Paper PO-10-2014

Modifying a structural equation model of child dietary intake using the Lagrange Multiplier test in SAS® PROC CALIS for MWSUG 2014

Lauren Cook, University of Southern California, Los Angeles, CA

Chih-Ping Chou, University of Southern California, Los Angeles, CA

Nicole Gatto, MPH, PhD, Loma Linda University, Loma Linda, CA

Jaimie N Davis, RD, PhD, University of Texas, Austin, TX

Donna Spruijt-Metz, MFA, PhD, University of Southern California, Los Angeles, CA

ABSTRACT

Structural equation models (SEM) allow for simultaneous evaluation of causal pathways between multiple predictors and outcomes, including latent variables, given a hypothesized theoretical model. If an initial SEM does not meet appropriate fit criteria, model modification can achieved through utilization of statistical results from the Lagrange Multiplier (LM) test and theoretical knowledge.

To demonstrate modification of a SEM, baseline data from a garden-based nutrition intervention with elementary school children were examined using SAS® PROC CALIS. The two outcome variables of interest are fruit and vegetable intake (individually). Several measured determinants are hypothesized to predict these practices, including two latent variables, willingness to try fruit and vegetables, which have six indicator variables each.

Initial model fit was not acceptable (χ^2: 407.9, 111 df, p<0.0001; RMSEA: 0.09, CFI: 0.89). From the LM statistic we identified several unaccounted for correlations between errors on indicators of willingness to try fruits and vegetables. These correlations are theoretically sensible given that items on these scales are worded identically, save for the words “fruit” and “vegetable”, and these modifications were made. Once correlations between comparable indicators were added to the SEM, model fit was improved (χ^2: 234.0, 105 df, p<0.0001; RMSEA: 0.06, CFI: 0.95). Additional relationships identified by the LM test were evaluated, but none were theoretically meaningful, and the second model was accepted as final.

Use of the LM test in PROC CALIS facilitates theoretically appropriate modification of an a priori SEM in order to improve overall model fit and produce more reliable parameter estimates.

INTRODUCTION

Structural equation modeling (SEM) is an analytical approach that allows for exploration of complex multivariate relationships. There are two primary benefits of this technique: 1) latent factors can be observed from knowledge of indicator variables, as with factor analysis, and 2) multiple relationships between independent and dependent variables can be simultaneously observed.

As with general linear regression, checks for model validity are an essential part of the modeling process. Unlike linear regression, statistical tests are performed to ensure the model meets criteria for acceptable fit. When overall model fit does not meet standards, the Lagrange Multiplier (LM) test can be employed to identify parameters that can be added to the model (either directional pathways or covariances between variables, factors or error terms) in order to improve fit. However, given its post-hoc nature, the LM test is a somewhat controversial approach, so must be used in conjunction with theoretical relevance.

This paper will demonstrate the use of the LM test to improve model fit in SEM using PROC CALIS. The specific application will be in exploring the determinants of fruit and vegetable intake and preferences in elementary school children, as part of a nutrition, cooking and gardening intervention for obesity prevention. [1,2]

VARIABLES

One structural model will include variables specific to fruits, and variables specific to vegetables (all data are cross-sectional). The model is structured in this way because fruits and vegetables have varying determinants,[3] and thus the strength of relationships may differ between predictors of fruit versus vegetables intake and preferences. Also, there may be some effects of fruit –related variables on vegetable- related variables, and vice versa. The initial hypothesized model is included in **Figure 1**.

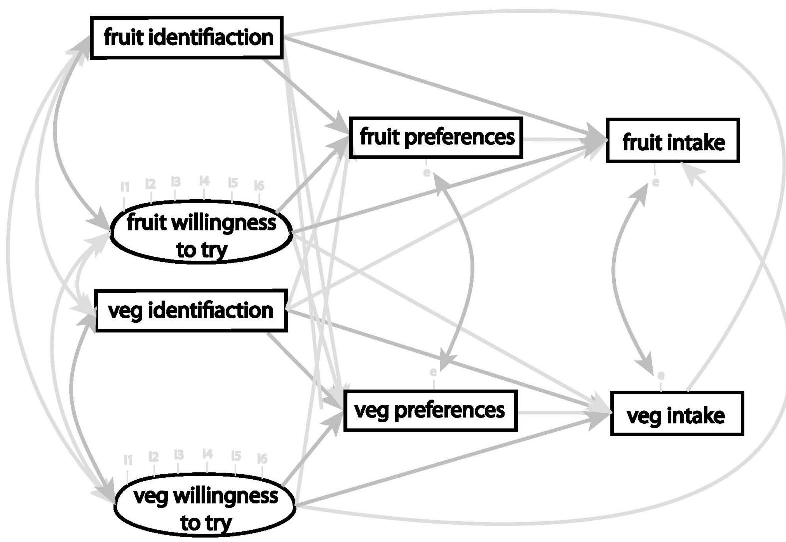

Figure 1: Initial hypothesized model

LIST OF VARIABLES

Fruit intake and **vegetable intake**: These variables are measured via the Block Kids Food Screener (last week version),[4] and intake of each of these food types is provided in cup equivalents.

Fruit identification and **vegetable identification**: Students were given a list of 8 fruit items and 17 vegetable items,[5] and were asked if they knew these items, or not. A standardized sum (accounting for unanswered questions) was obtained, with a range of 6-8 items answered about fruits, and 11-17 items answered about vegetables.

Fruit preferences and **vegetable preferences**: For each item identified from the above question, students were asked if they liked this food “A lot”, “A little”, or “Not at all”. Mean scores were used for fruit items and vegetable items (factor analysis was not used due to the large number of indicators).

Willingness to try fruit and **willingness to try vegetables**: The willingness to try scale[6] has six items each for fruits and vegetables, with questions such as “How much do you like tasting new vegetables?” and responses ranging from “Not at all” to “A lot”. The variables for the six items in each scale (specific to fruits or vegetables) served as indicators for a latent factor, either willingness to try fruits or willingness to try vegetables, respectively.

DATA ANALYSIS

DATA PREPARATION

Data appeared to be missing at random, so to account for missingness, the correlation matrix was analyzed instead of raw data. The following code was used to perform this task:

```
proc corr data=las2 out=las_corr nosimple noprob;
      var sex ethnicity bmiz age fruit_intake veg_intake fruit_id
            veg_id veg_willing_1 veg_willing_2 veg_willing_3 veg_willing_4
            veg_willing_5  veg_willing_6 fruit_willing_1 fruit_willing_2
            fruit_willing_3 fruit_willing_4 fruit_willing_5 fruit_willing_6
            fruit_pref veg_pref;
run;
```

MODEL 1 SPECIFICATION

The code below was used to fit the initial hypothesized model. In the PROC line we specify that the data form is a correlation matrix, specify that we would like to use maximum likelihood estimation, and the ‘mod’ command indicates we would also like to output the results of the LM test.

Following the ‘lineqs’ statement, each equation represents a path we would like to include in the model. Terms such as ‘b1’ give a name to the parameter to reference in the output, and each equation also includes an error term. The terms ‘f1’ and ‘f2’ refer to our two latent factors, willingness to try fruits and willingness to try vegetables, and note these must be indicated in this form (rather than creating a meaningful factor name, such as ‘fruit_willingness_factor’; it is acceptable in CALIS to use original variable names rather than renaming as ‘v1’, etc. for those included in the database).

In the ‘variance’ statement (or ‘std’, which performs the same function), we indicate the parameters of which we want to estimate the variance. Similarly in the ‘cov’ statement, we indicate the relationships of which we would like to estimate the covariances, and we indicate a name for this parameter. We can name parameters in the ‘variance’ statement, as well, but this is not necessary. Note that for dependent variables (fruit and vegetable preferences and intake), we estimate the covariances of the error terms, not the variables themselves.

```
proc calis corr data=las_corr method=ml mod;

lineqs
fruit_pref = b1 fruit_id + b2 veg_id + b3 f1 + b4 f2 + e1,
veg_pref = b5 fruit_id + b6 veg_id + b7 f1 + b8 f2 + e2,
fruit_willing_1 = a1 f1 + e10,
fruit_willing_2 = a2 f1 + e11,
fruit_willing_3 = a3 f1 + e12,
fruit_willing_4 = a4 f1 + e13,
fruit_willing_5 = a5 f1 + e14,
fruit_willing_6 = a6 f1 + e15,
veg_willing_1 = a7 f2 + e16,
veg_willing_2 = a8 f2 + e17,
veg_willing_3 = a9 f2 + e18,
veg_willing_4 = a10 f2 + e19,
veg_willing_5 = a11 f2 + e20,
veg_willing_6 = a12 f2 + e21,
fruit_intake = b9 fruit_id + b10 veg_id + b11 f1 + b12 f2
       + b13 fruit_pref + b14 veg_pref + e3,
veg_intake = b15 fruit_id + b16 veg_id + b17 f1 + b18 f2
       + b19 fruit_pref + b20 veg_pref + e4;

variance
e1, e2, e3, e4, e10, e11, e12, e13, e14,
e15, e16, e17, e18, e19, e20, e21;

cov
fruit_id veg_id= theta1,
f1 f2 = theta2,
fruit_id f1 = theta3,
veg_id f2 = theta4,
fruit_id f2 = theta5,
veg_id f1 = theta6,
e1 e2 = theta7,
e3 e4 = theta8;

run;
```

MODEL 1 OUTPUT

The following output summarizes the fit of this model. Fit of this model is not acceptable (we would like to see a value >0.95 for the Bentler Comparative Fit Index (CFI), and a value <0.05 for the Root Mean Square Error of Approximation (RMSEA) Index). Therefore, we look to the LM test results.

Fit Summary		
Modeling Info	N Observations	350
	N Variables	18
	N Moments	171
	N Parameters	60
	N Active Constraints	0
	Baseline Model Function Value	7.8962
	Baseline Model Chi-Square	2755.7718
	Baseline Model Chi-Square DF	153
	Pr > Baseline Model Chi-Square	<.0001
Absolute Index	Fit Function	1.1687
	Chi-Square	407.8929
	Chi-Square DF	111
	Pr > Chi-Square	<.0001
	Z-Test of Wilson & Hilferty	12.1838
	Hoelter Critical N	117
	Root Mean Square Residual (RMSR)	0.0500
	Standardized RMSR (SRMSR)	0.0500
	Goodness of Fit Index (GFI)	0.8678
Parsimony Index	Adjusted GFI (AGFI)	0.7963
	Parsimonious GFI	0.6296
	RMSEA Estimate	0.0875
	RMSEA Lower 90% Confidence Limit	0.0785
	RMSEA Upper 90% Confidence Limit	0.0967
	Probability of Close Fit	<.0001
	ECVI Estimate	1.5324
	ECVI Lower 90% Confidence Limit	1.3626
	ECVI Upper 90% Confidence Limit	1.7252
	Akaike Information Criterion	527.8929
	Bozdogan CAIC	819.3689
	Schwarz Bayesian Criterion	759.3689
	McDonald Centrality	0.6543
Incremental Index	Bentler Comparative Fit Index	0.8859
	Bentler-Bonett NFI	0.8520
	Bentler-Bonett Non-normed Index	0.8428
	Bollen Normed Index Rho1	0.7960
	Bollen Non-normed Index Delta2	0.8877
	James et al. Parsimonious NFI	0.6181

Display 1: Selected output from first structural equation model (fit statistics)

Results of the LM test are included below. The first two tables indicate additional pathways that can be created between observed variables (distinguishing them such that the first table includes both 'predictor' and 'outcome' variables, and the second table only includes 'predictor' variables (ie, identification and willingness variables only); technically our willingness indicators are dependent on the willingness factor), and ranks them by magnitude of change to model fit. However, inclusion of any of these pathways indicated in the first two tables is not desirable. Since we hypothesize a relationship between the willingness to try factors and preferences and intake, it would not make sense to include pathways in the model directly linking the factor indicators with those outcome variables.

The third table shows additional covariances between errors that could be added to the model, again ranked by magnitude. From examining these, we see that some of these pathways make sense. For example, 'e21' and 'e15' both refer to the sixth item on the willingness to try scales, and these questions are worded identically, with the exception of the words 'fruits' or 'vegetables'. The same is true for 'e10' and 'e16', and 'e20' and 'e14'.

Rank Order of the 10 Largest LM Stat for Paths from Endogenous Variables				
To	From	LM Stat	Pr > ChiSq	Parm Change
veg_willing_1	veg_pref	80.57136	<.0001	2.53750
veg_willing_2	fruit_pref	56.52670	<.0001	-1.86176
veg_willing_5	veg_intake	50.16653	<.0001	-0.60705
veg_willing_1	fruit_pref	26.50015	<.0001	0.45302
fruit_willing_1	fruit_pref	15.62408	<.0001	0.32701
veg_willing_6	fruit_pref	5.09153	0.0240	0.23337
veg_willing_6	veg_intake	3.14364	0.0762	0.13432
veg_willing_5	fruit_intake	3.14335	0.0762	-0.06888
fruit_willing_1	fruit_intake	1.90042	0.1680	0.06702
veg_willing_3	veg_intake	1.88023	0.1703	0.08664

Rank Order of the 7 Largest LM Stat for Paths from Exogenous Variables				
To	From	LM Stat	Pr > ChiSq	Parm Change
fruit_willing_2	fruit_id	19.18740	<.0001	1.80624
fruit_willing_4	fruit_id	5.41341	0.0200	-0.22117
veg_willing_1	fruit_id	3.87754	0.0489	0.15839
veg_willing_4	fruit_id	0.32662	0.5677	-0.03983
veg_willing_2	fruit_id	0.17771	0.6733	0.08149
fruit_willing_1	fruit_id	0.00521	0.9424	0.00511
veg_willing_6	fruit_id	0.0003511	0.9851	-0.00284

Rank Order of the 10 Largest LM Stat for Error Variances and Covariances				
Var1	Var2	LM Stat	Pr > ChiSq	Parm Change
e12	e10	562.26681	<.0001	5.70035
e21	e15	87.76411	<.0001	0.35950
e16	e10	76.56286	<.0001	0.31856
e14	e1	68.70051	<.0001	-0.69057
e20	e12	66.66359	<.0001	-0.54926
e14	e10	56.00892	<.0001	-0.65302
e20	e14	51.67631	<.0001	0.33683
e16	e14	37.46651	<.0001	-0.23992
e19	e13	36.23283	<.0001	0.22154
e13	e10	32.74577	<.0001	-0.31696

Display 2: Selected output from first structural equation model (Lagrange Multiplier test)

MODEL 2 SPECIFICATION

To modify the model, we add six additional covariance parameters to the model, such that correlations are included between comparable items on the fruit and vegetable willingness to try subscales (**Figure 2**).

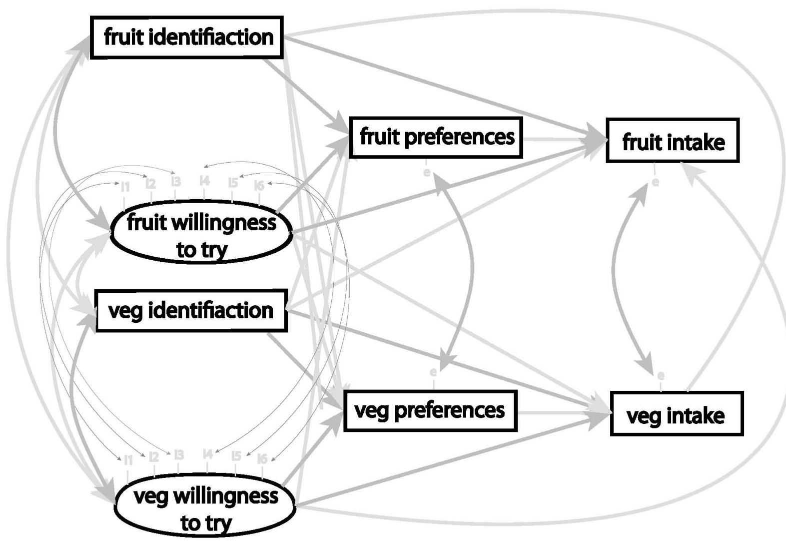

Figure 2: Modified structural model

The revised code is below, with the additional lines to the ‘cov’ statement being the only modifications made.

```
proc calis corr data=las_corr method=ml mod;
lineqs
fruit_pref = b1 fruit_id + b2 veg_id + b3 f1 + b4 f2 + e1,
veg_pref = b5 fruit_id + b6 veg_id + b7 f1 + b8 f2 + e2,
fruit_willing_1 = a1 f1 + e10,
fruit_willing_2 = a2 f1 + e11,
fruit_willing_3 = a3 f1 + e12,
fruit_willing_4 = a4 f1 + e13,
fruit_willing_5 = a5 f1 + e14,
fruit_willing_6 = a6 f1 + e15,
veg_willing_1 = a7 f2 + e16,
veg_willing_2 = a8 f2 + e17,
veg_willing_3 = a9 f2 + e18,
veg_willing_4 = a10 f2 + e19,
veg_willing_5 = a11 f2 + e20,
veg_willing_6 = a12 f2 + e21,
fruit_intake = b9 fruit_id + b10 veg_id + b11 f1
       + b12 f2 + b13 fruit_pref + b14 veg_pref + e3,
veg_intake = b15 fruit_id + b16 veg_id + b17 f1
       + b18 f2 + b19 fruit_pref + b20 veg_pref + e4;

variance
e1, e2, e3, e4, e10, e11, e12, e13, e14,
e15, e16, e17, e18, e19, e20, e21;

cov
fruit_id veg_id= theta1,
f1 f2 = theta2,
fruit_id f1 = theta3,
veg_id f2 = theta4,
fruit_id f2 = theta5,
veg_id f1 = theta6,
e1 e2 = theta7,
e3 e4 = theta8,
e10 e16 = theta9,
e15 e21 = theta10,
e13 e19 = theta11,
```

```
e11 e17 = theta12,
e12 e18 = theta13,
e14 e20 = theta14;
run;
```

MODEL 2 OUTPUT

This code gives us the following output for fit statistics. We see that the model fit is improved, such that the CFI is within an acceptable range (0.95), and the RMSEA value is borderline (0.059). Also the Chi-square value is improved from 407.9 to 234.0 (with six fewer degrees of freedom), although still with a p-value >0.05. Results of the LM test were again evaluated, but none were theoretically meaningful, so we consider this model acceptable to move forward with hypothesis testing.

Fit Summary		
Modeling Info	N Observations	350
	N Variables	18
	N Moments	171
	N Parameters	66
	N Active Constraints	0
	Baseline Model Function Value	7.8962
	Baseline Model Chi-Square	2755.7718
	Baseline Model Chi-Square DF	153
	Pr > Baseline Model Chi-Square	<.0001
Absolute Index	Fit Function	0.6705
	Chi-Square	233.9922
	Chi-Square DF	105
	Pr > Chi-Square	<.0001
	Z-Test of Wilson & Hilferty	6.7015
	Hoelter Critical N	194
	Root Mean Square Residual (RMSR)	0.0430
	Standardized RMSR (SRMSR)	0.0430
	Goodness of Fit Index (GFI)	0.9255
Parsimony Index	Adjusted GFI (AGFI)	0.8786
	Parsimonious GFI	0.6351
	RMSEA Estimate	0.0593
	RMSEA Lower 90% Confidence Limit	0.0491
	RMSEA Upper 90% Confidence Limit	0.0695
	Probability of Close Fit	0.0653
	ECVI Estimate	1.0705
	ECVI Lower 90% Confidence Limit	0.9526
	ECVI Upper 90% Confidence Limit	1.2119
	Akaike Information Criterion	365.9922
	Bozdogan CAIC	686.6158
	Schwarz Bayesian Criterion	620.6158
	McDonald Centrality	0.8317
Incremental Index	Bentler Comparative Fit Index	0.9504
	Bentler-Bonett NFI	0.9151
	Bentler-Bonett Non-normed Index	0.9278
	Bollen Normed Index Rho1	0.8763
	Bollen Non-normed Index Delta2	0.9513
	James et al. Parsimonious NFI	0.6280

Display 3: Selected output from second structural equation model (fit statistics)

The output below indicates the standardized estimates for the linear equations in the model, and the t-value indicates the significance of these associations. We see that fruit identification and willingness to try fruit are significant predictors of fruit preferences, and that willingness to try vegetables is a significant predictor of vegetable preferences. For fruit intake, only fruit identification is a significant predictor; and for vegetable intake, fruit identification, vegetables identification, and willingness to try vegetables are significant predictors.

Standardized Results for Linear Equations

fruit_pref	=	0.2258	*	fruit_id	+	-0.0948	*	veg_id	+	0.3290	*	f1	+	0.1392	*	f2	+	1.0000		e1							
Std Err		0.0524		b1		0.0542		b2		0.0746		b3		0.0720		b4											
t Value		4.3124				-1.7485				4.4085				1.9346													
veg_pref	=	-0.00427	*	fruit_id	+	-0.0459	*	veg_id	+	-0.1218	*	f1	+	0.5488	*	f2	+	1.0000		e2							
Std Err		0.0529		b5		0.0540		b6		0.0763		b7		0.0667		b8											
t Value		-0.0808				-0.8505				-1.5973				8.2286													
fruit_willing_1	=	0.5501	*	f1	+	1.0000		e10																			
Std Err		0.0417		a1																							
t Value		13.2048																									
fruit_willing_2	=	0.6278	*	f1	+	1.0000		e11																			
Std Err		0.0383		a2																							
t Value		16.3878																									
fruit_willing_3	=	0.7276	*	f1	+	1.0000		e12																			
Std Err		0.0326		a3																							
t Value		22.3219																									
fruit_willing_4	=	0.5811	*	f1	+	1.0000		e13																			
Std Err		0.0408		a4																							
t Value		14.2578																									
fruit_willing_5	=	0.6809	*	f1	+	1.0000		e14																			
Std Err		0.0350		a5																							
t Value		19.4423																									
fruit_willing_6	=	0.6417	*	f1	+	1.0000		e15																			
Std Err		0.0368		a6																							
t Value		17.4157																									
veg_willing_1	=	0.7614	*	f2	+	1.0000		e16																			
Std Err		0.0251		a7																							
t Value		30.3433																									
veg_willing_2	=	0.7760	*	f2	+	1.0000		e17																			
Std Err		0.0244		a8																							
t Value		31.8369																									
veg_willing_3	=	0.8244	*	f2	+	1.0000		e18																			
Std Err		0.0208		a9																							
t Value		39.6661																									
veg_willing_4	=	0.7181	*	f2	+	1.0000		e19																			
Std Err		0.0286		a10																							
t Value		25.0965																									
veg_willing_5	=	0.8085	*	f2	+	1.0000		e20																			
Std Err		0.0218		a11																							
t Value		37.0371																									
veg_willing_6	=	0.7652	*	f2	+	1.0000		e21																			
Std Err		0.0249		a12																							
t Value		30.6789																									
fruit_intake	=	-0.1302	*	fruit_id	+	0.1004	*	veg_id	+	0.0879	*	f1	+	0.0541	*	f2	+	0.0172	*	fruit_pref	+	0.0477	*	veg_pref	+	1.0000	e3
Std Err		0.0592		b9		0.0591		b10		0.0898		b11		0.0878		b12		0.0658		b13		0.0661		b14			
t Value		-2.1980				1.6991				0.9796				0.6161				0.2607				0.7217					
veg_intake	=	-0.2106	*	fruit_id	+	0.1868	*	veg_id	+	-0.0633	*	f1	+	0.2303	*	f2	+	0.0535	*	fruit_pref	+	0.0795	*	veg_pref	+	1.0000	e4
Std Err		0.0564		b15		0.0562		b16		0.0863		b17		0.0835		b18		0.0631		b19		0.0634		b20			
t Value		-3.7372				3.3219				-0.7333				2.7575				0.8479				1.2532					

Display 4: Selected output from second structural equation model (linear equations)

CONCLUSION

These results indicate that identification and willingness to try fruits and vegetables are predictors of preferences and intake, but preference is not predictive of intake, as has previously been demonstrated in the literature.[7] A structural equation model is a useful way to examine these data because preferences can be evaluated as a mediator between intake and the predictors identification and willingness to try, but we see from these results that it is not one.

When specifying the model, the LM test is a helpful tool for model modification, to ensure that overall model fit is appropriate and that parameter estimates are reliable. However, prudence must be exercised when employing this approach so that models remain theoretically sound. Just as inappropriate model fit can diminish findings, so can too an illogical model.

REFERENCES

1. Davis JN, Ventura EE, Cook LT, Gyllenhammer LE, Gatto NM. LA Sprouts: a gardening, nutrition, and cooking intervention for Latino youth improves diet and reduces obesity. *YJADA*. 2011;111(8):1224–1230. doi:10.1016/j.jada.2011.05.009.
2. Gatto NM, Ventura EE, Cook LT, Gyllenhammer LE, Davis JN. LA Sprouts: a garden-based nutrition intervention pilot program influences motivation and preferences for fruits and vegetables in Latino youth. *J Acad Nutr Diet*. 2012;112(6):913–920. doi:10.1016/j.jand.2012.01.014.
3. Glasson C, Chapman K, James E. Fruit and vegetables should be targeted separately in health promotion programmes: differences in consumption levels, barriers, knowledge and stages of readiness for change. *Public Health Nutr*. 2010;14(04):694–701. doi:10.1017/S1368980010001643.
4. Garcia-Dominic O, Trevino RP, Echon RM, et al. Improving Quality of Food Frequency Questionnaire Response in Low-Income Mexican American Children. *Health Promotion Practice*. 2012;13(6):763–771. doi:10.1177/1524839911405847.
5. Domel SB, Baranowski T, Davis H, Leonard SB, Riley P, Baranowski J. Measuring fruit and vegetable preferences among 4th- and 5th-grade students. *Preventive Medicine*. 1993;22(6):866–879. doi:10.1006/pmed.1993.1078.
6. Pliner P, Hobden K. Development of a scale to measure the trait of food neophobia in humans. *Appetite*. 1992;19(2):105–120.
7. McClain AD, Chappuis C, Nguyen-Rodriguez ST, Yaroch AL, Spruijt-Metz D. Psychosocial correlates of eating behavior in children and adolescents: a review. *Int J Behav Nutr Phys Act*. 2009;6(1):54. doi:10.1186/1479-5868-6-54.

ACKNOWLEDGMENTS

This project is funded by NIH R21 DK094066-02.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Lauren Cook
Enterprise: University of Southern California
Address: 2250 Alcazar St, CSC-200
City, State ZIP: Los Angeles, CA, 90089
Work Phone: 323-442-2735
Fax: 323-442-4103
E-mail: laurenco@usc.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.