Association between agricultural production and food insecurity in GA, USA in 2017: ecological study

DOCUMENT: SAR-2024-004-HS-v01

From: Felipe Figueiredo **To:** Hamidah Sharif

2022-02-12

TABLE OF CONTENTS

Association between agricultural production and food insecurity in GA, USA in 2017: ecological study

Document version

Version	Alterations
01	Initial version

1 ABBREVIATIONS

- AIC: Akaike Information Criterion
- CI: confidence interval
- IQR: interquartile range
- SD: standard deviation

2 CONTEXT

2.1 Objectives

Assess the association between the vegetable productive area of Georgia rural counties and the proportion of people under food insecurity conditions in 2017.

3 METHODS

The data procedures, design and analysis methods used in this report are fully described in the annex document **SAP-2024-004-HS-v02**.

This analysis was performed using statistical software R version 4.3.2.

4 RESULTS

4.1 Status of Georgia counties in 2017

There were 3143 USA counties in the original database merged from all sources. After f those, 159 were Georgia counties. After the inclusion criterion was applied there were N = 108 rural counties included in the study sample.

The average proportion of females in counties in the study sample is 50.3% and the average proportion of people younger than 18 years old is 23.1%. Table 1 shows the characteristics of rural counties in Georgia in 2017 (or other years, where specified).

Table 1 *Status of Georgia counties in 2017.*

Characteristic	N = 108
% Food insecurity	
Median (IQR)	17.9 (14.6, 21.3)
Mean (SD)	18.0 (4.3)
Range	9.1, 29.8
Acres of vegetables harvested	
Median (IQR)	45 (17, 285)
Mean (SD)	981 (2,839)
Range	4, 17,770
(Missing)	22
% Low access to stores	
Median (IQR)	8 (4, 16)
Mean (SD)	12 (14)

Characteristic	**N = 108**
Range	0, 76
(Missing)	1
% Limited access to healthy foods	
Median (IQR)	3.8 (1.5, 6.8)
Mean (SD)	6.0 (8.0)
Range	0.0, 43.9
% Rural	
Median (IQR)	74 (62, 99)
Mean (SD)	77 (17)
Range	50, 100
% Poverty rate, 2015	
Median (IQR)	23 (18, 27)
Mean (SD)	23 (7)
Range	7, 42
% Child poverty rate, 2015	
Median (IQR)	35 (29, 40)
Mean (SD)	34 (9)
Range	9, 56

Characteristic	N = 108
Population estimate, 2015	
Median (IQR)	16,501 (9,073, 24,943)
Mean (SD)	18,686 (12,395)
Range	1,638, 63,038
Diabetes prevalence	
Median (IQR)	13.00 (11.95, 14.40)
Mean (SD)	13.14 (1.73)
Range	9.30, 17.40
Adult obesity	
Median (IQR)	31.50 (29.58, 33.33)
Mean (SD)	31.40 (2.58)
Range	25.90, 36.70
Poor or fair health	
Median (IQR)	19.5 (17.4, 22.7)
Mean (SD)	20.0 (4.1)
Range	11.8, 34.1
% Native Hawaiian	
Median (IQR)	0.06 (0.02, 0.15)

Characteristic	N = 108
Mean (SD)	0.12 (0.23)
Range	0.00, 2.01
% Hispanic	
Median (IQR)	4.0 (2.6, 6.1)
Mean (SD)	5.6 (5.0)
Range	1.1, 29.6
% Non-Hispanic White	
Median (IQR)	65 (56, 78)
Mean (SD)	65 (17)
Range	25, 95
% African American	
Median (IQR)	27 (11, 38)
Mean (SD)	27 (17)
Range	1, 72
% Asian	
Median (IQR)	0.70 (0.56, 0.93)
Mean (SD)	0.81 (0.50)
Range	0.22, 4.18

Characteristic	**N = 108**
% American Indian/Alaskan Native	
Median (IQR)	0.42 (0.33, 0.56)
Mean (SD)	0.50 (0.35)
Range	0.10, 3.09
% younger than 18	
Median (IQR)	22.19 (20.64, 24.84)
Mean (SD)	22.38 (3.18)
Range	13.55, 29.10
% Female	
Median (IQR)	50.8 (49.7, 51.7)
Mean (SD)	50.0 (3.2)
Range	34.8, 56.7

Two variables in the dataset had no variability, showing only a single value for the whole state. The % SNAP participants, 2016 had a value of 16.41 and the % WIC participants, 2015 had 2.59.

Figure 1 shows the exposures % Food insecurity against Acres of vegetables harvested and % Low access to stores. Both panels do not show an evident trend in the scatter. In order to visualize the scatter this graph shows the x axis in log-10 scale and this strategy was applied in all future steps of this analysis (see section 8.2 in the Appendix).

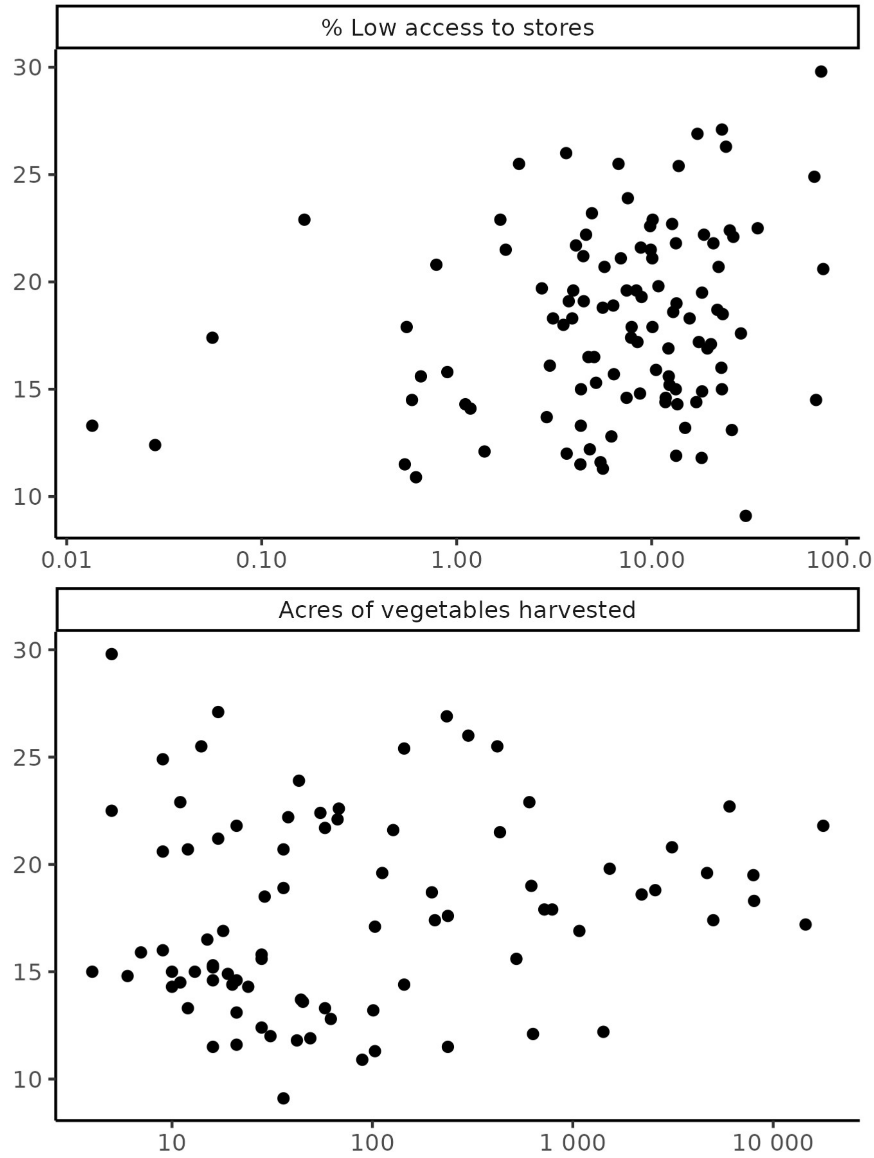

Figure 1 *Scatter plot of food insecurity against vegetable productive area and low access to stores.*

4.2 Food insecurity in rural Georgia

Five models were fit in the search for predictors of food insecurity. Models 1 to 3 were the best estimates to test the hypothesis of association with Acres of vegetables harvested and % Low access to stores. Model 1 is a crude estimate that includes only the exposures of interest above. Model 2 includes only covariates that do not correlate with the exposures and is used as an intermediate step before Model 3, that was subject to a stepwise selection of variables, under the constraint that the final specification kept the two exposures of interest (section 4.2.1).

Model 4 frees the stepwise algorithm to pick the best predictors in the dataset, even if the two exposures are included (section 4.2.2.1). This model specification was also tested in the full dataset to test the sensitivity of the analysis to the inclusion criterion (section 4.2.2.2).

4.2.1 Association with agricultural production and low access to stores

After removing missing values the complete cases dataset had N = 131 observations available for the analysis.

The simplest model evaluated (Model 1) had an AIC = 498.68 (coefficients are shown in section 8.2). This model explains 5.22% of the variance observed in food insecurity in rural counties. This crude estimate provides a very poor explanatory power.

Many predictors are correlated with the productive area, with the proportion of low access to stores, or with both (Figure 2). Poverty rate, child poverty rate, obesity, fair or poor health and proportion of Asians were significantly correlated with Acres of vegetables harvested. Limited access to healthy foods, population, Hispanic, White, African American, Asian, younger than 18 and female were significantly correlated with % Low access to stores.

After removing all variables correlated with either of the two exposures of interest, only Diabetes prevalence, % Native Hawaiian and % American Indian/Alaskan Native were added to Model 2 (AIC = 458.40). The inclusion of these additional predictors increased the explanatory power of the model to 42.92% of the variance in food insecurity (coefficients are shown in section 8.2). Both metrics show an improvement over the previous model, but this specification is still an intermediate step in the modeling strategy employed here.

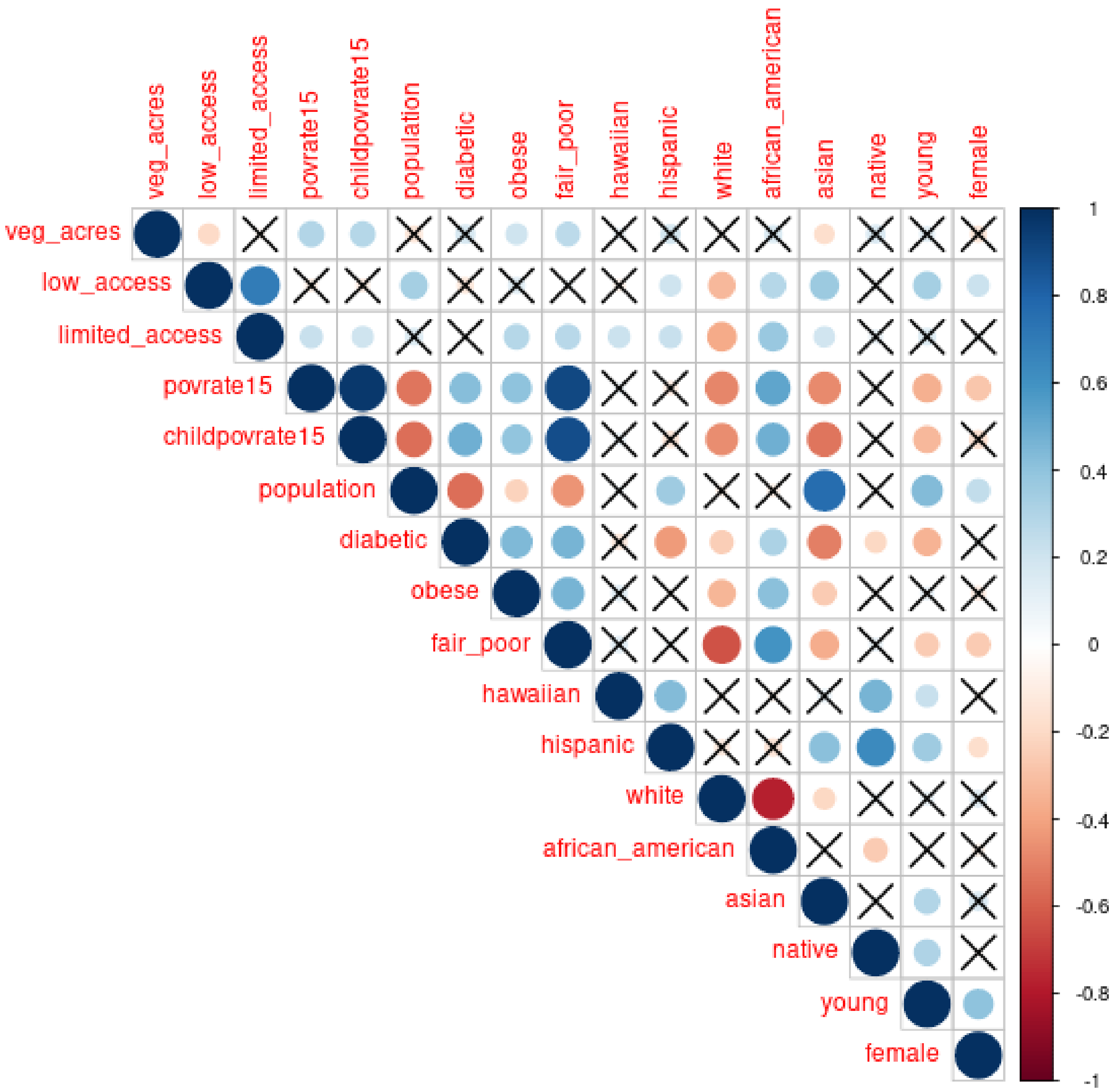

Figure 2 *Correlation matrix plot between predictors. Crosses indicate non-significant correlations.*

After applying the stepwise selection to this model, the proportion of Hawaiian natives was removed producing Model 3. This model has AIC = 458.19 and explains 42.44% of the variance in food insecurity. Since both metrics are similar for models 2 and 3, the simplest one is preferred and model 3 is shown in Table 2.

No association was found with Acres of vegetables harvested (β = 0.69, 95% CI: -0.15, 1.5), but there is a weak association with % Low access to stores. For each 10% increase in the proportion of low access to stores we can expect an increase of β = 1.4 percentage points in food insecurity (95% CI: 0.24, 2.5) This CI is wide and the lower confidence level is close to the null hypothesis, indicating poor evidence in the strength of this association (also shown in Figure 1). This significant CI could be a spurious association, fruit of remaining confounding. As seen in figure 2, limited access to healthy foods is positively correlated with low access to food and it could be a better predictor of food insecurity in this dataset.

Table 2 *Association with agricultural production and low access to stores.*

Characteristic	**Beta**[1]	**95% CI**[2]	**p-value**
Acres of vegetables harvested	0.69	-0.15 to 1.5	0.107
% Low access to stores	1.4	0.24 to 2.5	**0.018**
Diabetes prevalence	41	28 to 55	**<0.001**
% American Indian/Alaskan Native	-3.5	-7.1 to 0.18	0.062

[1]Stepwise, forcing exposures (adjusted R^2 = 42.44%)

[2]CI = Confidence Interval

Model 3 offers a minute decrease in the AIC and explains only a small fraction (less than 50%) of the proportion of food insecurity in rural counties so it is worthwhile to investigate better models. We will do that in the next sections.

4.2.2 Unconstrained predictors of food insecurity

4.2.2.1 Predictors in rural counties

Model 4 was constructed by applying the stepwise selection procedure to all log10-transformed covariates. This model selected a larger number of predictors like poverty rate, population, fair or poor health, a few ethnicity's, age and sex, but did not include any of the exposures of interest (Table 3). It shows a sharp decrease in the AIC metric indicating a better fit to the data compared with the previous models (AIC = 248.05, the best between all models for rural counties). This model explains 95.35% of the variance in food insecurity, which is also the best explanatory power among the models evaluated.

This makes it the best hypothesis for the prediction of factors associated with food insecurity in rural counties in 2017.

Table 3 *Unconstrained predictors of food insecurity in rural counties.*

Characteristic	Beta[1]	95% CI[2]	p-value
% Poverty rate, 2015	8.0	4.2 to 12	**<0.001**
Population estimate, 2015	1.6	0.69 to 2.5	**<0.001**
Poor or fair health	21	14 to 28	**<0.001**
% Hispanic	-3.5	-4.5 to -2.5	**<0.001**
% Non-Hispanic White	-14	-17 to -11	**<0.001**
% American Indian/Alaskan Native	-3.5	-4.7 to -2.3	**<0.001**
% younger than 18	3.0	-1.2 to 7.1	0.158
% Female	15	5.2 to 24	**0.003**

[1]Stepwise, free to remove exposures (adjusted R^2 = 95.35%)

[2]CI = Confidence Interval

No association was found with either productive area or low access to stores (Table 3). Each 10% increase in the proportion of white Americans is associated with a decrease of β=-14 percentage points in food insecurity (95% CI: -17, -11). Similarly, for each 10% increase in the poverty rate we could expect an increase of β=8.0 percentage points in food insecurity (95% CI: 4.2, 12), and a 10% increase in poor or fair health is associated with an increase of β=21 percentage points in food insecurity (95% CI: 14, 28).

4.2.2.2 Predictors in the full dataset

By applying the same specification of Model 4 to the full data we can assess how robust the analysis is to the selection of rural counties as opposed to the full state. Since this new model does not originate from the same data, an AIC comparison is not possible. This new model (Table 4) explains 94.00% of the variance in food insecurity.

Table 4 *Unconstrained predictors of food insecurity in all counties.*

Characteristic	Beta[1]	95% CI[2]	p-value
% Poverty rate, 2015	9.2	6.0 to 12	**<0.001**

Characteristic	Beta[1]	95% CI[2]	p-value
Population estimate, 2015	1.1	0.58 to 1.7	**<0.001**
Poor or fair health	21	15 to 28	**<0.001**
% Hispanic	-4.1	-5.1 to -3.2	**<0.001**
% Non-Hispanic White	-11	-14 to -9.2	**<0.001**
% American Indian/Alaskan Native	-3.4	-4.7 to -2.0	**<0.001**
% younger than 18	2.7	-1.5 to 6.9	0.205
% Female	12	2.6 to 22	**0.013**

[1]Full data, same specification as Model 3 (adjusted R^2 = 94.00%)

[2]CI = Confidence Interval

There are only minimal changes in both the estimates and the CI's, when compared with Table 4 so the results are robust to the choice of using rural counties in the analysis. This, combined with its similar adjusted R^2 of over 90%, reinforces the hypothesis that the best model found before has a good predictive power to explain food insecurity in Georgia, 2017.

5 OBSERVATIONS AND LIMITATIONS

Mismatch in data collection periods

Although most of the data used in this analysis reflect the status of counties in 2017, some variables were only available from either 2015 or 2016. It is unknown how much change would be expected if all measurements were respective of the same year. There is risk of bias of information, since the data was not measured at the same time. This limitation should be taken into account when using the results.

Recommended reporting guideline

The EQUATOR network reporting guidelines (http://www.equator-network.org/) have seen increasing adoption by scientific journals. All observational studies are recommended to be reported following the STROBE guideline (von Elm et al, 2014).

6 CONCLUSIONS

- There is no association between exposure to larger area of vegetable production and food insecurity in rural counties;
- There is weak evidence of association between exposure to low access to stores and food insecurity in rural counties;
- There is strong evidence of association with other predictors that are correlated with these exposures in rural counties;
- This strong association is also robust to the choice of rural counties vs the entire state with 2017 data.

7 REFERENCES

- **SAP-2024-004-HS-v02** – Analytical Plan for Association between agricultural production and food insecurity in GA, USA in 2017: ecological study
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg. 2014 Dec;12(12):1495-9 (https://doi.org/10.1016/j.ijsu.2014.07.013).

8 APPENDIX

8.1 Exploratory data analysis

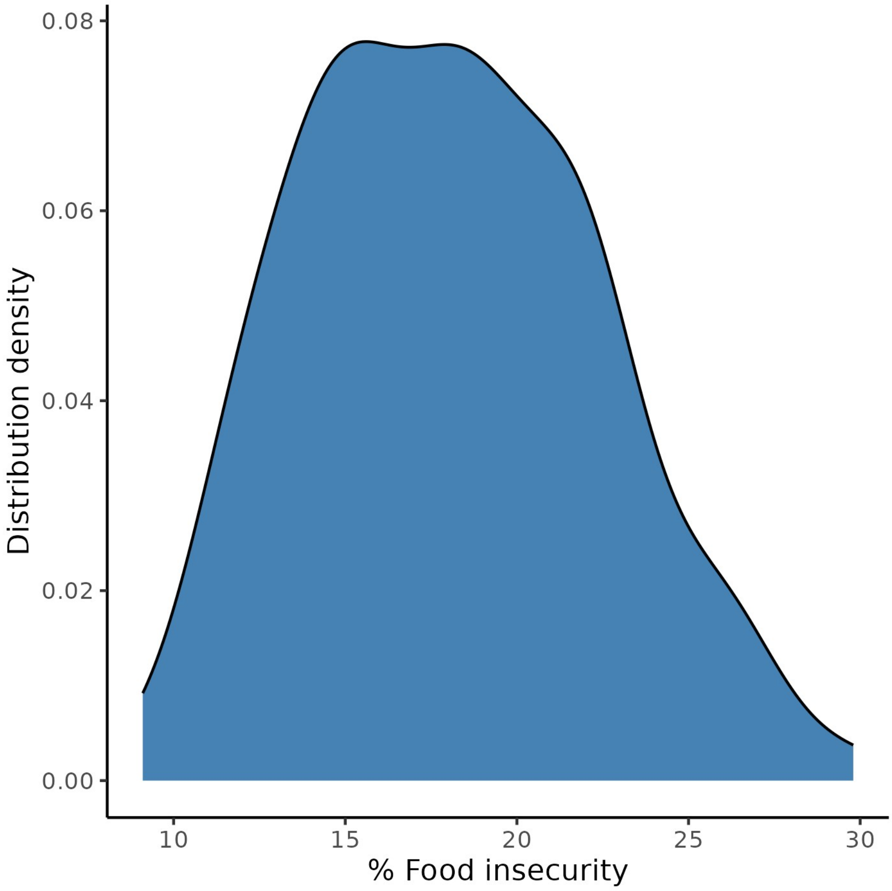

Figure A1 *Distribution density of food insecurity.*

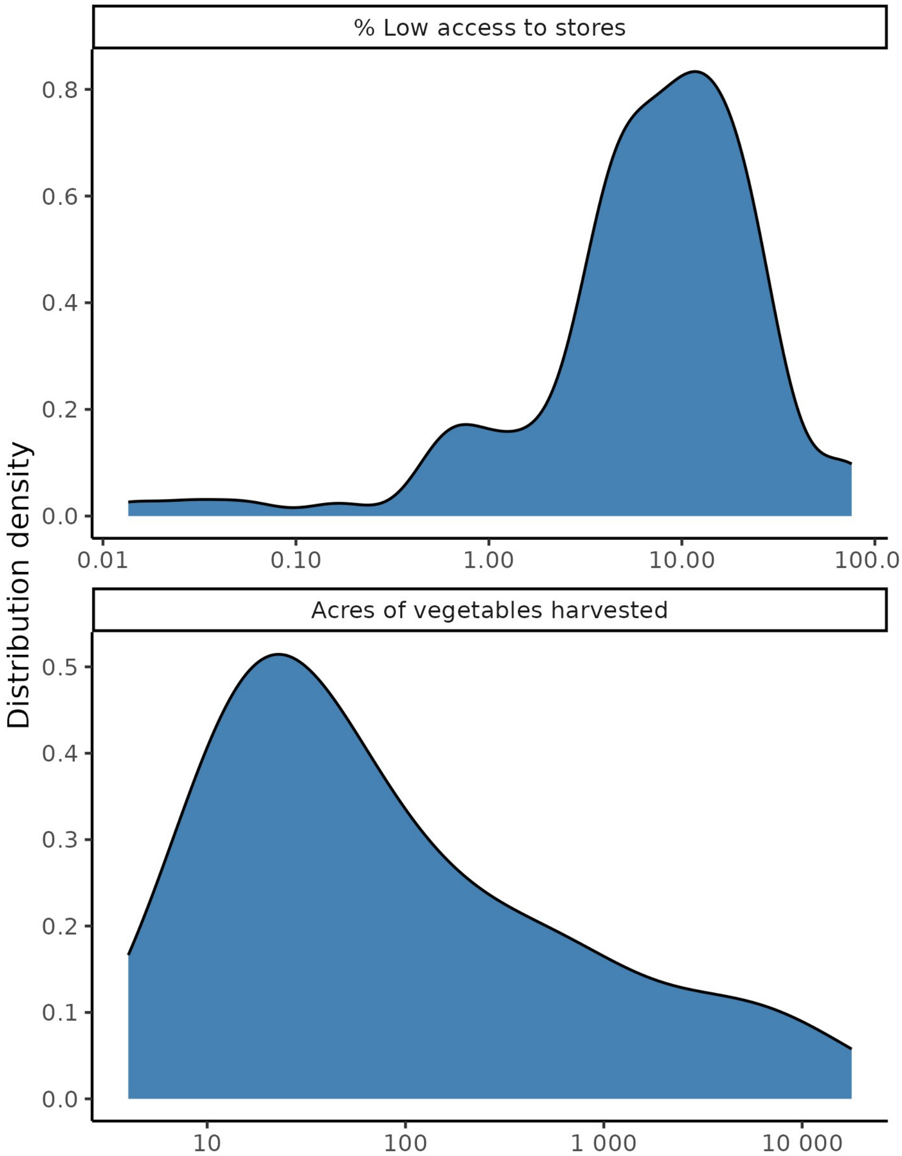

Figure A2 *Distribution density of the exposures.*

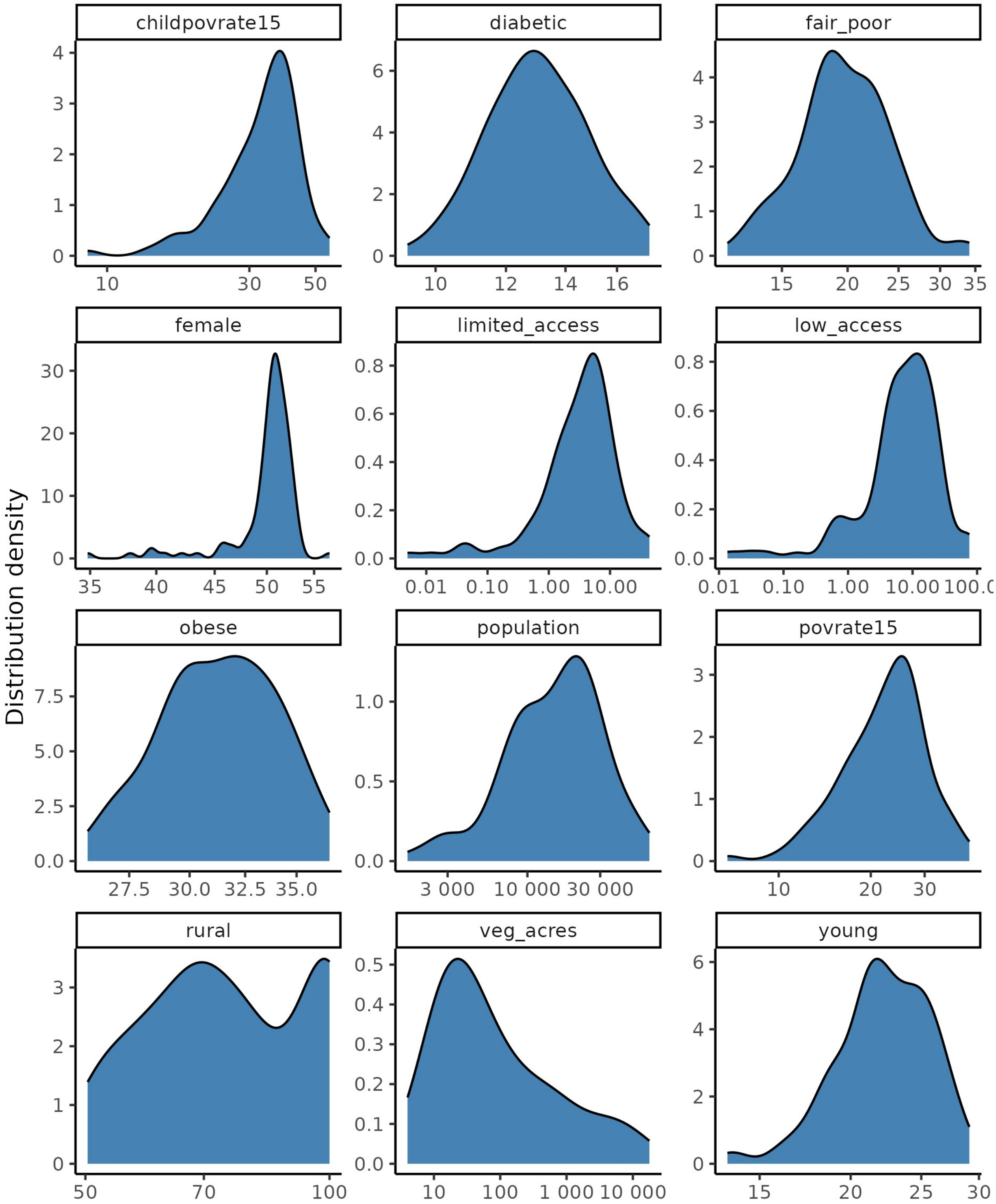

Figure A3 *Distribution density of the non-ethnicity predictors.*

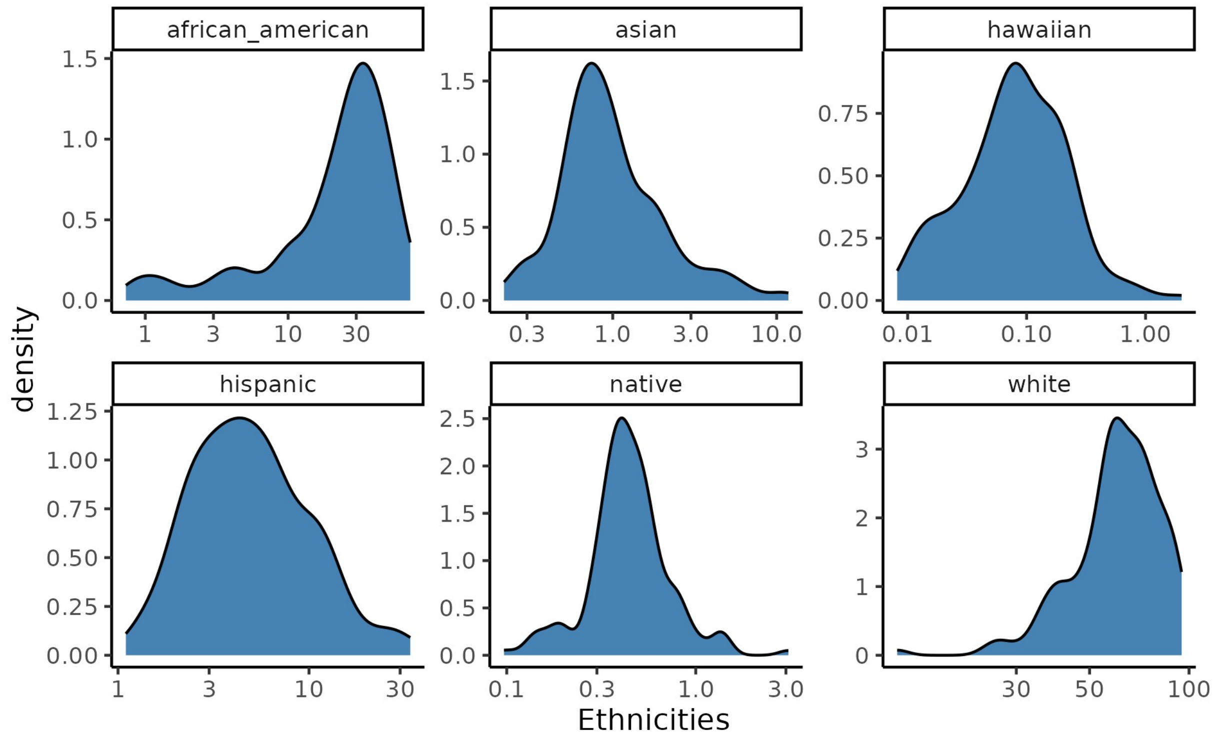

Figure A4 *Distribution density of ethnicity.*

8.2 Modeling strategy

A log-10 transformation was applied in all predictors, with the exceptions of % Limited access to healthy foods and % Native Hawaiian. Both of these predictors had some nil values, and their log transformation would be undefined. As such a shift-by-1 operation was used in these variables to avoid non-numeric results and keep all data points present in the analysis.

Table A1 *Estimates from all models.*

	Model 1			Model 2			Model 3			Model 4			Model 5		
Characteristic	**Beta[1]**	**95% CI[2]**	**p-value**	**Beta[3]**	**95% CI[2]**	**p-value**	**Beta[4]**	**95% CI[2]**	**p-value**	**Beta[5]**	**95% CI[2]**	**p-value**	**Beta[6]**	**95% CI[2]**	**p-value**
Acres of vegetables harvested	0.65	-0.39 to 1.7	0.219	0.70	-0.13 to 1.5	0.099	0.69	-0.15 to 1.5	0.107						
% Low access to stores	1.6	0.19 to 3.0	**0.027**	1.5	0.37 to 2.6	**0.010**	1.4	0.24 to 2.5	**0.018**						
Diabetes prevalence				42	28 to 55	**<0.001**	41	28 to 55	**<0.001**						

	Model 1			Model 2			Model 3			Model 4			Model 5		
Characteristic	**Beta**[1]	**95% CI**[2]	**p-value**	**Beta**[3]	**95% CI**[2]	**p-value**	**Beta**[4]	**95% CI**[2]	**p-value**	**Beta**[5]	**95% CI**[2]	**p-value**	**Beta**[6]	**95% CI**[2]	**p-value**
% Native Hawaiian				8.5	-4.5 to 21	0.199									
% American Indian/Alaskan Native				-4.8	-9.0 to -0.63	**0.025**	-3.5	-7.1 to 0.18	0.062	-3.5	-4.7 to -2.3	**<0.001**	-3.4	-4.7 to -2.0	**<0.001**
% Poverty rate, 2015										8.0	4.2 to 12	**<0.001**	9.2	6.0 to 12	**<0.001**
Population estimate, 2015										1.6	0.69 to 2.5	**<0.001**	1.1	0.58 to 1.7	**<0.001**
Poor or fair health										21	14 to 28	**<0.001**	21	15 to 28	**<0.001**
% Hispanic										-3.5	-4.5 to -2.5	**<0.001**	-4.1	-5.1 to -3.2	**<0.001**
% Non-Hispanic White										-14	-17 to -11	**<0.001**	-11	-14 to -9.2	**<0.001**
% younger than 18										3.0	-1.2 to 7.1	0.158	2.7	-1.5 to 6.9	0.205
% Female										15	5.2 to 24	**0.003**	12	2.6 to 22	**0.013**

[1]Crude estimate (adjusted R^2 = 5.22%)

[2]CI = Confidence Interval

[3]Predictors selected based on correlations (adjusted R^2 = 42.92%)

[4]Stepwise, forcing exposures (adjusted R^2 = 42.44%)

[5]Stepwise, free to remove exposures (adjusted R^2 = 95.35%)

[6]Full data, same specification as Model 3 (adjusted R^2 = 94.00%)

Table A2 *AIC of all models.*

model	df	AIC
Model 1	4	498.68
Model 2	7	458.40
Model 3	6	458.19
Model 4	10	248.05

8.3 Availability

All documents from this consultation were included in the consultant's Portfolio.

The portfolio is available at:

https://philsf-biostat.github.io/SAR-2024-004-HS/

8.4 Analytical dataset

Table A3 shows the structure of the analytical dataset.

Table A3 *Analytical dataset structure*

county	food_insecure	veg_acres	low_access	limited_access	rural	is_rural	snap16	wic15	povrate15	childpovrate15	population	diabetic	obese	fair_poor	hawaiian	hispanic	white	african_american	asian	native	young	female
1																						
2																						
3																						
…																						
108																						

Due to confidentiality the data-set used in this analysis cannot be shared online in the public version of this report.