	32
female Asian 20-39 some college or more	109.07	1.44	176
female Asian 40-59 HS or less	126.38	2.50	79
female Asian 40-59 some college or more	121.60	2.12	153
female Asian 60+ HS or less	138.94	2.47	73
female Asian 60+ some college or more	134.05	2.36	86
female Black 20-39 HS or less	115.20	1.63	118
female Black 20-39 some college or more	116.05	1.80	234
female Black 40-59 HS or less	133.55	2.63	119
female Black 40-59 some college or more	129.07	2.42	252
female Black 60+ HS or less	143.54	3.00	146
female Black 60+ some college or more	140.06	2.67	175
female Hispanic 20-39 HS or less	110.22	1.59	218
female Hispanic 20-39 some college or more	110.90	1.39	209
female Hispanic 40-59 HS or less	125.47	2.44	254
female Hispanic 40-59 some college or more	123.01	2.20	182
female Hispanic 60+ HS or less	136.50	2.46	272
female Hispanic 60+ some college or more	131.95	2.47	126
female other 20-39 HS or less	114.46	1.01	21
female other 20-39 some college or more	114.87	1.50	59
female other 40-59 HS or less	127.12	3.93	14
female other 40-59 some college or more	122.64	2.78	45
female other 60+ HS or less	140.94	3.04	17
female other 60+ some college or more	136.09	2.20	31
female white 20-39 HS or less	111.50	1.16	127
female white 20-39 some college or more	111.67	1.41	365
female white 40-59 HS or less	124.49	1.86	151
female white 40-59 some college or more	120.01	1.99	326
female white 60+ HS or less	138.32	2.82	273
female white 60+ some college or more	133.71	2.71	361
male Asian 20-39 HS or less	120.89	1.50	33
male Asian 20-39 some college or more	116.09	1.46	164
male Asian 40-59 HS or less	125.09	2.41	52
male Asian 40-59 some college or more	122.85	1.88	179
male Asian 60+ HS or less	140.57	2.82	53
male Asian 60+ some college or more	133.89	2.69	76
male Black 20-39 HS or less	121.01	1.63	135

male Black 20-39 some college or more	123.20	2.08	155
male Black 40-59 HS or less	135.58	2.51	140
male Black 40-59 some college or more	131.63	2.55	159
male Black 60+ HS or less	139.51	3.07	174
male Black 60+ some college or more	136.90	2.79	175
male Hispanic 20-39 HS or less	121.74	1.51	230
male Hispanic 20-39 some college or more	119.65	1.31	141
male Hispanic 40-59 HS or less	125.77	2.05	221
male Hispanic 40-59 some college or more	125.50	1.95	104
male Hispanic 60+ HS or less	134.11	2.32	261
male Hispanic 60+ some college or more	134.77	2.57	125
male other 20-39 HS or less	120.39	1.40	34
male other 20-39 some college or more	124.22	2.28	46
male other 40-59 HS or less	120.81	2.15	32
male other 40-59 some college or more	124.86	1.80	49
male other 60+ HS or less	132.25	2.05	29
male other 60+ some college or more	137.94	2.44	27
male white 20-39 HS or less	120.06	1.58	188
male white 20-39 some college or more	120.80	1.59	269
male white 40-59 HS or less	125.34	2.39	175
male white 40-59 some college or more	125.04	1.94	287
male white 60+ HS or less	133.19	2.53	299
male white 60+ some college or more	131.29	2.29	418

2c. Main effects regression (non-intersectional method) results

Note that results presented here are for individual coefficient estimates, but outcome estimates for each intersection, as presented in the main text, were calculated using the "predict" function in R.

eTable 5. Main effects regression results

	Estimate	SE	P-value
Intercept	117.23	0.51	< 0.001
Gender (ref = Male)			
Female	-2.44	0.36	< 0.001
Race/ethnicity (ref = White)			
Black	5.84	0.49	< 0.001
Hispanic	0.31	0.48	0.508
Asian	-0.18	0.59	0.757
Other	1.86	0.91	0.04
Education (ref = Highschool or less)			
College	-2.11	0.37	< 0.001
Age (ref = 20 to 39)			
40-59	9.49	0.45	< 0.001
60 plus	19.27	0.44	< 0.001

2d. Regression (saturated) results

Note that results presented here are for individual coefficient estimates, but outcome estimates for each intersection, as presented in the main text, were calculated using the "predict" function in R.

eTable 6. Saturated regression results

	Estimate	SE	P-value
Intercept	120.05	1.23	< 0.001
Gender (ref = Male)			
Female	-8.55	1.94	< 0.001
Race/ethnicity (ref = White)			
Black	0.95	1.91	0.618
Hispanic	1.69	1.66	0.310
Asian	0.83	3.19	0.794
Other	0.34	3.15	0.915
Education (ref = Highschool or less)			
College	0.74	1.61	0.644
Age (ref = 20 to 39)			
40-59	5.29	1.78	0.003
60 plus	13.14	1.58	< 0.001
Gender*Race/ethnicity			
Female*Black	2.75	2.89	0.341
Female*Hispanic	-2.98	2.52	0.237
Female*Asian	-5.55	4.63	0.231
Female*Other	2.62	5.08	0.606
Gender*Education			
Female*College	-0.58	2.37	0.806
Gender*Age			
Female*40-59	7.70	2.70	0.004
Female*60 plus	13.67	2.41	< 0.001
Race/ethnicity*Education			
Black*College	1.44	2.56	0.573
Hispanic*College	-2.83	2.42	0.242
Asian*College	-5.54	3.61	0.125
Other*College	3.08	4.15	0.458
Race/ethnicity*Age			
Black*40-59	9.28	2.71	0.001
Hispanic*40-59	-1.27	2.39	0.595
Asian*40-59	-1.09	4.17	0.794
Other*40-59	-4.87	4.53	0.283
Black*60 plus	5.36	2.50	0.032
Hispanic*60 plus	-0.77	2.20	0.725
Asian*60 plus	6.54	4.07	0.108

Other*60 plus	-1.28	4.56	0.779
Education*Age			
College*40-59	-1.05	2.29	0.645
College*60 plus	-2.65	2.06	0.197
Gender*Race/ethnicity*Education			
Female*Black*College	-0.76	3.64	0.834
Female*Hispanic*College	3.35	3.40	0.324
Female*Asian*College	7.65	5.16	0.138
Female*Other*College	-2.83	6.23	0.649
Gender*Race/ethnicity*Age			
Female*Black*40-59	-3.92	4.04	0.332
Female*Hispanic*40-59	3.53	3.51	0.314
Female*Asian*40-59	7.69	5.84	0.188
Female*Other*40-59	4.54	7.67	0.554
Female*Black*60 plus	-3.84	3.73	0.304
Female*Hispanic*60 plus	0.24	3.24	0.941
Female*Asian*60 plus	-1.20	5.72	0.834
Female*Other*60 plus	0.95	7.39	0.898
Gender*Education*Age			
Female*College*40-59	-3.59	3.32	0.279
Female*College*60 plus	-2.12	3.02	0.483
Race/ethnicity*Education*Age			
Black*College*40-59	-5.08	3.61	0.160
Hispanic*College*40-59	2.87	3.54	0.417
Asian*College*40-59	3.61	4.77	0.449
Other*College*40-59	1.27	5.89	0.829
Black*College*60 plus	-2.15	3.39	0.526
Hispanic*College*60 plus	5.40	3.30	0.102
Asian*College*60 plus	0.78	4.88	0.874
Other*College*60 plus	4.51	6.28	0.472
Gender*Race/ethnicity*Education*Age			
Female*Black*College*40-59	4.40	5.10	0.389
Female*Hispanic*College*40-59	-1.37	4.87	0.779
Female*Asian*College*40-59	-6.02	6.68	0.368
Female*Other*College*40-59	-1.52	9.26	0.870
Female*Black*College*60 plus	2.60	4.86	0.592
Female*Hispanic*College*60 plus	-5.86	4.67	0.209
Female*Asian*College*60 plus	-3.17	6.82	0.642
Female*Other*College*60 plus	-5.01	9.43	0.595

2e. MAIHDA results

The MAIHDA model used to create the final intersection predictions is what is referenced in eTable 6 as the "Full model". The "Null model" is a model fitted with only random effects, and no fixed effects. The null model was only used to calculate estimates of discriminatory accuracy, the variance partition coefficient (VPC), and the proportional change in variance (PCV). The formulas for calculating VPC and PCV are as follows, where $\sigma^2_{u(0)}$ is between-stratum variance of the "null model", $\sigma^2_{u(0)}$ is within-stratum variance of the "null model", and $\sigma^2_{u(1)}$ is from the "full model":

$$\text{VPC} = \frac{\sigma^2_{u(0)}}{\sigma^2_{u(0)}+\sigma^2_{e(0)}} \text{ x } 100\% \qquad \text{PCV} = \frac{\sigma^2_{u(0)}-\sigma^2_{u(1)}}{\sigma^2_{u(0)}} \text{ x } 100\%$$

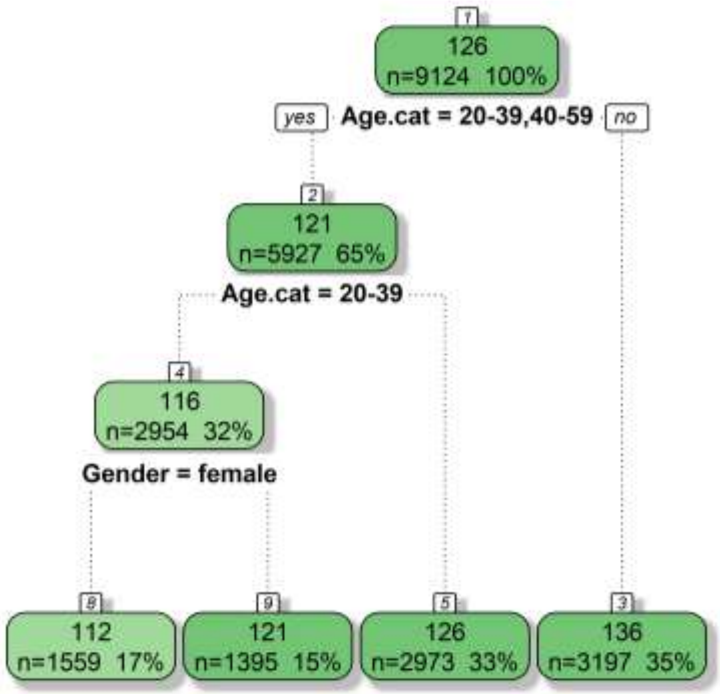
Note that results presented here are for coefficient estimates, but outcome estimates for each intersection, as presented in the main text, were calculated using the "predict" function in R.

eTable 7. MAIHDA results

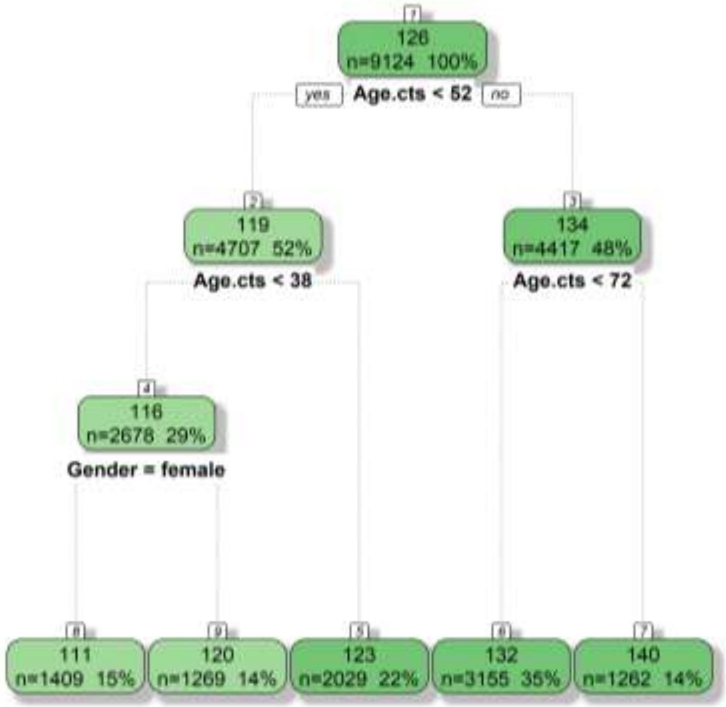
	MAIHDA unweighted		
Fixed Effects (Full model)	Est	95% CI	
Intercept	116.95	114.59	119.21
Gender (ref = Male)			
Female	-2.50	-4.30	-1.03
Race/ethnicity (ref = White)			
Black	5.60	3.28	7.95
Hispanic	0.40	-2.06	3.01
Asian	-0.01	-2.36	2.58
Other	1.61	-1.23	4.61
Education (ref = Highschool or less)			
College	-1.85	-3.44	-0.40
Age (ref = 20 to 39)			
40 to 59	9.81	7.89	11.95
60 plus	19.99	17.84	22.15
Random Effects (Null model)	Est		
Between-Stratum Variance	81.96		
Within-Stratum Variance	289.65		
VPC (%)	22.06%		
PCV (%)	90.10%		
Random Effects (Full model)	Est		
Between-Stratum Variance	8.11		
Within-Stratum Variance	289.66		
VPC (%)	2.72%		

2f. CART results

Outcome estimates for each intersection, as presented in the main text, were calculated for the model shown in eFigure3, using the “predict” function in R.



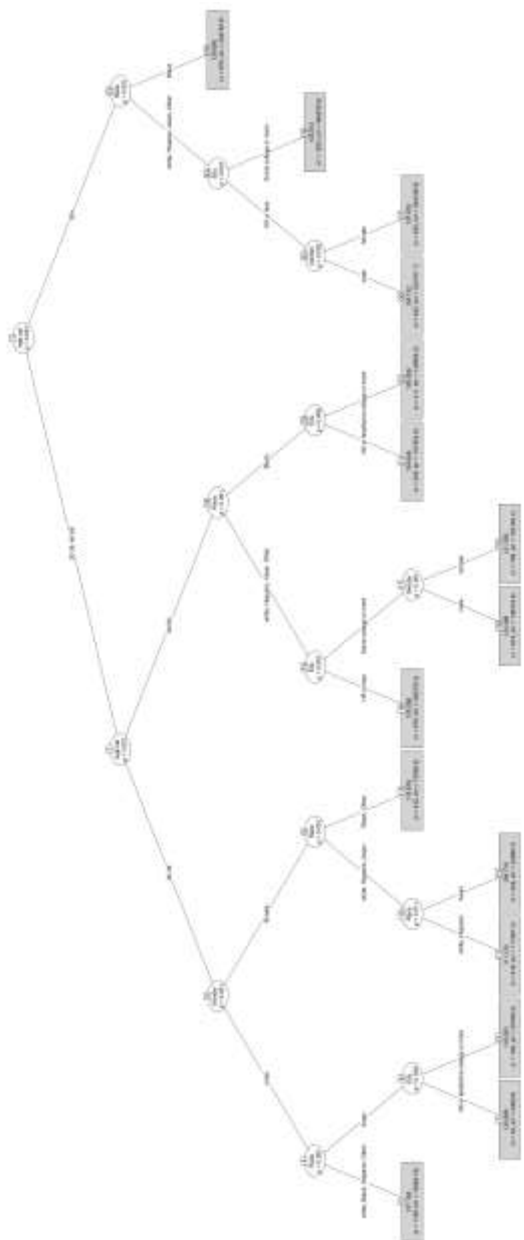
eFigure 3. CART model



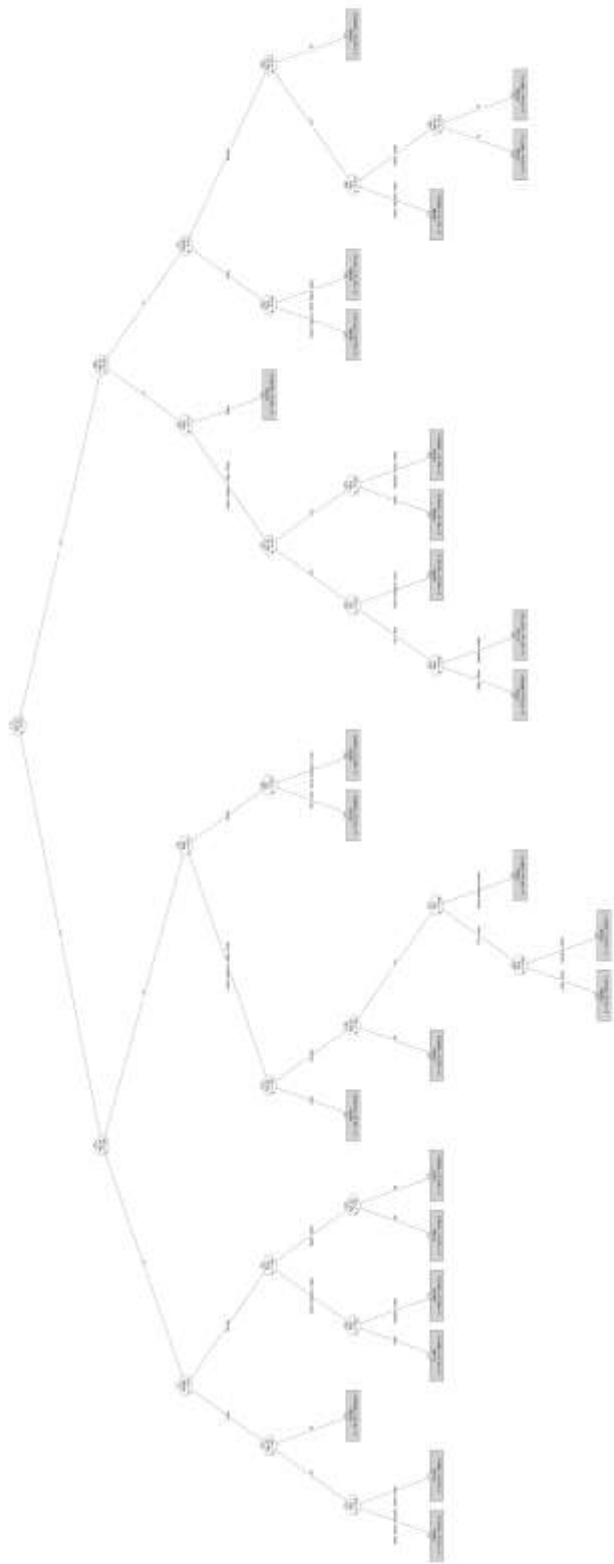
eFigure 4. CART model using continuous age variable

2g. CTree results

The CTree images below can be zoomed in to view the individual splitting patterns. For ease of interpretation, composition of the final subgroups created by the CTree model in Supplementary eFigure 5 has been summarized in table form. Outcome estimates for each intersection, as presented in the main text, were calculated for the model shown in eFigure5, using the “predict” function in R.



eFigure 5. CTree model



eFigure 6. CTree model using continuous age variable

eTable 8. CTree model subgroup results

Age	Gender	Race	Education	Mean SBP (mm Hg)
20-39	male	white, Black, Hispanic, other		121.183
20-39	male	Asian	HS or less	120.889
20-39	male	Asian	some college or more	116.093
20-39	female	white, Hispanic		111.125
20-39	female	Asian		108.716
20-39	female	Black, other		115.578
40-59		white, Hispanic, Asian, other	HS or less	125.288
40-59	male	white, Hispanic, Asian, other	some college or more	124.468
40-59	female	white, Hispanic, Asian, other	some college or more	121.293
40-59		Black	HS or less	134.646
40-59		Black	some college or more	130.059
60+		Black		139.850
60+		white, Hispanic, Asian, other	some college or more	133.012
60+	male	white, Hispanic, Asian, other	HS or less	134.132
60+	female	white, Hispanic, Asian, other	HS or less	137.679

eAppendix 3: Additional methods citations

For readers unfamiliar with the novel methods included in this study (CART, CTree, random forest, and MAIHDA), we provide a starter list of example applications and methods papers. Papers are labelled as follows:

[a] Methods citation

[b] Other methods papers

[c] Example analysis

MAIHDA:

- Evans CR, Williams DR, Onnela JP, Subramanian SV. A multilevel approach to modeling health inequalities at the intersection of multiple social identities. *Soc Sci Med* 2018;203:64-73. [a]
- Lizotte DJ, Mahendran M, Churchill SM, Bauer GR. Math versus meaning in MAIHDA: a commentary on multilevel statistical models for quantitative intersectionality. *Soc Sci Med* 2020;245:112500. [b]
- Bell A, Holman D, Jones K. Using shrinkage in multilevel models to understand intersectionality. *Methodology* 2019;15(2):88-96. [b]
- Evans CR, Erickson N. Intersectionality and depression in adolescence and early adulthood: a MAIHDA analysis of the national longitudinal study of adolescent to adult health, 1995–2008. *Soc Sci Med*. 2019 1;220:1-1. [c]
- Persmark A, Wemrell M, Zettermark S, Leckie G, Subramanian SV, Merlo J. Precision public health: Mapping socioeconomic disparities in opioid dispensations at Swedish pharmacies by Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy (MAIHDA). *PloS one*. 2019 27;14(8):e0220322.

CART:

- Breiman L, Friedman J, Olshen R, Stone C. Classification and regression trees. New York: Chapman and Hall/CRC, 1984. [a]
- Villanti AC, Gaalema DE, Tidey JW, Kurti AN, Sigmon SC, Higgins ST. Co-occurring vulnerabilities and menthol use in US Young adult cigarette smokers: findings from wave 1 of the PATH Study, 2013–2014. *Prev. Med.* 2018;117:43-51. [c]
- Cairney J, Veldhuizen S, Vigod S, Streiner DL, Wade TJ, Kurdyak P. Exploring the social determinants of mental health service use using intersectionality theory and CART analysis. *J Epidemiol Community Health.* 2014;68(2):145-50. [c]

CTree:

- Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. *J Comput Graph Stat* 2006;15(3):651-74. [a]
- Venkatasubramaniam A, Wolfson J, Mitchell N, Barnes T, JaKa M, French S. Decision trees in epidemiological research. *Emerg. Themes Epidemiol*. 2017 Dec 1;14(1):11. [b, c]

Random Forest:

- Breiman L. Random forests. *Machine learning*. 2001;45(1):5-32. [a]
- Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics* 2010;26(10):1340-7. [b]
- Nayak S, Hubbard A, Sidney S, Syme SL. A recursive partitioning approach to investigating correlates of self-rated health: The CARDIA Study. *SSM-population health.* 2018 1;4:178-88.