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

DOCUMENT: SAP-2024-004-HS-v01

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

2024-02-02

TABLE OF CONTENTS

Analytical Plan for 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 counties and the proportion of people under food insecurity conditions in 2017.

2.2 Hypotheses

1. The proportion of people in food insecurity situation is associated with the total area of vegetable production in Georgia counties in 2017.
2. The proportion of people in food insecurity situation is associated with the limited access to food markets in Georgia counties in 2017.

3 DATA

3.1 Raw data

County data was gathered from seven databases (food environment assistance, food environment access, food environment socioeconomic, food environment supplemental, health rankings ranked, health rankings additional and NASS). The first six databases had FIPS, state and county as keys, while NASS only has state and county. Given that the FIPS code uniquely identifies counties across the entire country, this was used as the main key for joining the first six tables, followed by a join by state and county with the NASS database.

This merging procedure produced the original raw data table, with 375 variables collected on 3143 counties.

3.2 Analytical dataset

The cleaning process selected only the variables relevant to the analysis, and only the records for counties in the state of Georgia. After the cleaning process 22 variables were included in the analysis with data from 159 counties.

The total number of observations excluded due to incompleteness and exclusion criteria will be reported in the analysis. Table 1 shows the structure of the analytical dataset.

Table 1 *Analytical dataset structure after variable selection and cleaning.*

county	food_insecure	veg_acres	low_access	limited_access	rural	snap16	wic15	povrate15	childpovrate15	population	diabetic	obese	fair_poor	hawaiian	hispanic	white	african_american	asian	native	young	female
1																					
2																					
3																					
…																					
159																					

All variables in the analytical set were labeled according to the raw data provided and values were labeled according to the data dictionary for the preparation of production-quality results tables and figures.

4 STUDY PARAMETERS

4.1 Study design

Ecological design, with aggregate data at the county level.

Not all data reflect counties status in 2017 (see Observations and Limitations).

4.2 Inclusion and exclusion criteria

N/A

4.3 Exposures

1. Acres of vegetables harvested
2. % Low access to stores

4.4 Outcomes

Specification of outcome measures (Zarin, 2011):

1. (Domain) Food insecurity
2. (Specific measurement) Proportion of food insecure in 2017
3. (Specific metric) End value
4. (Method of aggregation) Mean

Primary outcome

Percentage of food insecurity in the county, in 2017.

4.5 Covariates

1. % Limited access to healthy foods
2. % Rural
3. % Poverty rate, 2015
4. % Child poverty rate, 2015
5. Population estimate, 2015
6. Diabetes prevalence
7. Adult obesity
8. Poor or fair health
9. % Native Hawaiian
10. % Hispanic
11. % Non-Hispanic White
12. % African American
13. % Asian
14. % American Indian/Alaskan Native
15. % Female

16. % younger than 18
17. % SNAP participants, 2016
18. % WIC participants, 2015

5 STATISTICAL METHODS

5.1 Statistical analyses

5.1.1 Descriptive analyses

Demographic and clinical variables will be described as mean (SD) and median (IQR) or as counts and proportions (%), as appropriate. The distributions of counties' characteristics will be summarized in tables and visualized in exploratory plots.

5.1.2 Inferential analyses

A correlation matrix, with Pearson correlation tests, will be calculated to assess if covariates (section 4.5) are correlated to each other. These correlation tests will be used as a decision criteria to remove variables from the candidate pool of covariates for the statistical analysis (described in the next section).

5.1.3 Statistical modeling

To assess an association between the proportion of food insecurity (section 4.4) and the exposures (acres and proportion of low access to stores) (section 4.3) a multiple linear regression model will be fit to the data.

Given the large number of covariates to consider for inclusion a stepwise selection algorithm will be used to find the set of predictors that best fits the data, according to the adjusted R^2. It is intended that this stepwise selection approach be used in two ways: both exposures will be kept in the first run, and other predictors will be included automatically by the algorithm; in the second run the algorithm will be free to remove the exposures if they don't fit the data well.

All models found will be compared with a goodness of fit test in an analysis of variance, ranked by AIC and the final model will be selected based on the ranking.

5.1.4 Missing data

No missing data imputation will be performed. All evaluations will be performed as complete case analyses. Missing data counts and proportions will be reported in tables.

5.2 Significance and Confidence Intervals

All analyses will be performed using the significance level of 5%. All significance hypothesis tests and confidence intervals computed will be two-tailed.

5.3 Study size and Power

N/A

5.4 Statistical packages

This analysis will be performed using statistical software R version 4.3.2.

6 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).

7 REFERENCES

- **SAR-2024-004-HS-v01** – Association between agricultural production and food insecurity in GA, USA in 2017: ecological study
- Zarin DA, et al. The ClinicalTrials.gov results database – update and key issues. N Engl J Med 2011;364:852-60 (https://doi.org/10.1056/NEJMsa1012065).
- Gamble C, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337–2343 (https://doi.org/10.1001/jama.2017.18556).
- 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

This document was elaborated following recommendations on the structure for Statistical Analysis Plans (Gamble, 2017) for better transparency and clarity.

8.1 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/