Supplementary Information:

Determinants of non-primary, primary, and exclusive use of liquefied petroleum gas (LPG) vary significantly in a nationally representative survey in Ghana

Abhishek Kar*[1], Theresa Tawiah[2], Linnea Graham[1], Georgette Owusu-Amankwah[1], Misbath Daouda[1], Flavio Malagutti[3], Steve Chillrud[4], Seidu Iddrisu[2], Edward A. Apraku[2], Richard Tetteh[2], Kelsey Jack[3], Sulemana W. Abubakari[2], Darby Jack[1], Kwaku P. Asante[2]

[1]Columbia World Projects, Columbia University, New York, USA
[2]Kintampo Health Research Centre, Research and Development Division, Ghana Health Service, Kintampo North Municipality, Bono East Region, Ghana
[3]Bren School of Environmental Science and Management, University of California, Santa Barbara, USA
[4]Lamont-Doherty Earth Observatory, Columbia University, New York, USA

SI1: Sociodemographic and economic relationship with LPG use level

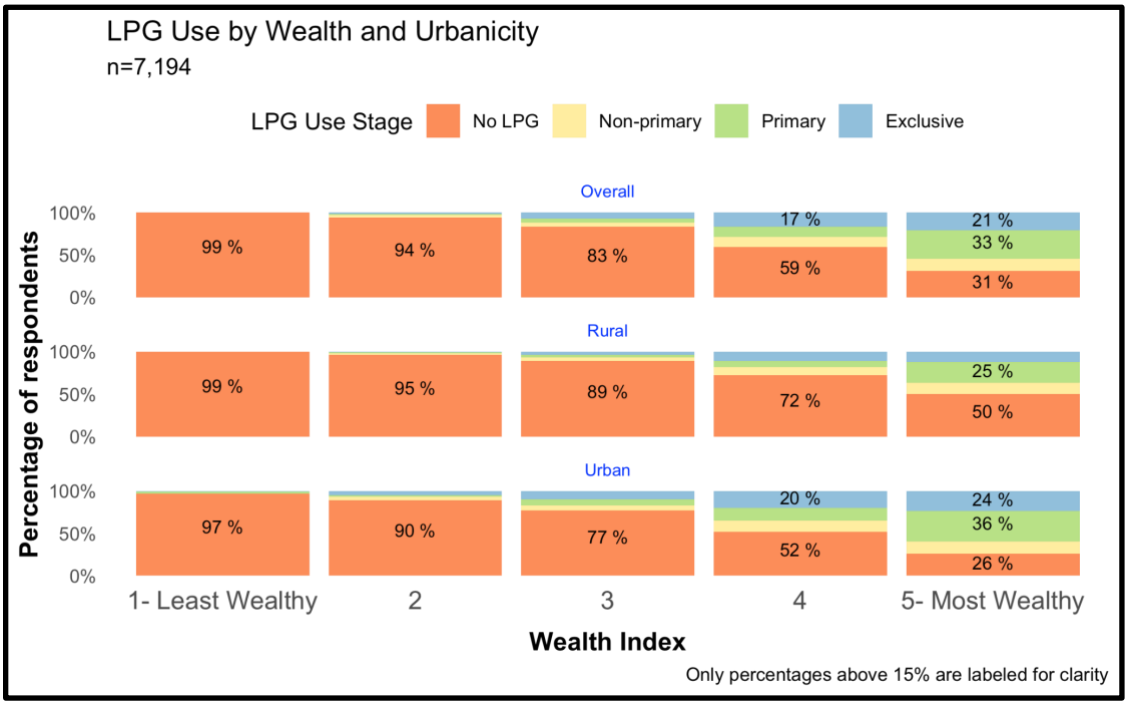


Figure SI1: LPG Use levels by Wealth & Geography: 100% Stacked bar of LPG use by stage when respondents are grouped by wealth index score quintiles

SI2: Regional variation in LPG use as main cooking fuel

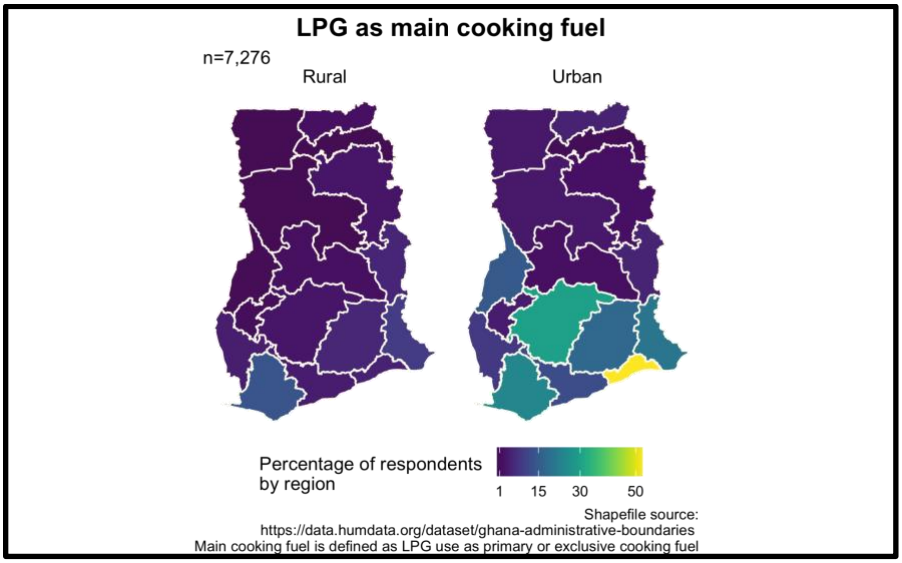


Figure SI2: Use of LPG as main fuel by region and urbanicity. Main fuel implies primary or exclusive use. "Only urban Accra has a majority of households using LPG as main fuel".

SI3: Correlation among quantitative indicators of LPG use

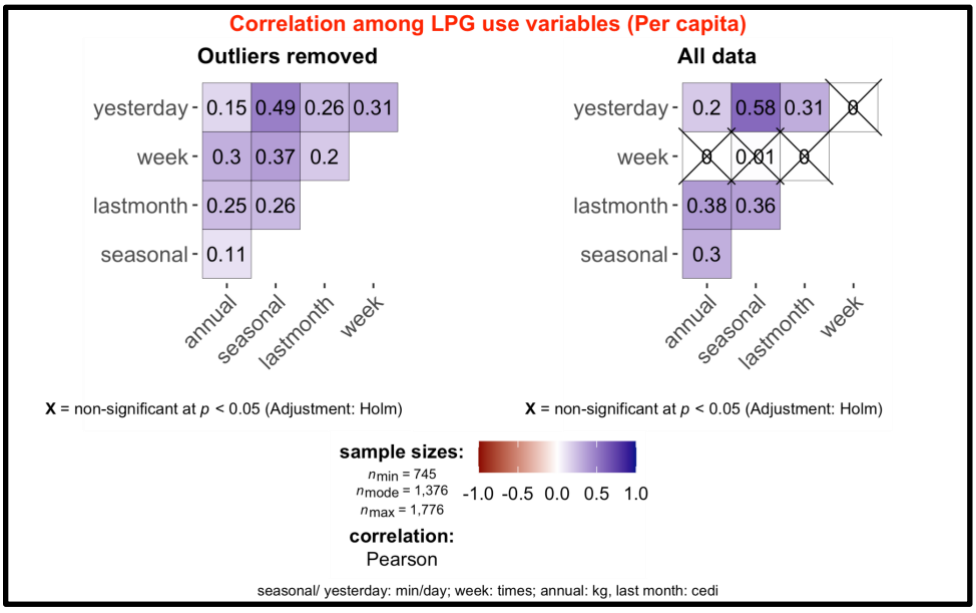


Figure SI3: Correlation among five quantitative LPG use indicators. While there is no negative correlation, the strength of the relationship improves from weak to moderate only when outliers are removed.

SI4: Behavioral perception on advantages and disadvantages of LPG use

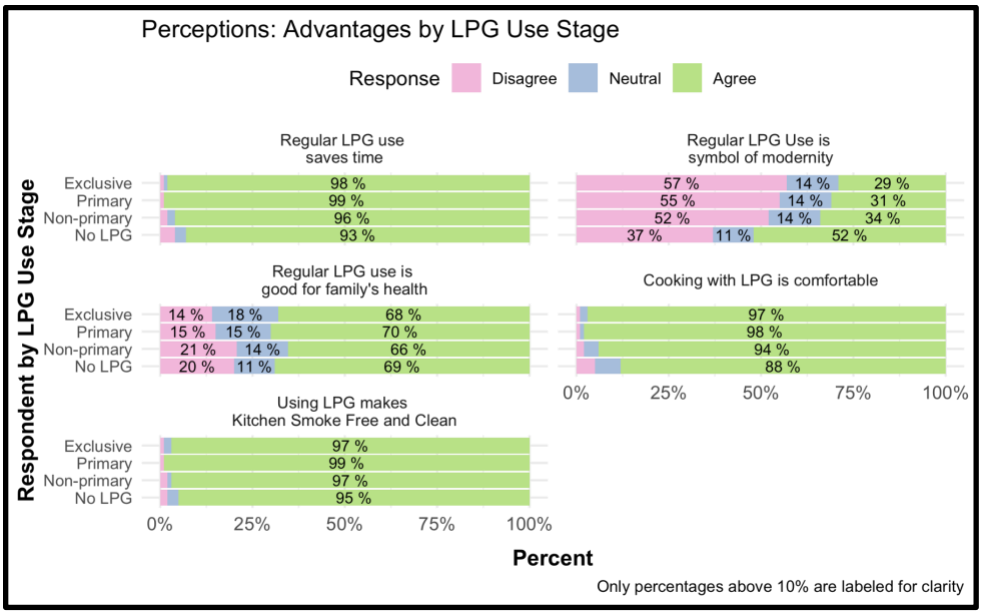


Figure SI4: Responses to five statements on advantages of LPG when respondents grouped by stage of LPG use. There is limited variation in response by stage.

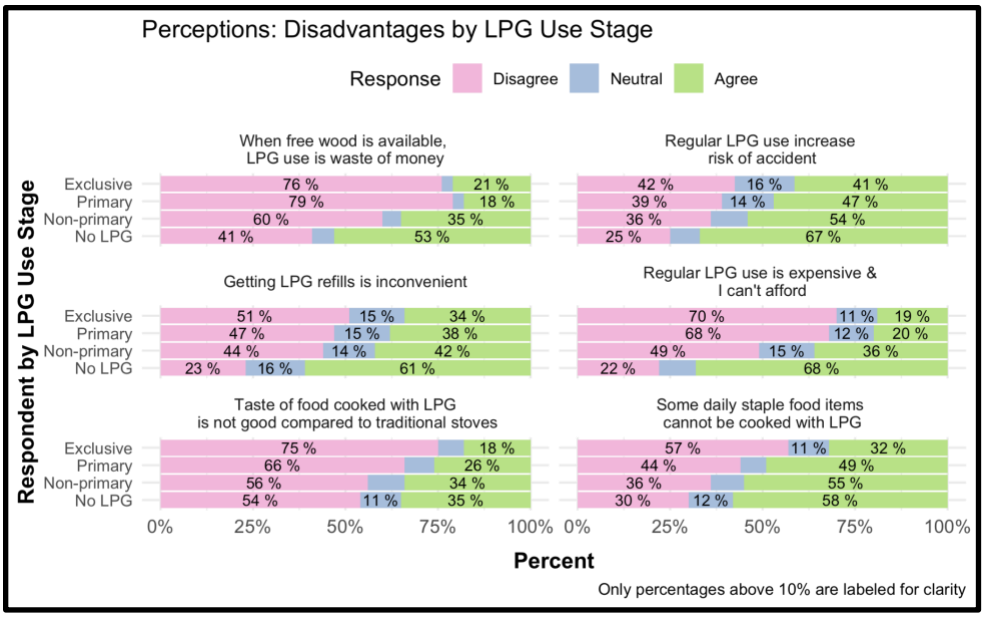


Figure SI5: Responses to five statements on disadvantages of LPG when respondents grouped by stage of LPG use. There is limited variation in response by stage.

Processing of Perception Scores

Advan_total = stove_gas_pref_3 + stove_gas_pref_5+ stove_gas_pref_7 + stove_gas_pref_9 + stove_gas_pref_11

The five variables represent the 5 statements from Figure SI5

#1: Agree becomes +1 points for agreeing with a positive sentiment about LPG; 2: Neutral becomes 0 points for neutral with a positive sentiment; #3: Disagree becomes -1 point for disagreeing with a positive sentiment; Higher score signals more favorable feeling towards LPG. A higher advantages total score indicates a more positive feeling toward LPG. The range of the *Advan_total* variable is +5 to -5

disadvan_total = stove_gas_pref_1 + stove_gas_pref_2 + stove_gas_pref_4 + stove_gas_pref_6+ stove_gas_pref_8 + stove_gas_pref_10

The six variables represent the 6 statements from Figure SI4

#1: Agree becomes -1 points for agreeing with the negative sentiment about LPG; 2: Neutral becomes 0 points for neutral with a negative sentiment; #3: Disagree becomes +1 point for disagreeing with negative sentiment (double negative= positive); Higher

score signals more favorable feelings toward LPG: negating the negative feeling about LPG. The range of the *Disadvan_total* variable is +6 to -6

SI4: Binary logistic regression for three processes

Choice of independent variables

We excluded a few frequently mentioned factors in the literature for three reasons. First, some factors were already captured more reliably by the listed factors. For example, we consider wealth index over income/ expenditure metrics due to reliability issues around direct response to income and expenditure estimates. Studies have shown that income estimates are often subject to social desirability bias as well as recall bias (Kar et al., 2020). Second, some factors were dropped due to restrictions imposed by statistical analysis models. For example, the autonomy factor related to 'Primary Cook is also household head" is dropped. The factor is accounted for indirectly as we include the age & gender of the household head in the analysis. When the primary cook is the same as the household head, there will be an absolute correlation between primary cook age/ gender and household head age/gender when the value of the third variable (Primary Cook= Household Head) is 1. It would violate core assumptions around multicollinearity. Third, we made a judgment call that some factors were not particularly relevant in the study context, i.e. they are not likely to play a differentiating role in the transition process. For example, social strata/ caste are more relevant in the Indian context compared to Ghana. We also did not use factors like the price of LPG as all survey respondents were subject to the market price (and its fluctuations) in Ghana. Fourth, we applied some factors conditionally to the three sub-transition process analyses. For example, the burner number is 0 for those who don't have LPG. If a household does not use LPG, the LPG burner number (=0) is a result of them not using LPG indicating reverse causality. Hence, we did not include them for the *uptake* sub-transition but as an important predictor for the other processes.

Processing of raw variables

While some predictor variables were used as-is from the survey data like age and gender, others like wealth index and family size were processed from the raw data. Detailed notes on the processing of variables are available in the companion R codes. For a few variables we have used two ways: original form as continuous, such as age, and categorical by splitting age into three groups (Table SI6).

<table>
<tr><th>Factor Description</th><th>Type used for regression</th><th>Data Processing, if any</th></tr>
<tr><td>1. Wealth Index Score</td><td>Continuous and Categorical</td><td>Wealth Index Score Calculation is detailed in Annexure 1 and appended with this Supplementary Information. When a categorical version of the wealth index score is used, we categorize the 1st to 3rd quantile range as “middle class”. Lower and higher values beyond the range are categorized as “Poor” and “Rich” respectively</td></tr>
<tr><td>2. Family/ Household Size</td><td>Continuous and Categorical</td><td>We adapt the "Standard adult" equivalence factors defined in terms of sex and age when calculating household size from the list of household members as per the Kitchen Performance Test Protocol (Bailis, 2007)

<table><tr><td>Gender and age</td><td>Fraction of Standard Adult</td></tr><tr><td>Child: 0-14 years</td><td>0.5</td></tr><tr><td>Female: over 14 years</td><td>0.8</td></tr><tr><td>Male: 15- 59 years</td><td>1.0</td></tr><tr><td>Male: Over 59 years</td><td>0.8</td></tr></table>
When a categorical version of the household size is used, we categorize the 1st to 3rd quantile range as “Medium”. Lower and higher values beyond the range are categorized as “Small” and “Large” respectively</td></tr>
<tr><td>3. Age of primary cook</td><td>Continuous and Categorical</td><td>As is from the survey database; when a categorical version of the age of primary cook is used, we categorize the 1st to 3rd quantile range as “middle age”. Lower and higher values beyond the range are categorized as “Young” and “Old” respectively</td></tr>
<tr><td>4. Gender of primary cook</td><td>Categorical</td><td>As is from survey database</td></tr>
</table>

<table>
<tr><td>5. Education of primary cook</td><td>Categorical</td><td>Respondents chose from the following options:
-10 None
1 Primary school: uncompleted
2 Primary school: completed
3 Middle/Junior HS: uncompleted
4 Middle/Junior HS: completed
5 SHS/Commercial/Technical: uncompleted
6 SHS/Commercial/Technical: completed
7 Post-middle training: teachers, secretarial, etc.
8 Post-sec training: Nursing, Teacher, Polytechnic,etc.
9 University: undergraduate completed
10 University: undergraduate uncompleted
11 University: masters or higher uncompleted
12 University: masters or higher completed
-99 I Don't Know
-88 I Don't Want to Answer

We categorized 1-5 & -10 as “No Higher Education”, 6-12 as “Higher Education” and -88, -99 & NA (missing) as NA (missing)</td></tr>
<tr><td>6. Age of household head</td><td>Continuous and Categorical</td><td>As is from the survey database, when a categorical version of the age of hh-head is used, we categorize the 1st to 3rd quantile range as “middle age”. Lower and higher values beyond the range are categorized as “Young” and “Old” respectively</td></tr>
<tr><td>7. Gender of household head</td><td>Categorical</td><td>As is from survey database</td></tr>
<tr><td>8. Education of household head</td><td>Categorical</td><td>As is from survey database</td></tr>
<tr><td>9. Positive perceptions about regular LPG use</td><td>Continuous and Categorical</td><td>Detailed in section SI3. When a categorical version of the positive perception score is used, we can’t use the quantile data as the 3rd quantile and maximum value are the same. So, after review of the decile data, we manually decided on the following:
<table><tr><td>Min</td><td>Max</td><td>Category</td></tr><tr><td>-5</td><td>2</td><td>Low</td></tr><tr><td>3</td><td>3</td><td>Medium</td></tr><tr><td>4</td><td>5</td><td>High</td></tr></table></td></tr>
</table>

10. Negative perceptions of regular LPG use	Continuous and Categorical	Detailed in section SI3. When a categorical version of the negative perception score is used, we categorize the 1st to 3rd quantile range as “Medium”. Lower and higher values beyond the range are categorized as “Small” and “High” respectively.
11. No. of burners in LPG stove	Continuous and Categorical	As is from the survey database. When a categorical version of the burner size is used, we categorize them as “Single”, “Double”, & “Triple or more”
12. Urban or Rural	Categorical	As is from survey database
13. Convenience to access LPG fuel	Categorical	Respondents chose from the following options: 1 Home delivered by LPG supplier 2 Send the cylinder with someone else: PAYS NO transportation charge 3 Send the cylinder with someone else: PAYS A transportation charge 4 Travels less than 1 km 5 Travels 1-3 km 6 Travels 4-6 km 7 Travels more than 6 km Those with options 1-4, were categorized as “Yes” and 5-7 with “No”. -88, -99 & NA (missing) were coded as NA (missing)

Table SI6: Processing of thirteen variables identified as drivers and barriers

Regression results

We have done two broad types of regression: binary logistic regression and ordered categorical regression as detailed below.

A. Binary Logistic Regression

We conducted four rounds of generalized linear model-based logistic regression to test the influence of predictors on the outcome (success or failure) for each of the three transition choices: *uptake*, *main*, and *exclusive*.

Round 1: All-mixed

First, we conducted analyses with all (for uptake: 11, others: 13) factors as predictors for regression analysis. These are mixed models with both continuous (e.g.,

primary cook age), and categorical (e.g. primary cook gender) predictors. These *all-mix* models are called models 1, 5, and 9 for *uptake*, *main*, and *exclusive* respectively.

Round 2: Significant-mixed

Second, we only selected the significant ($p<0.05$) factors for an updated estimate of their influence to re-run the regression analysis. In other words, we removed all the predictors with non-significant results. These *significant-mix* models are called models 2, 6, and 10 for *uptake*, *main*, and *exclusive* respectively. For these models, we then ran the full battery of tests to check how well the underlying assumptions for the binary logistic regression models hold true. The seven assumptions are detailed below in Table SI07. In the case of if an independent variable is statistically significant (i.e., $p \leq 0.05$), it implies the presence of non-linearity between the independent variable and the logit of the outcome variable.

Assumptions	**Significant-mixed**	**Significant-categorical**
1: One dependent variable that is dichotomous (Study design)	Yes	Yes
2: One or more independent variables that are measured on either a continuous or nominal scale (Study design)	Yes	Yes
3. Independence of observations: Predictors and dependent variables should be mutually exclusive and exhaustive (Study design)	Yes	Yes
4. Minimum of 50 cases per independent variable (Study implementation)	Yes	Yes
5. The linear relationship between the continuous independent variables and the log-odds of the dependent variable (Study result)	Partial	Yes
6. No multicollinearity; occurs when two or more independent variables are highly correlated with each other.	Yes	Yes
7. No significant outliers, high leverage points, or highly influential points.	Partial	Yes

Table SI07: Regression assumption tests

We find that the underlying assumption of the linear relationship between continuous predictor variables and the logit of the outcome was not fully satisfied for all

three models 2,6, & 10. In the Box-Tidwell test, if an independent variable is statistically significant (i.e., $p \leq 0.05$), it implies the presence of non-linearity between the independent variable and the logit of the outcome variable. We found 2/4, 2/5, and 2/3 continuous variables for *uptake*, *main*, and *exclusive* respectively failed to meet this assumption. We also found 1 influential outlier datapoint for *uptake* (model 2) and no influential outlier datapoint for *main*, and *exclusive* transition choices.

Round 3: All-categorical

Third, we converted all continuous variables into three categories and re-ran the analyses. The three categories are interquartile range (IQR: medium), <IQR as low, and >IQR: high. We manually adjusted the category cutoff when the 3rd quartile and max were identical in a couple of cases. These *all-categorical* models are called models 3, 7, and 11 for *uptake*, *main*, and *exclusive* respectively.

Round 4: Significant-categorical

Finally, we only selected the significant ($p<0.05$) factors for an updated estimate of their influence to re-run the regression analysis. In other words, we removed all the categorical predictors with non-significant results. If the results were significant for one level but not for the other, we removed the predictor. For example, Primary Cook: Middle Age (vs. Young) as a predictor variable for main is significant, but Primary Cook: Old (vs. Young) was not significant. So, we dropped Primary Cook Age (categorical) in this round. These *significant-categorical* models are called models 4, 8, and 12 for *uptake*, *main*, and *exclusive* respectively. Again, for these models, we then ran the full battery of tests to check how well the underlying assumptions for the binary logistic regression models hold true (Table SI07). They have satisfied all assumptions when applicable. For example, we did not run the Box-Tidwell test as there were no more continuous predictors. The binary logistic regression results for *uptake*, *main*, and *exclusive* transition choices are shown in Tables SI8, SI9, and SI10 respectively. The tables were created using the *stargazer* package in R platform (Hlavac, 2022).

	Binary Logistic Regression for LPG Uptake			
	LPG Uptake (Binary dependent variable): Yes (1), No (0)			
	Model 1	Model 2	Model 3	Model 4
Wealth Index Score	0.510***	0.509***		
	(0.024)	(0.024)		
Wealth Index: MiddleClass (vs. Poor)			1.723***	1.736***
			(0.215)	(0.214)
Wealth Index: Rich (vs. Poor)			3.314***	3.337***
			(0.223)	(0.222)
Primary Cook Age	0.005			
	(0.005)			
Primary Cook: Middle Age (vs. Young)			0.028	
			(0.103)	
Primary Cook: Old (vs. Young)			-0.149	
			(0.145)	
Primary Cook: Male (vs. Female)	-0.041		0.198	
	(0.148)		(0.146)	
Primary Cook HigherEducation: Yes (vs. No)	0.593***	0.549***	0.652***	0.689***
	(0.120)	(0.115)	(0.115)	(0.110)
Household Head Age	-0.028***	-0.025***		
	(0.004)	(0.003)		
Household Head: Middle Age (vs. Young)			-0.421***	-0.435***
			(0.109)	(0.096)
Household Head: Old (vs. Young)			-0.727***	-0.833***
			(0.149)	(0.120)
Household Head = Female? Yes	-0.119		0.019	
	(0.097)		(0.094)	
Household Head HigherEducation: Yes (vs. No)	0.237**	0.284**	0.438***	0.421***
	(0.115)	(0.110)	(0.110)	(0.107)
Family Size	-0.355***	-0.349***		
	(0.032)	(0.029)		
Family Size: Medium (vs. Small)			-0.418***	-0.439***
			(0.101)	(0.093)
Family Size: Large (vs. Small)			-1.130***	-1.157***
			(0.133)	(0.124)
Location: Urban (vs. Rural)	0.840***	0.832***	0.918***	0.916***
	(0.088)	(0.087)	(0.086)	(0.085)
Positive perception Score	0.034			
	(0.136)			
Positive perception: Middle (vs. Small)			0.256**	
			(0.105)	
Positive perception: High (vs. Small)			-0.047	
			(0.107)	
Negative perception Score	1.320***	1.322***		
	(0.083)	(0.083)		
Negative perception: Medium (vs. Small)			0.540***	0.576***
			(0.108)	(0.107)
Negative perception: High (vs. Small)			1.612***	1.651***
			(0.112)	(0.112)
Observations	4,604	4,604	4,604	4,604
Log Likelihood	-1,875.566	-1,876.836	-1,967.094	-1,975.259
Akaike Inf. Crit.	3,775.132	3,769.673	3,970.187	3,974.518
Bayesian Inf. Crit.	3852.3	3821.2	4086	4051.7
McFadden pseduo R^2	0.398	0.397	0.368	0.366
Hosmer Lemeshow goodness of fit- p value	0.15	0.721	0.073	0.162
Note:				*p<0.1; **p<0.05; ***p<0.01

Table SI8: Binary Logistic regression for *uptake* transition choice outcome with eleven predictors

	Binary Logistic Regression for LPG Use as main fuel			
	LPG mainuse (Binary dependent variable): Yes (1), No (0)			
	Model 5	Model 6	Model 7	Model 8
Wealth Index Score	0.113***	0.115***		
	(0.030)	(0.030)		
Wealth Index: MiddleClass (vs. Poor)			-1.407**	
			(0.645)	
Wealth Index: Rich (vs. Poor)			-0.966	
			(0.650)	
Primary Cook Age	-0.005			
	(0.007)			
Primary Cook: Middle Age (vs. Young)			-0.341**	
			(0.164)	
Primary Cook: Old (vs. Young)			-0.377	
			(0.238)	
Primary Cook: Male (vs. Female)	1.067***	1.090***	1.141***	1.125***
	(0.271)	(0.259)	(0.271)	(0.248)
Primary Cook HigherEducation: Yes (vs. No)	0.499***	0.539***	0.550***	0.726***
	(0.170)	(0.139)	(0.169)	(0.129)
Household Head Age	-0.025***	-0.029***		
	(0.007)	(0.005)		
Household Head: Middle Age (vs. Young)			-0.024	
			(0.172)	
Household Head: Old (vs. Young)			-0.935***	
			(0.236)	
Household Head = Female? Yes	-0.098		-0.070	
	(0.146)		(0.148)	
Household Head HigherEducation: Yes (vs. No)	0.031		0.183	
	(0.168)		(0.167)	
Family Size	-0.394***	-0.375***		
	(0.052)	(0.048)		
Family Size: Medium (vs. Small)			-0.998***	-0.896***
			(0.165)	(0.145)
Family Size: Large (vs. Small)			-1.582***	-1.528***
			(0.220)	(0.191)
Convenient Access to Fuel: Yes (vs. No)	0.064		0.058	
	(0.139)		(0.140)	
Location: Urban (vs. Rural)	0.487***	0.504***	0.541***	0.484***
	(0.148)	(0.147)	(0.150)	(0.140)
Positive perception Score	0.295			
	(0.237)			
Positive perception: Middle (vs. Small)			0.102	
			(0.166)	
Positive perception: High (vs. Small)			0.013	
			(0.177)	
Negative perception Score	0.904***	0.883***		
	(0.125)	(0.124)		
Negative perception: Medium (vs. Small)			0.766***	0.731***
			(0.195)	(0.182)
Negative perception: High (vs. Small)			1.286***	1.272***
			(0.191)	(0.179)
Number of LPG Burners	0.463***	0.460***		
	(0.065)	(0.064)		
LPG Burners: Double (vs. Single)			0.857***	0.884***
			(0.175)	(0.166)
LPG Burners: Triple(vs. Single)			1.291***	1.210***
			(0.169)	(0.150)
Observations	1,741	1,762	1,741	1,788
Log Likelihood	-784.066	-794.839	-780.853	-846.887
Akaike Inf. Crit.	1,596.132	1,607.678	1,603.706	1,713.774
Bayesian Inf. Crit.	1672.6	1656.9	1718.4	1768.7
McFadden pseduo R^2	0.214	0.213	0.218	0.18
Hosmer Lemeshow goodness of fit- p value	0.668	0.592	0.877	0.553
Note:				*p<0.1; **p<0.05; ***p<0.01
				Only LPG users are considered
	Perception questions around LPG use only answered by respondents who have seen/ operated LPG stove			

Table SI9: Binary Logistic regression for *main* transition choice outcome with thirteen predictors

210

	Binary Logistic Regression for LPG Use as Exclusive Fuel			
	LPG as exclusive cooking fuel (Binary dependent variable): Yes (1), No (0)			
	Model 9	Model 10	Model 11	Model 12
Wealth Index Score	-0.085***	-0.078***		
	(0.027)	(0.023)		
Wealth Index: MiddleClass (vs. Poor)			-0.156	
			(0.595)	
Wealth Index: Rich (vs. Poor)			-0.779	
			(0.603)	
Primary Cook Age	-0.011			
	(0.008)			
Primary Cook: Middle Age (vs. Young)			0.001	
			(0.156)	
Primary Cook: Old (vs. Young)			-0.174	
			(0.254)	
Primary Cook: Male (vs. Female)	1.240***	1.224***	1.273***	1.280***
	(0.202)	(0.189)	(0.197)	(0.183)
Primary Cook HigherEducation: Yes (vs. No)	0.259		0.304*	
	(0.185)		(0.182)	
Household Head Age	0.005			
	(0.008)			
Household Head: Middle Age (vs. Young)			0.032	
			(0.165)	
Household Head: Old (vs. Young)			-0.095	
			(0.269)	
Household Head = Female? Yes	0.353**	0.328**	0.340**	0.330**
	(0.152)	(0.140)	(0.152)	(0.138)
Household Head HigherEducation: Yes (vs. No)	0.437**	0.633***	0.414**	0.579***
	(0.188)	(0.129)	(0.184)	(0.123)
Family Size	-0.409***	-0.425***		
	(0.066)	(0.061)		
Family Size: Medium (vs. Small)			-0.853***	-1.040***
			(0.147)	(0.133)
Family Size: Large (vs. Small)			-1.424***	-1.543***
			(0.271)	(0.237)
Convenient Access to Fuel: Yes (vs. No)	-0.149		-0.152	
	(0.141)		(0.142)	
Location: Urban (vs. Rural)	0.094		0.154	
	(0.160)		(0.163)	
Positive perception Score	-0.353			
	(0.253)			
Positive perception: Middle (vs. Small)			-0.260	
			(0.169)	
Positive perception: High (vs. Small)			-0.344*	
			(0.180)	
Negative perception Score	0.509***	0.448***		
	(0.131)	(0.124)		
Negative perception: Medium (vs. Small)			-0.173	-0.265
			(0.248)	(0.231)
Negative perception: High (vs. Small)			0.330	0.173
			(0.235)	(0.219)
Number of LPG Burners	0.011			
	(0.060)			
LPG Burners: Double (vs. Single)			0.289*	
			(0.173)	
LPG Burners: Triple(vs. Single)			0.102	
			(0.161)	
Observations	1,288	1,400	1,288	1,411
Log Likelihood	-749.344	-820.802	-750.356	-838.010
Akaike Inf. Crit.	1,526.689	1,655.604	1,542.713	1,692.020
Bayesian Inf. Crit.	1672.6	1656.9	1718.4	1768.7
McFadden pseduo R^2	0.214	0.213	0.218	0.18
Hosmer Lemeshow goodness of fit- p value	0.668	0.592	0.877	0.553
Note:				*p<0.1; **p<0.05; ***p<0.01
				Only main (primary/ exclusive) LPG users are considered
				Perception questions around LPG use only answered by respondents who have seen/ operated LPG stove

211

212 Table SI10: Binary Logistic regression for *exclusive* transition choice outcome with thirteen predictors

Notably for *uptake*, two predictors of convenience to LPG fuel, and the number of burners were not used as they are discussed in the choice of predictors. Please note that sample size numbers can slightly vary as for each regression, any blank (missing) value in any factor leads to that entire respondent (observation) getting dropped.

For *uptake*, we do not find any difference between model 2 and model 4 in terms of significant predictors- it shows that when continuous predictors are converted to categorical, it did not change the key results. For *main*, two predictors Household Head Age and Wealth Index Score that were statistically significant ($p<0.05$) in model 6 were dropped in model 8. The odds ratio for Household Head Age and Wealth Index Score were 0.97 and 1.12 respectively in model 6. For *exclusive*, two predictors Negative perception Score and Wealth Index Score that were statistically significant ($p<0.05$) in model 10 were dropped in model 12. The odds ratio for the Negative perception Score and Wealth Index Score were 1.56 and 0.92 respectively in model 10.

We also conducted four tests for model fit assessment which are shown in the tables SI 8,9, and 10. One, McFadden pseudo-squared R which mirrors the adjusted R-squared in OLS by penalizing a model for including too many predictors. Values between 0.2 and 0.4 imply an excellent fit. Two, the Hosmer-Lemeshow test which is a statistical test for goodness of fit for the logistic regression model. Small p-values mean that the model is a poor fit. We also used Probabilistic model selection (or “information criteria”) which provides an analytical technique for scoring and choosing among candidate models. Three, we calculate the Akaike information criterion (AIC). AIC is used to compare different models and determine which one is the best fit for the data. Lower AIC values are better as AIC penalizes models that use more parameters. So if two models explain the same amount of variation, the one with fewer parameters will have a lower AIC score and will be the better-fit model. Four, we also calculate the Bayesian information criterion (BIC). Unlike the AIC, the BIC penalizes the model more for its complexity, meaning that more complex models will have a worse (larger) score and will, in turn, be less likely to be selected.

We select model 2 over model 4 for *uptake* as it scores better in 4 out of 4 goodness of fit tests though it was partially successful in two underlying assumptions tests. (Both models have a similar statistically significant set of predictors wherein the continuous predictors in model 2 were cut into three levels of categorical predictors).

We select model 6 over model 8 for *main* as it scores better in 4 out of 4 goodness of fit tests though it was partially successful in one underlying assumptions test. Notably, model 6 has a better fit in spite of having two additional continuous predictors that were not significant in model 8 when converted to categorical predictors (for both levels).

We select model 10 over model 12 for *exclusive* as it scores better in 4 out of 4 goodness of fit tests though it was partially successful in one underlying assumptions test. Notably, model 10 has a better fit in spite of having two additional continuous predictors that were not significant in model 12 when converted to categorical predictors (for both levels).

Hence, in the main paper, we present the findings from models 2, 6, and 10 in Figure 4, but present the mixed model results in Table SI11.

	Transition Choice Outcome: Binary Logistic Regression					
	LPG Uptake: Yes (1), No (0)		LPG as Main Fuel: Yes (1), No (0)		LPG as Exclusive Fuel: Yes (1), No (0)	
	All	Significant	All	Significant	All	Significant
Wealth Index Score	0.51***	0.51***	0.11***	0.11***	-0.09***	-0.08***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)
Primary Cook Age	0.01		-0.00		-0.01	
	(0.00)		(0.01)		(0.01)	
Primary Cook = Male? Yes	-0.04		1.07***	1.09***	1.24***	1.22***
	(0.15)		(0.27)	(0.26)	(0.20)	(0.19)
Primary Cook HigherEducation? Yes	0.59***	0.55***	0.50***	0.54***	0.26	
	(0.12)	(0.12)	(0.17)	(0.14)	(0.19)	
Household Head Age	-0.03***	-0.02***	-0.03***	-0.03***	0.00	
	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	
Household Head = Female? Yes	-0.12		-0.10		0.35**	0.33**
	(0.10)		(0.15)		(0.15)	(0.14)
Household Head HigherEducation? Yes	0.24**	0.28**	0.03		0.44**	0.63***
	(0.11)	(0.11)	(0.17)		(0.19)	(0.13)
Family Size	-0.35***	-0.35***	-0.39***	-0.38***	-0.41***	-0.43***
	(0.03)	(0.03)	(0.05)	(0.05)	(0.07)	(0.06)
Convenient LPG Fuel Access? Yes			0.06		-0.15	
			(0.14)		(0.14)	
Location= Urban? Yes	0.84***	0.83***	0.49***	0.50***	0.09	
	(0.09)	(0.09)	(0.15)	(0.15)	(0.16)	
Positive perception Score	0.03		0.29		-0.35	
	(0.14)		(0.24)		(0.25)	
Negative perception Score	1.32***	1.32***	0.90***	0.88***	0.51***	0.45***
	(0.08)	(0.08)	(0.13)	(0.12)	(0.13)	(0.12)
Number of burners in LPG stove			0.46***	0.46***	0.01	
			(0.06)	(0.06)	(0.06)	
Observations	4,604	4,604	1,741	1,762	1,288	1,400
Log Likelihood	-1,875.57	-1,876.84	-784.07	-794.84	-749.34	-820.80
Akaike Inf. Crit.	3,775.13	3,769.67	1,596.13	1,607.68	1,526.69	1,655.60
Bayesian Inf. Crit.	3852.3	3821.2	1672.6	1656.9	1598.9	1692.3
McFadden pseduo R^2	0.398	0.397	0.214	0.213	0.159	0.153
Hosmer Lemeshow goodness of fit- p value	0.15	0.721	0.668	0.592	0.178	0.839
Note:						*p<0.1; **p<0.05; ***p<0.01
						Main fuel implies LPG use as primary or exclusive fuel

Table SI11: Summary of binary logistic regression for all three transition nodes using a mix of categorical and continuous variables

We also present the graphical plot (Figure SI12) and summary (Table SI13) of binary logistic regression for all three transition nodes using only categorical variables that satisfy all regression assumptions completely. The figures were created using the *dotwhisker* package in R platform (Solt et al., 2022).

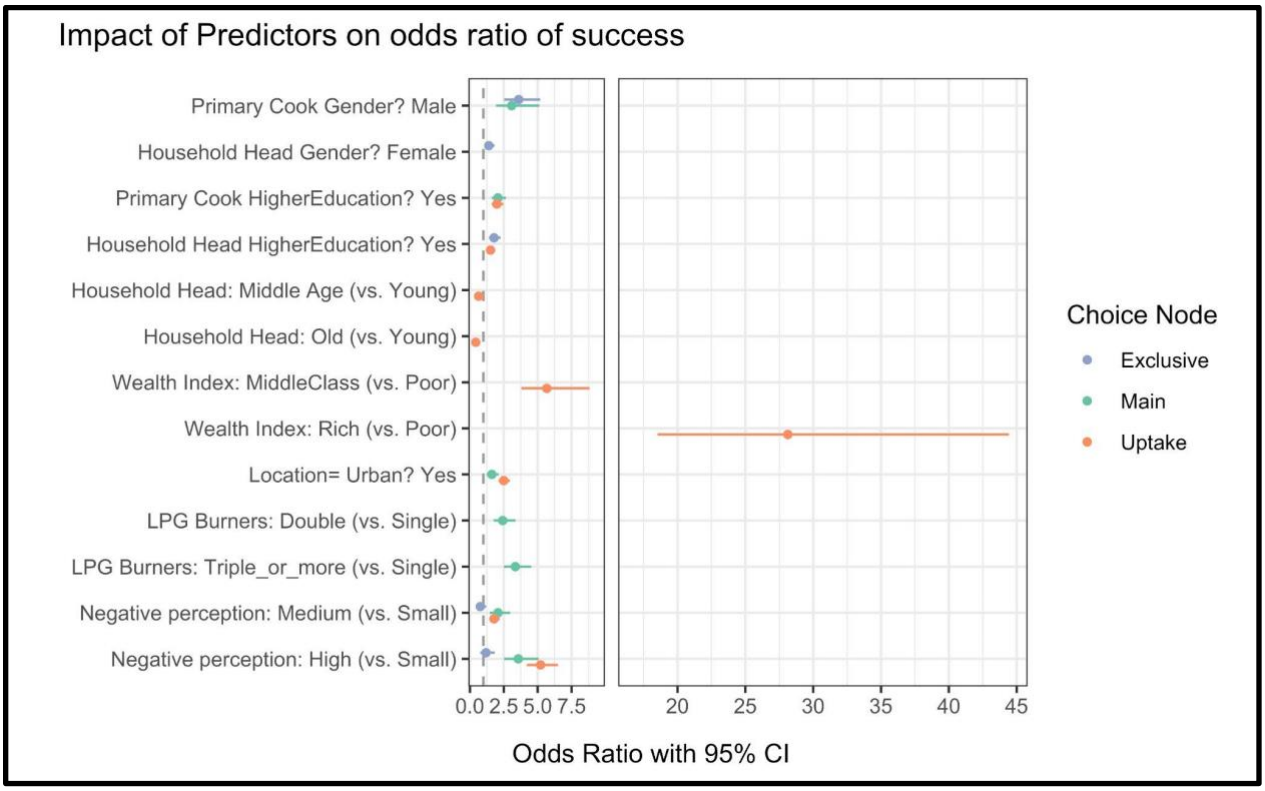


Figure SI12: Coefficient plot for all the three processes (ref: Figure 1) of LPG transition only showing predictors that were found to be statistically significant for the three choices. exp(Estimate) provides the odds value of the factor using the exponential of the estimate (coefficient value) of the binary logistic regression. The circles represent the odds value while the line shows the 95% confidence interval around the odds value.

	Transition Choice Outcome: Binary Logistic Regression					
	LPG Uptake: Yes (1), No (0)		LPG as Main Fuel: Yes (1), No (0)		LPG as Exclusive Fuel: Yes (1), No (0)	
	All	Significant	All	Significant	All	Significant
Wealth Index: MiddleClass (vs. Poor)	1.72***	1.74***	-1.41**		-0.16	
	(0.21)	(0.21)	(0.65)		(0.59)	
Wealth Index: Rich (vs. Poor)	3.31***	3.34***	-0.97		-0.78	
	(0.22)	(0.22)	(0.65)		(0.60)	
Primary Cook: Middle Age (vs. Young)	0.03		-0.34**		0.00	
	(0.10)		(0.16)		(0.16)	
Primary Cook: Old (vs. Young)	-0.15		-0.38		-0.17	
	(0.14)		(0.24)		(0.25)	
Primary Cook: Male (vs. Female)	0.20		1.14***	1.12***	1.27***	1.28***
	(0.15)		(0.27)	(0.25)	(0.20)	(0.18)
Primary Cook HigherEducation: Yes (vs. No)	0.65***	0.69***	0.55***	0.73***	0.30*	
	(0.11)	(0.11)	(0.17)	(0.13)	(0.18)	
Household Head: Middle Age (vs. Young)	-0.42***	-0.44***	-0.02		0.03	
	(0.11)	(0.10)	(0.17)		(0.16)	
Household Head: Old (vs. Young)	-0.73***	-0.83***	-0.94***		-0.10	
	(0.15)	(0.12)	(0.24)		(0.27)	
Household Head = Female? Yes	0.02		-0.07		0.34**	0.33**
	(0.09)		(0.15)		(0.15)	(0.14)
Household Head HigherEducation: Yes (vs. No)	0.44***	0.42***	0.18		0.41**	0.58***
	(0.11)	(0.11)	(0.17)		(0.18)	(0.12)
Family Size: Medium (vs. Small)	-0.42***	-0.44***	-1.00***	-0.90***	-0.85***	-1.04***
	(0.10)	(0.09)	(0.17)	(0.14)	(0.15)	(0.13)
Family Size: Large (vs. Small)	-1.13***	-1.16***	-1.58***	-1.53***	-1.42***	-1.54***
	(0.13)	(0.12)	(0.22)	(0.19)	(0.27)	(0.24)
Convenient Access to Fuel: Yes (vs. No)			0.06		-0.15	
			(0.14)		(0.14)	
Location: Urban (vs. Rural)	0.92***	0.92***	0.54***	0.48***	0.15	
	(0.09)	(0.08)	(0.15)	(0.14)	(0.16)	
Positive perception: Middle (vs. Small)	0.26**		0.10		-0.26	
	(0.11)		(0.17)		(0.17)	
Positive perception: High (vs. Small)	-0.05		0.01		-0.34*	
	(0.11)		(0.18)		(0.18)	
Negative perception: Medium (vs. Small)	0.54***	0.58***	0.77***	0.73***	-0.17	-0.27
	(0.11)	(0.11)	(0.19)	(0.18)	(0.25)	(0.23)
Negative perception: High (vs. Small)	1.61***	1.65***	1.29***	1.27***	0.33	0.17
	(0.11)	(0.11)	(0.19)	(0.18)	(0.23)	(0.22)
LPG Burners: Double (vs. Single)			0.86***	0.88***	0.29*	
			(0.18)	(0.17)	(0.17)	
LPG Burners: Triple(vs. Single)			1.29***	1.21***	0.10	
			(0.17)	(0.15)	(0.16)	
Observations	4,604	4,604	1,741	1,788	1,288	1,411
Log Likelihood	-1,967.09	-1,975.26	-780.85	-846.89	-750.36	-838.01
Akaike Inf. Crit.	3,970.19	3,974.52	1,603.71	1,713.77	1,542.71	1,692.02
Bayesian Inf. Crit.	4086	4051.7	1718.4	1768.7	1651.1	1734
McFadden pseduo R^2	0.368	0.366	0.218	0.18	0.157	0.142
Hosmer Lemeshow goodness of fit- p value	0.073	0.162	0.877	0.553	0.218	0.982
Note:					*p<0.1; **p<0.05; ***p<0.01	
					Main fuel implies LPG use as primary or exclusive fuel	

Table SI13: Summary of binary logistic regression for all three transition nodes using only categorical variables

SI5: Inspection of the independent variables and the outcome of LPG transition choices

In addition to wealth index score and urbanicity, we have plotted the association (and statistical results) among the other eleven predictor variables and LPG transition choice outcomes as shown here. The figures were created using the *ggstatsplot* package in R platform (Patil, 2021).

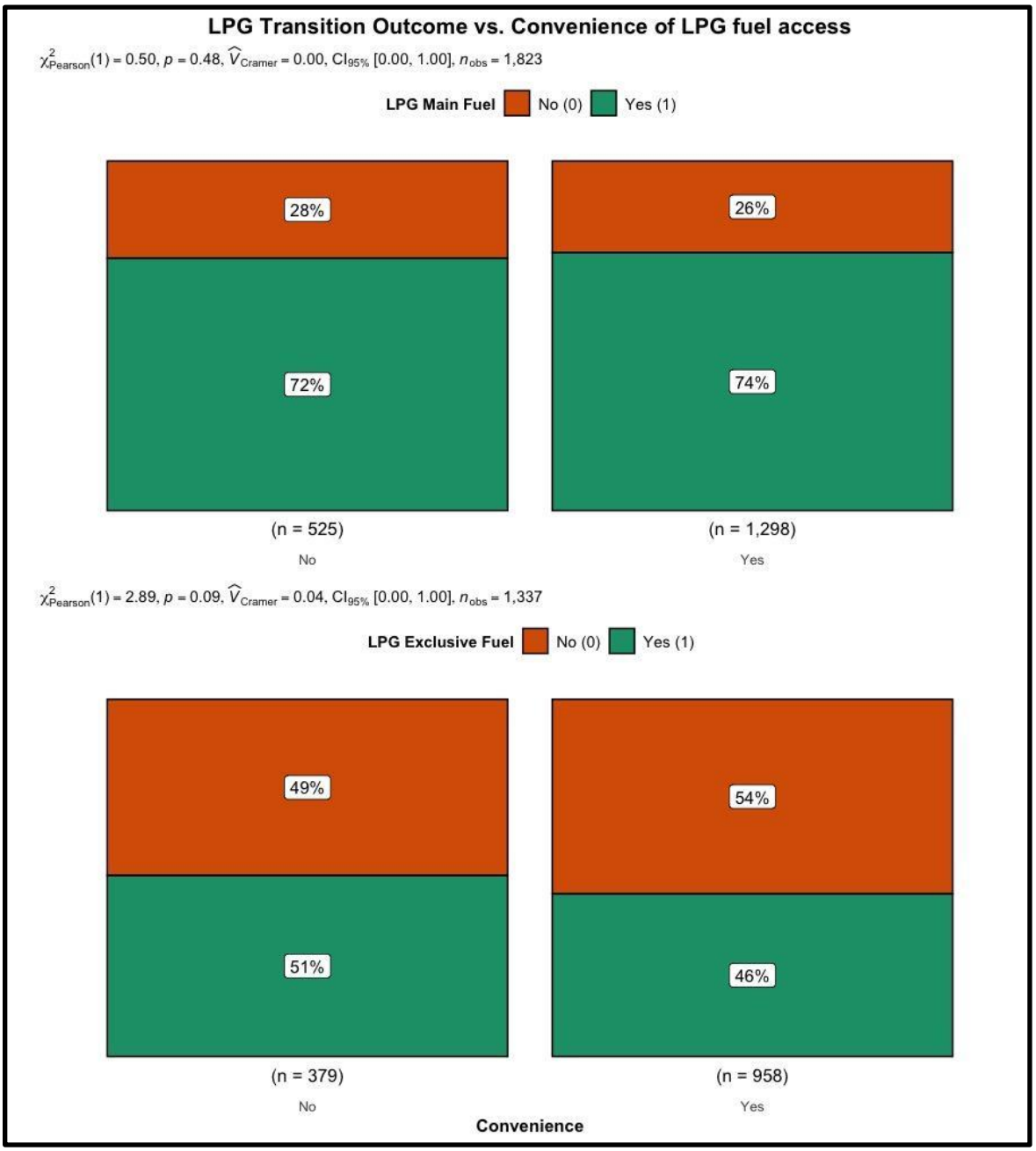


Figure SI14: Association between convenience to access LPG fuel and LPG transition choice outcomes. Only LPG users are considered.

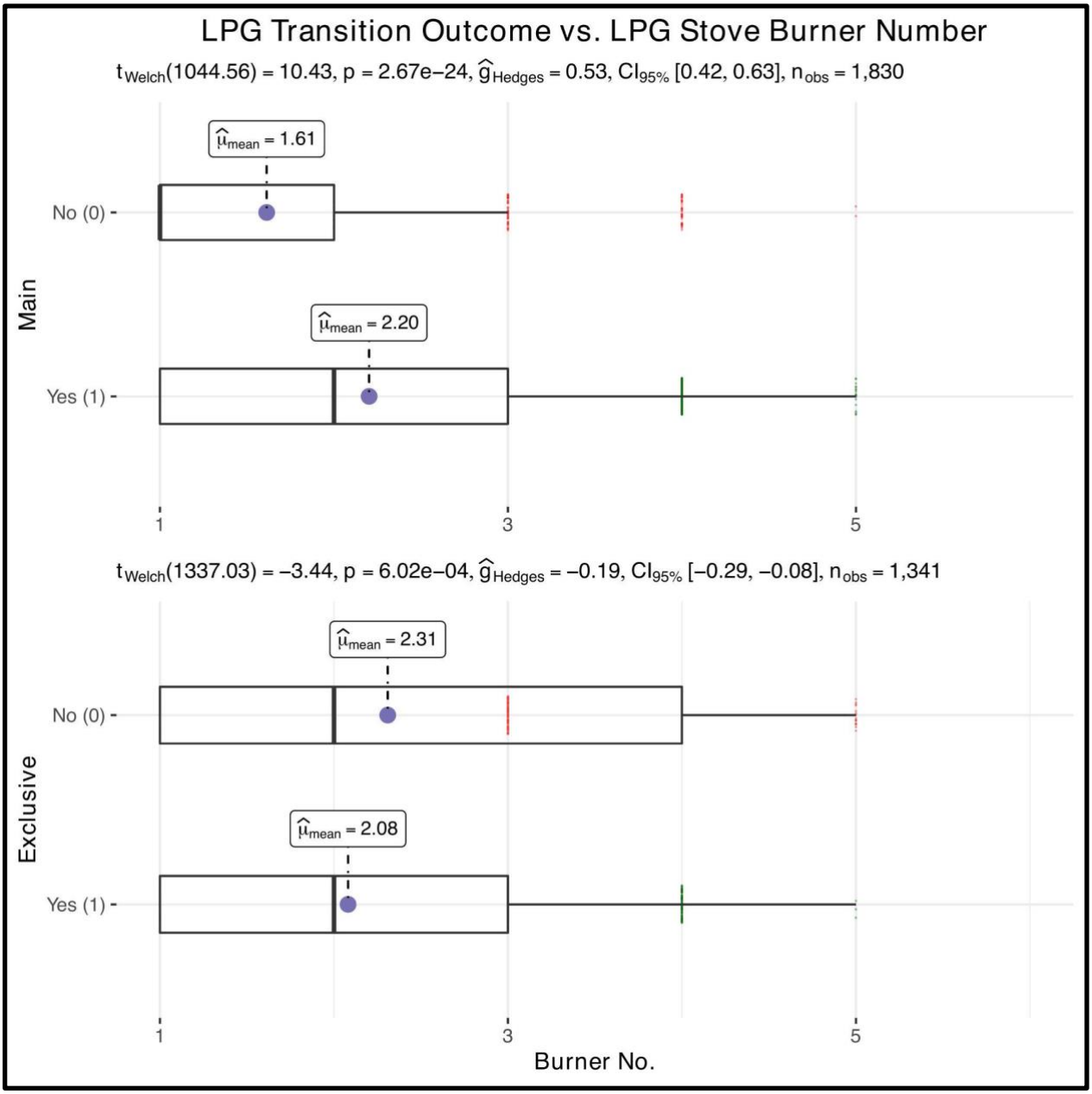


Figure SI15: Association between the number of burners in LPG stove and LPG transition choice outcomes. Only considered for LPG users.

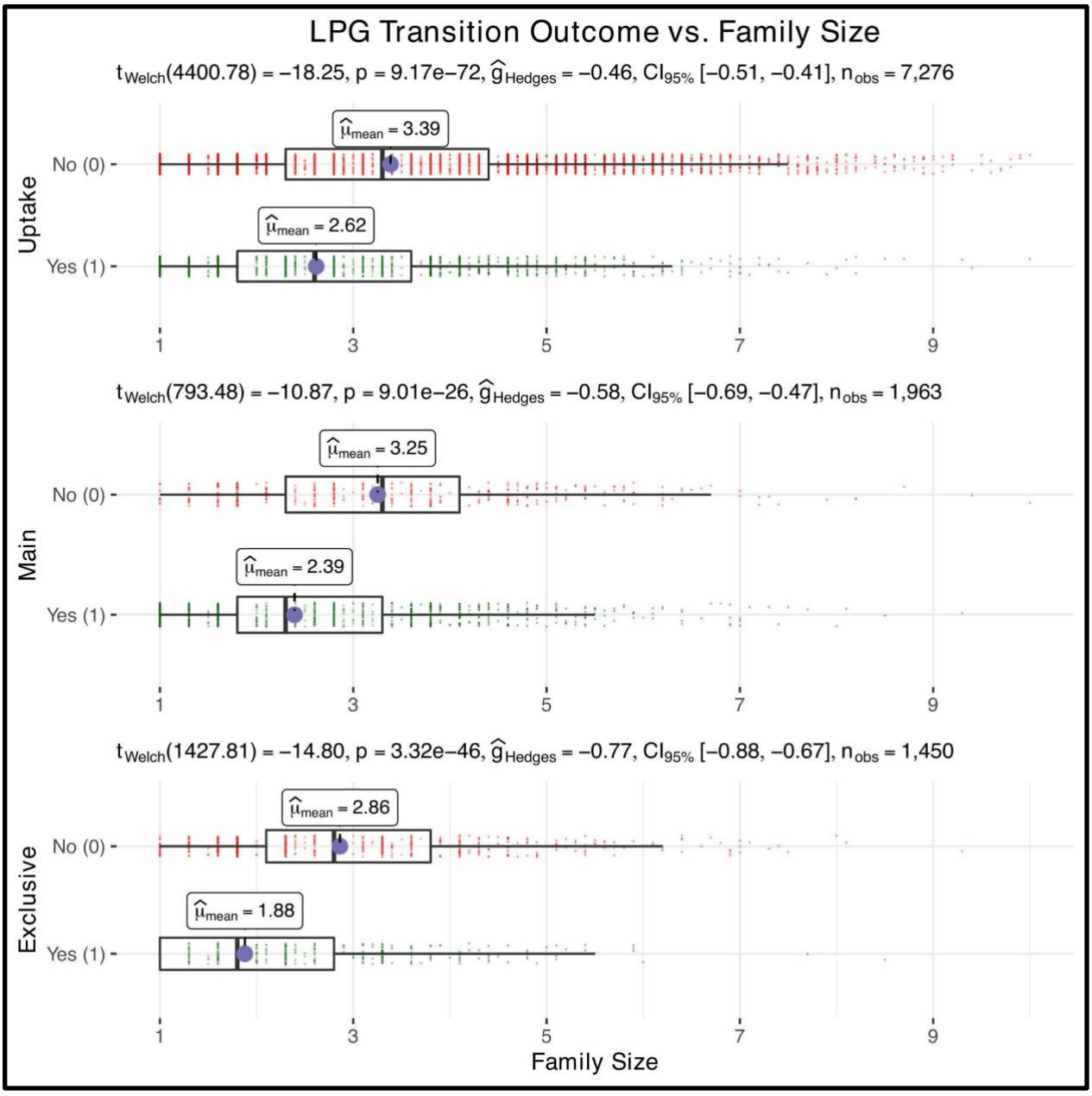


Figure SI16: Association between family size (standard adult) and LPG transition choice outcomes

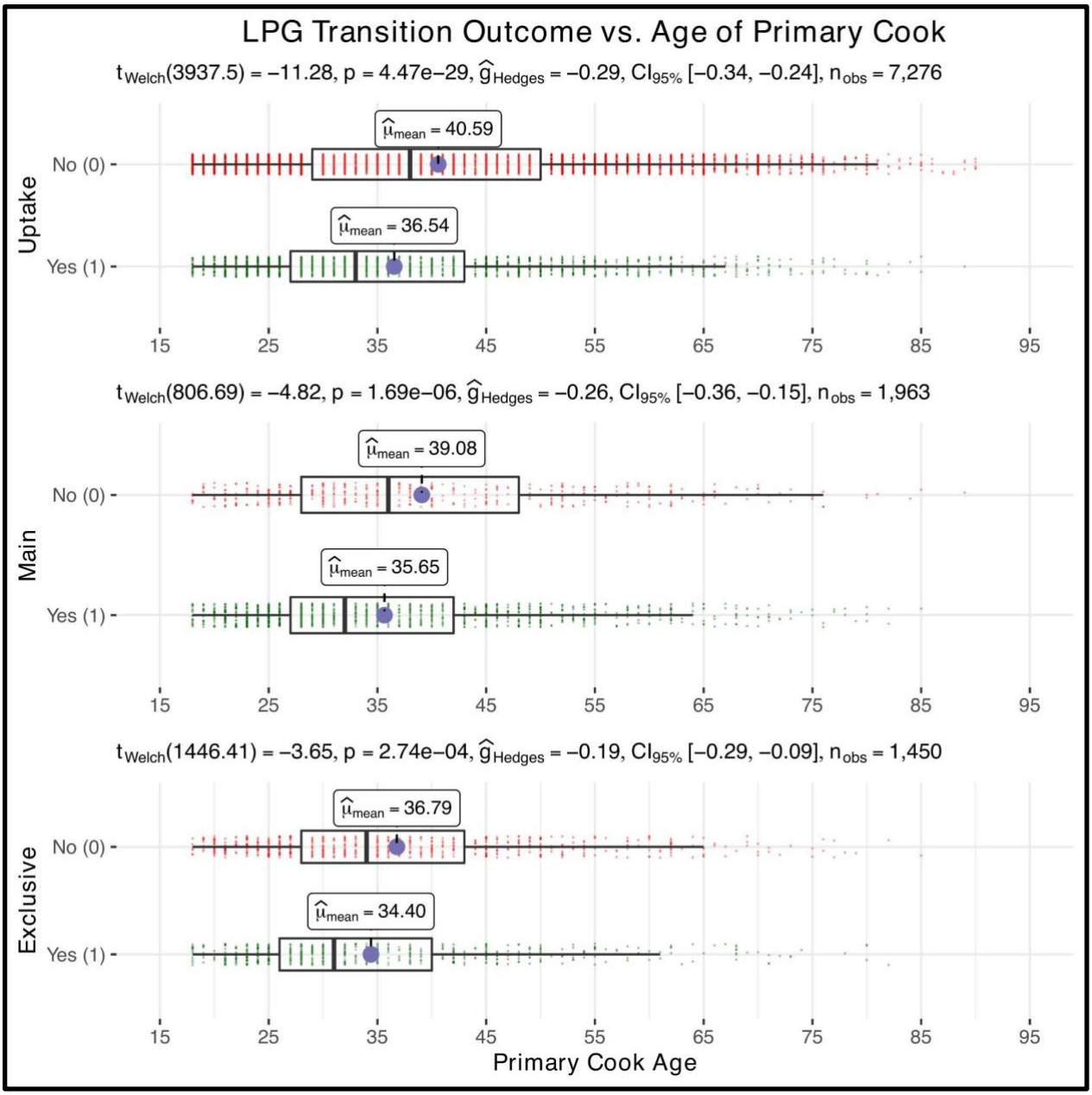


Figure SI17: Association between primary cook age and LPG transition choice outcomes

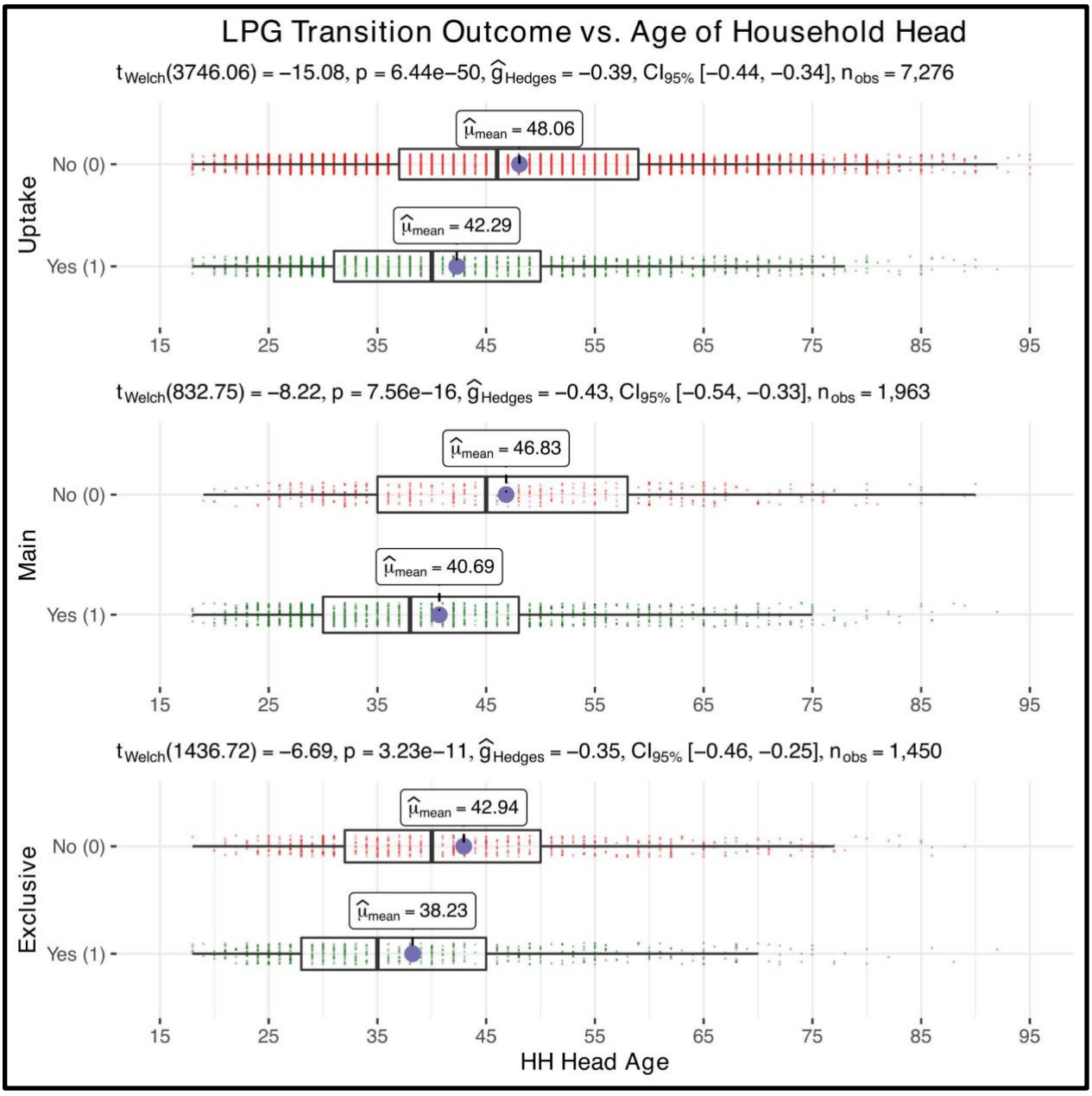


Figure SI18: Association between Household Head Age and LPG transition choice outcomes

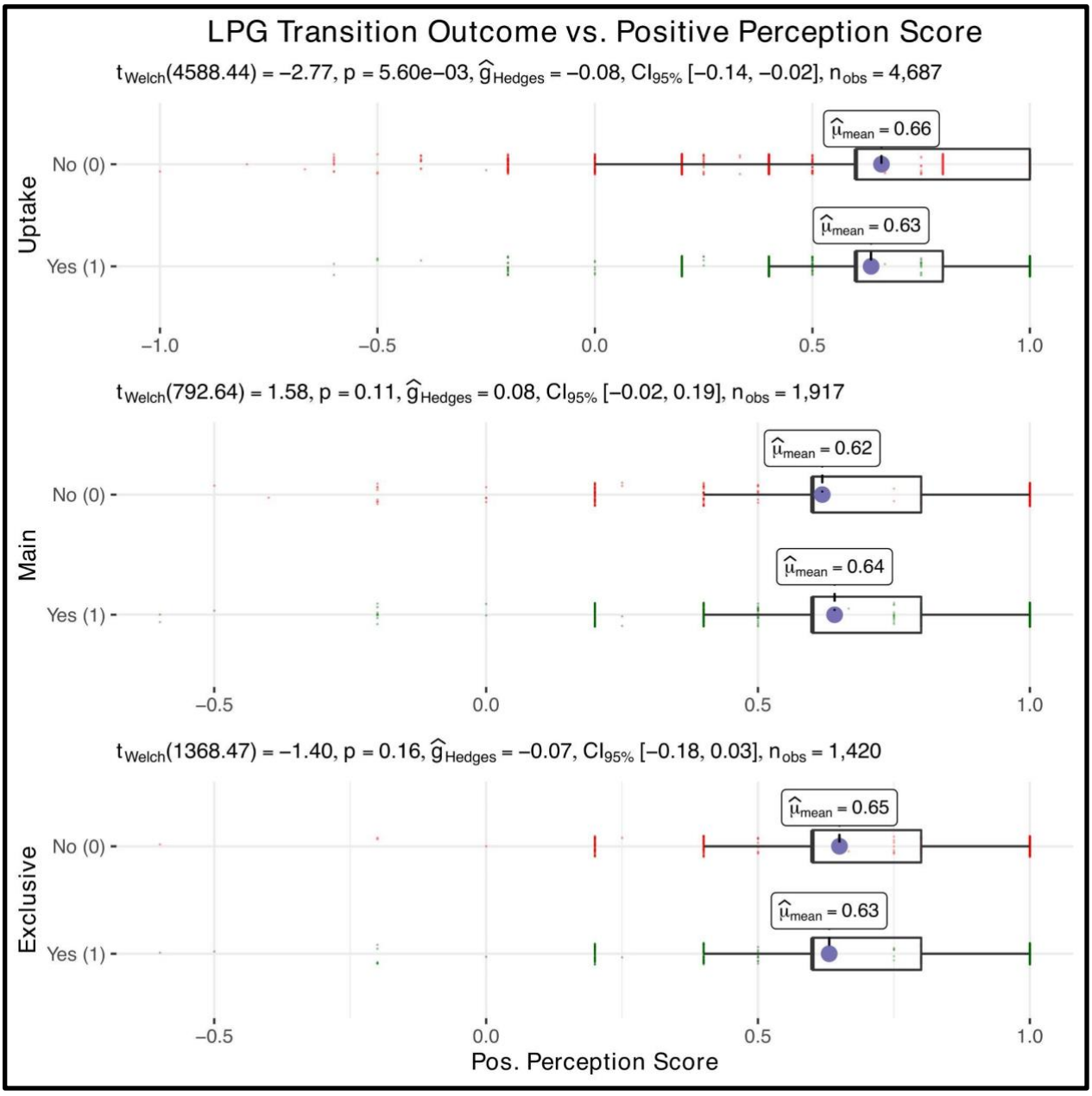


Figure SI19: Association between positive perception score and LPG transition choice outcomes

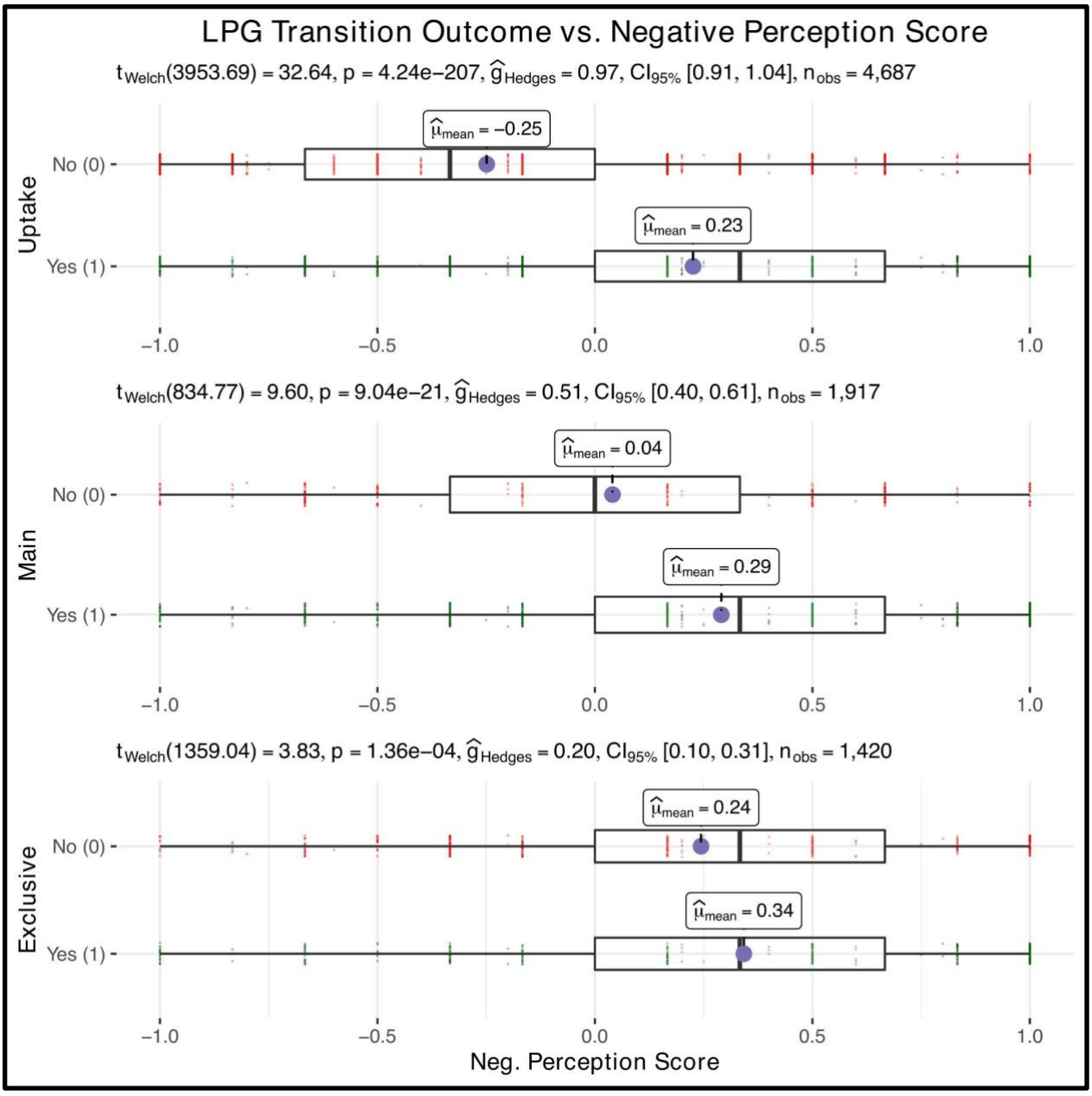


Figure SI20: Association between negative perception score and LPG transition choice outcomes

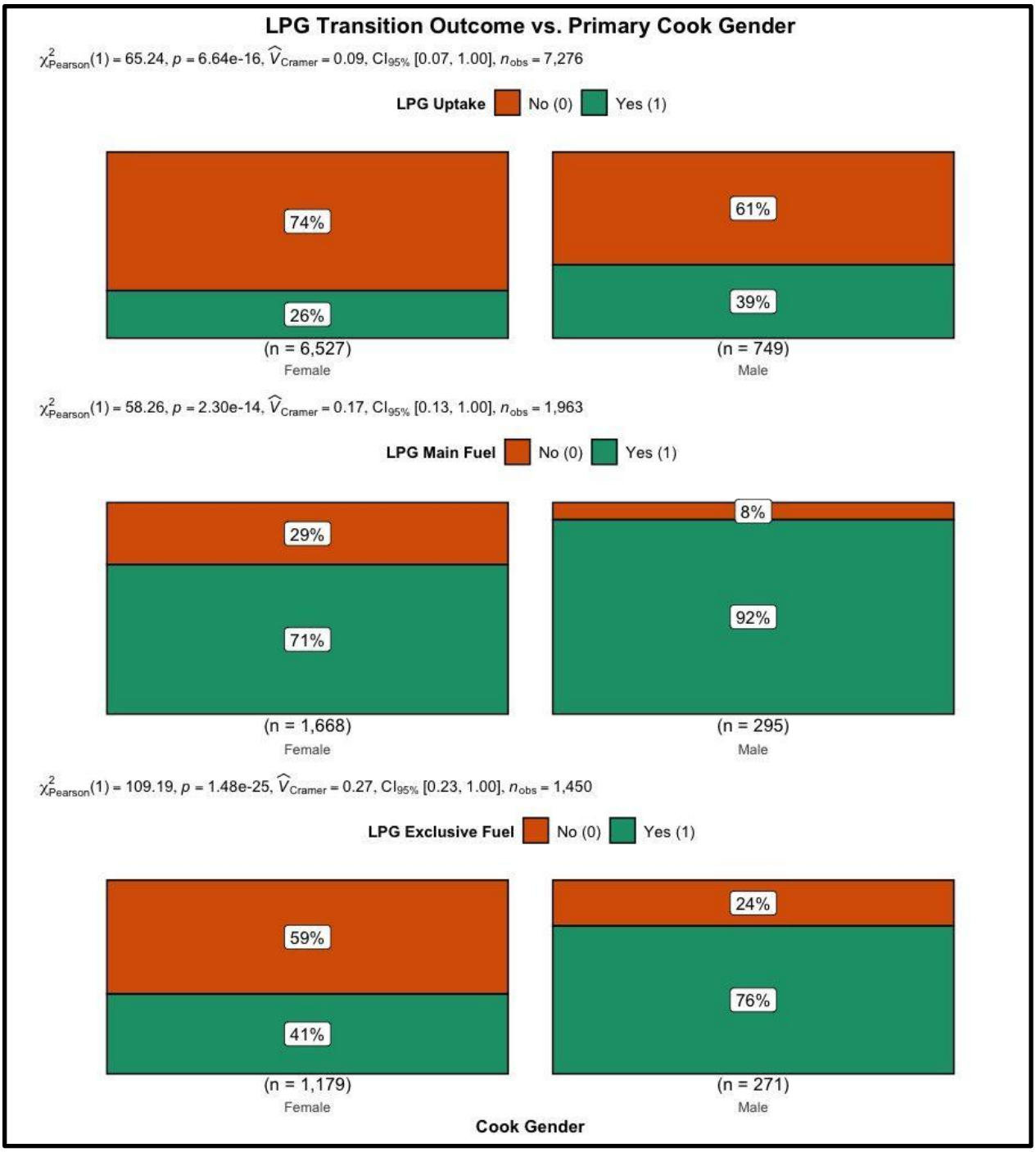


Figure SI21: Association between primary cook gender and LPG transition choice outcomes

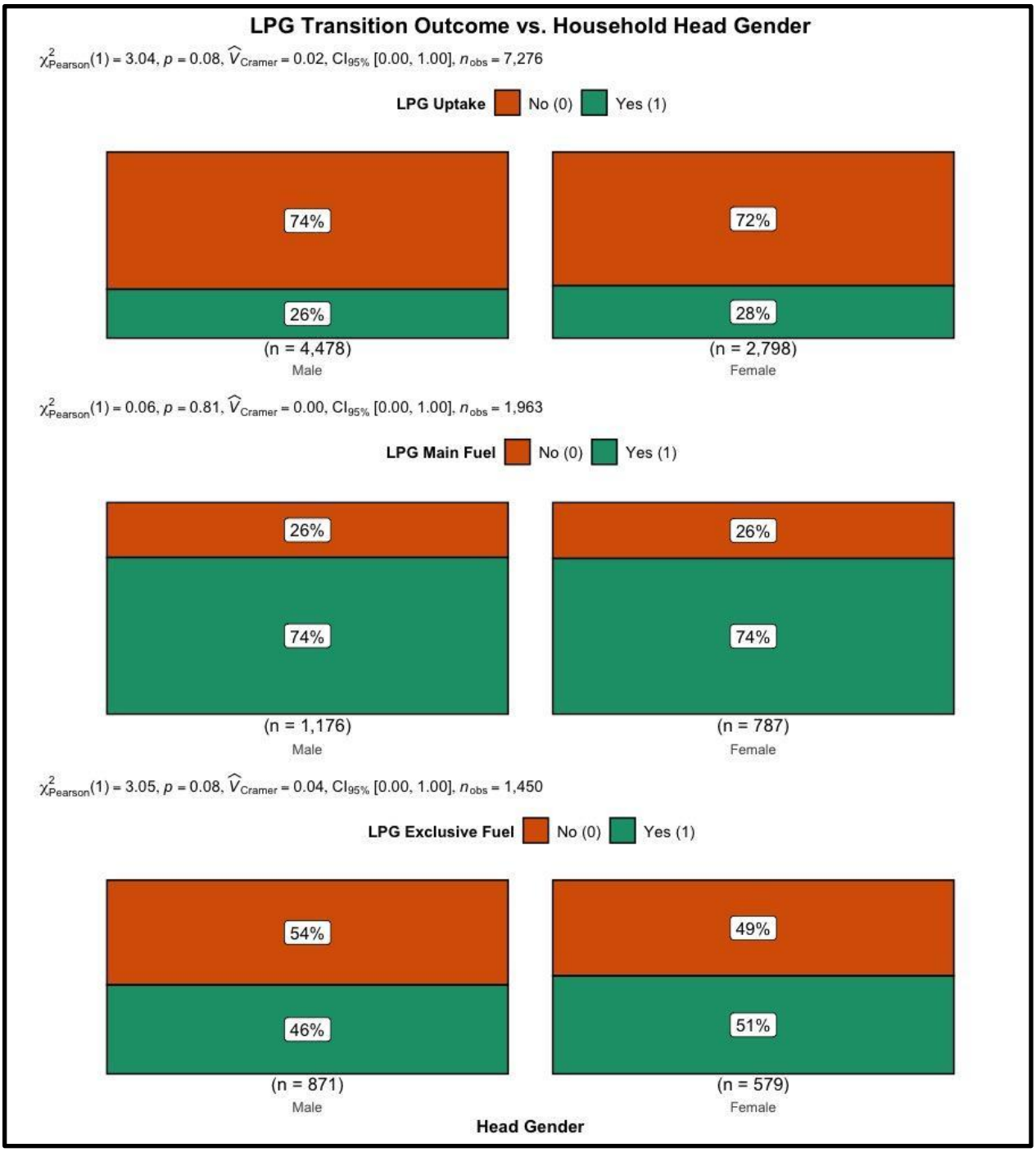


Figure SI22: Association between household head gender and LPG transition choice outcomes

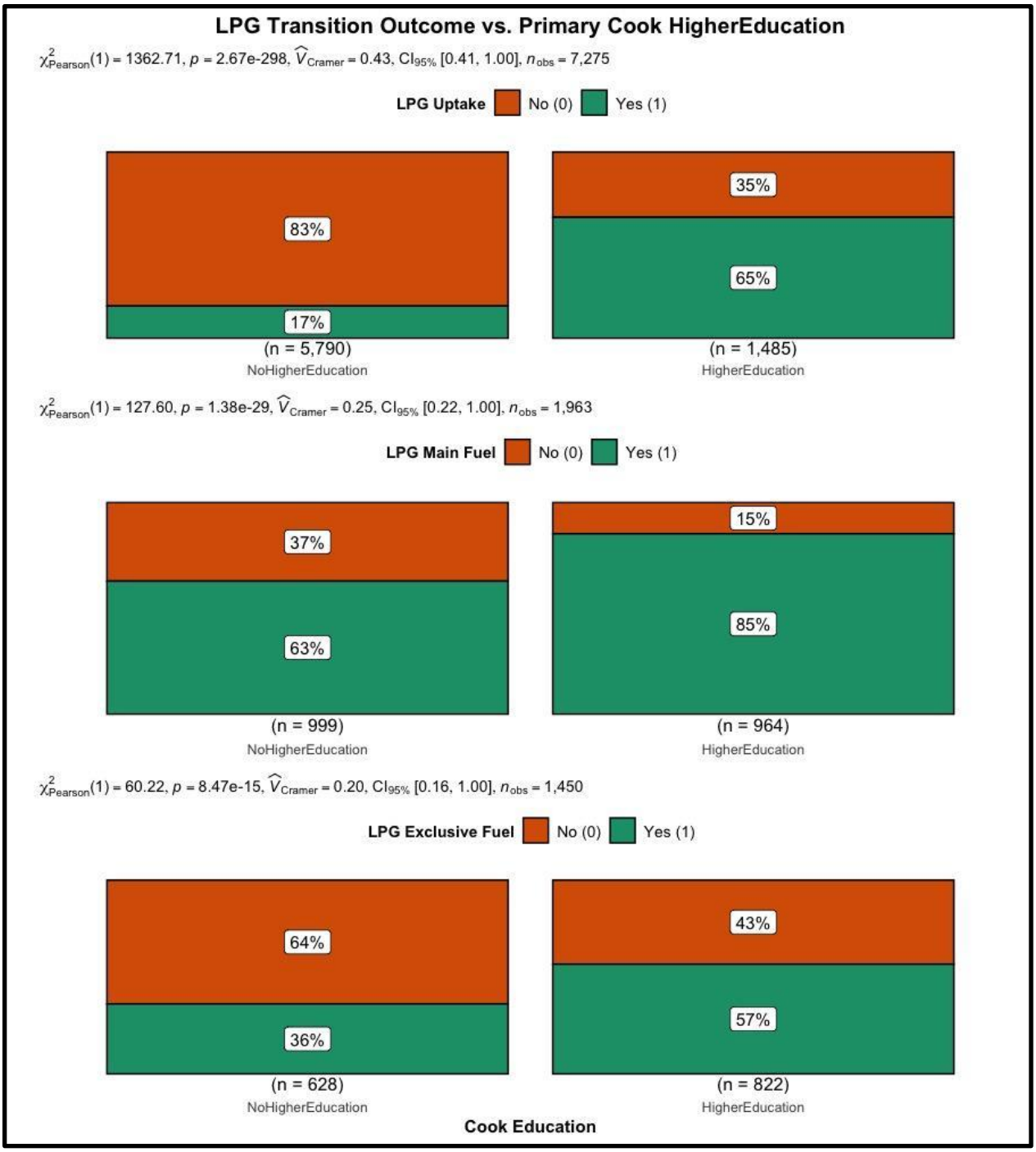


Figure SI23: Association between primary cook higher education and LPG transition choice outcomes

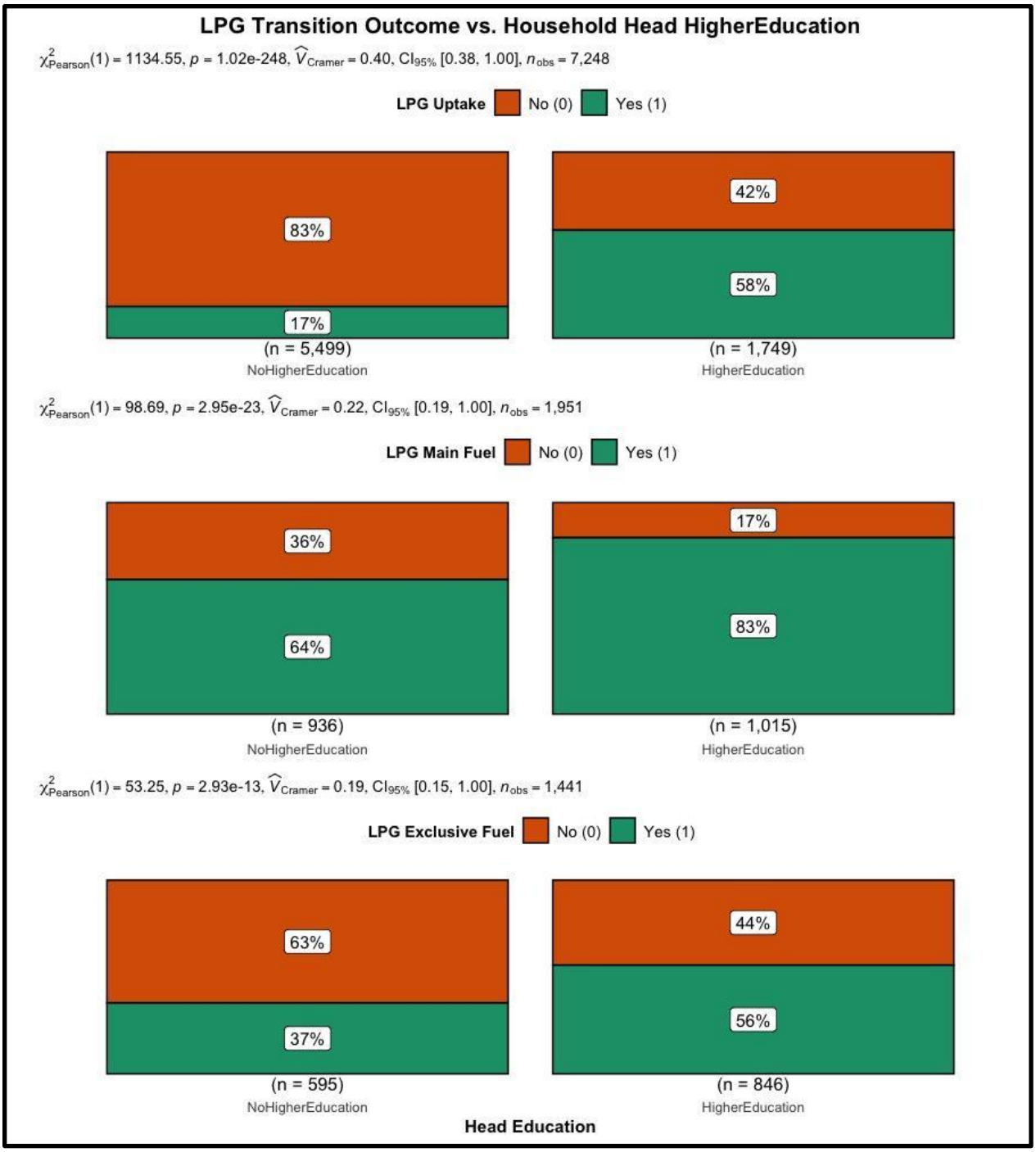


Figure SI24: Association between household head education and LPG transition choice outcomes

SI6: Discussion on LPG stacking

The national (Figure SI27) and urban/rural (Figure SI28) breakup of cooking technology stacking pattern provide insight into stove stacking in Ghana. Literature suggests that household cooking in low- and middle-income countries often exhibits considerable stove stacking (when households have two or more stoves at home). To our knowledge, our data collection effort is the first to characterize stacking in Ghana in a nationally representative sample. We measured stacking by asking primary cooks for the quantity and type of every functioning stove owned by the household. Following this information, we asked respondents to rank their stoves from the most to least commonly used in their day-to-day cooking activities. We find that the majority of households that cook primarily with three-stone or charcoal stoves do not stack stoves (66% and 71%, respectively), meaning that they do not report using a secondary stove. Households that cook primarily with a three-stone stove most often report a charcoal stove as their secondary stove type, while charcoal stove users are roughly equally divided between those who stack with LPG and with three-stone stoves. Finally, LPG owners are the most likely to have a secondary stove, usually charcoal. Figure 3 depicts these results.

Our sample is about equally split between rural and urban households. We observe more stove stacking in urban areas. The main difference happens with charcoal users who have a secondary stove. Those in urban areas generally use LPG stoves as their second option, whereas in rural areas, three-stone stoves are more prevalent.

Given that the vast majority of households use a charcoal stove, three-stone fire, or LPG stove, we combined the remaining options (such as electric stoves) into an “other stove” category. Notably, discussions on stove stacking are limited to primary and secondary stoves as 97% of households report that they do not have any tertiary stoves. Also, while these questions were asked with respect to stove use during the dry season when the surveys were conducted, they remained representative of year-round stove use patterns in Ghana, as 94% of respondents reported not changing their primary stove in the wet season.

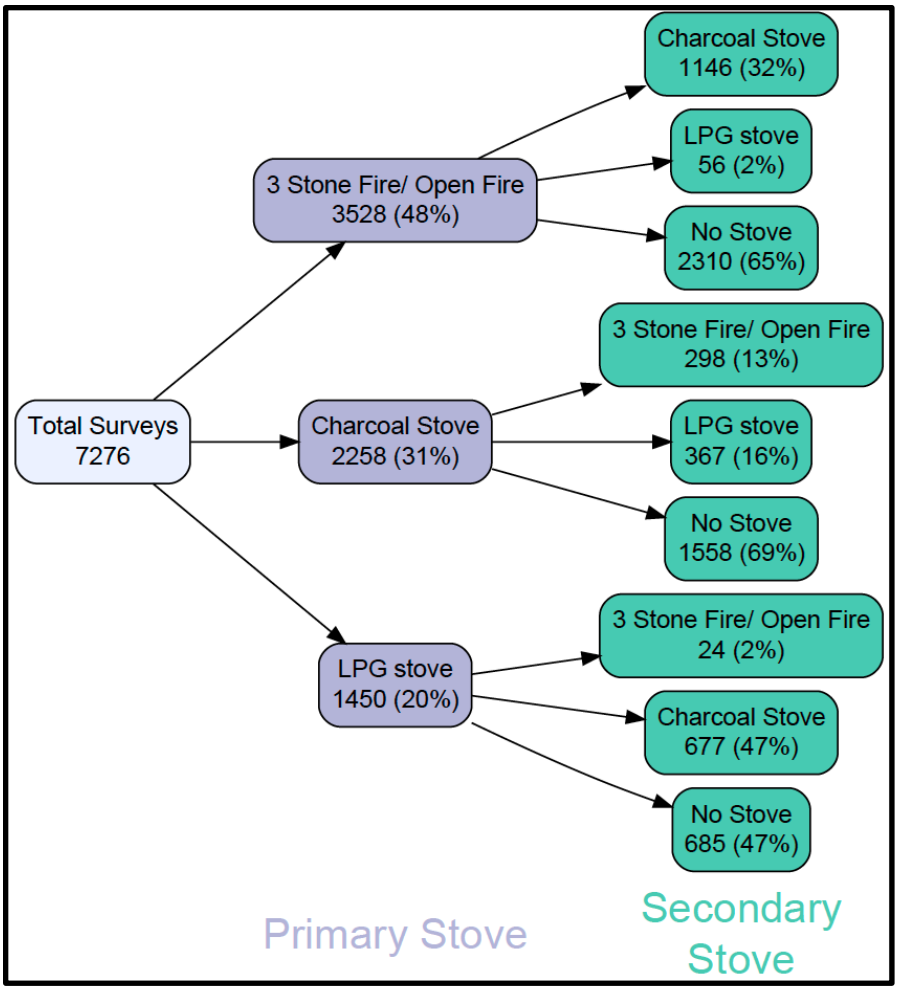


Figure SI25: Breakup of stove use pattern at the national level. For the purpose of readability, the figure does not display choices that are less than 1%. Hence, only three main cooking technologies are shown as the primary stove. Other cooking technologies and 'No Stove/ No Data' is not displayed for the primary stove. Similarly, other cooking technologies and 'primary/secondary same stove' are not displayed for the secondary stove.

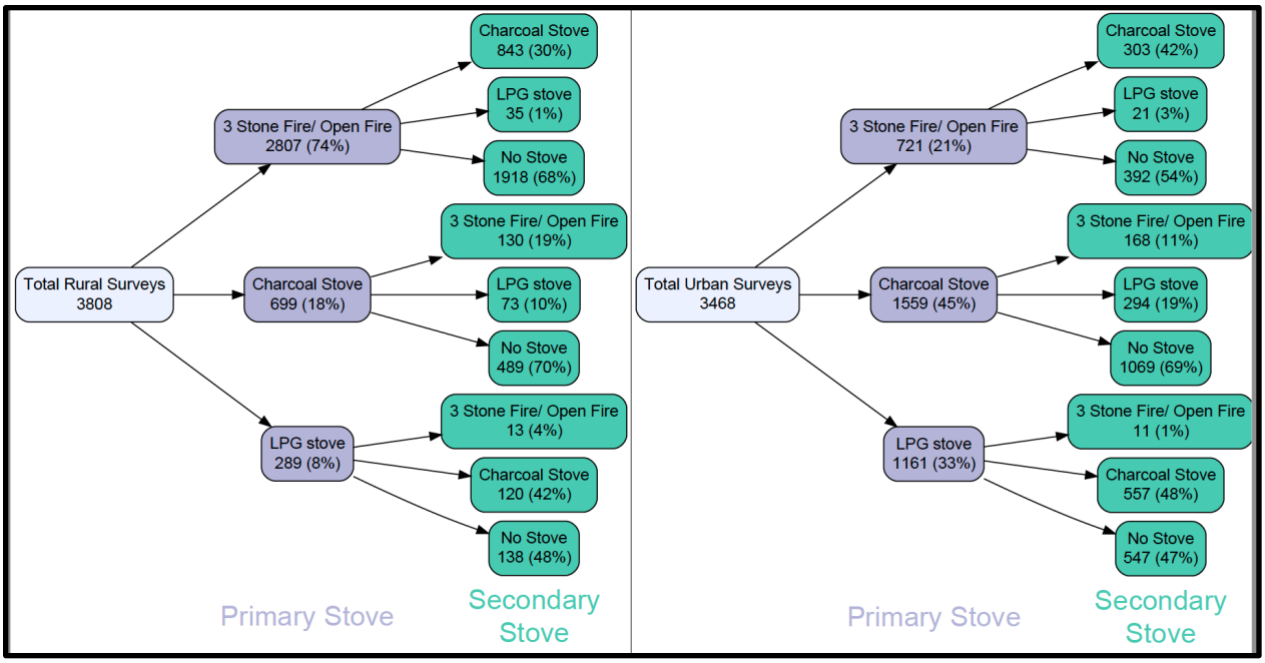


Figure SI26: Breakup of stove use pattern at urban (left panel) and rural (right panel) levels. For the purpose of readability, the figure does not display choices that are less than 1%. Hence, only three main cooking technologies are shown as the primary stove. Other cooking technologies and 'No Stove/ No Data' is not displayed for the primary stove. Similarly, other cooking technologies and 'primary/secondary same stove' are not displayed for the secondary stove.

References

Bailis, R. (2007). *Kitchen Performance Test (KPT) Version 3.0*. https://cleancooking.org/binary-data/DOCUMENT/file/000/000/83-1.pdf

Hlavac, M. (2022). *Stargazer: Well-Formatted Regression and Summary Statistics Tables. Social Policy Institute, Bratislava, Slovakia. R package version 5.2.3*. https://CRAN.R-project.org/package=stargazer

Kar, A., Brauer, M., Bailis, R., & Zerriffi, H. (2020). The risk of survey bias in self-reports vs. Actual consumption of clean cooking fuels. *World Development Perspectives*, *18*, 100199. https://doi.org/10.1016/j.wdp.2020.100199

Patil, I. (2021). Visualizations with statistical details: The “ggstatsplot” approach. *Journal of Open Source Software*, *6*(61), 3167. https://doi.org/10.21105/joss.03167

412 Solt, F., Hu, Y., Keyes, O., Bolker, B., Müller, S., Leeper, T., Wallace, C., & Warshaw, C.
413 (2022). *dotwhisker: Dot-and-Whisker Plots of Regression Results*. https://cran.r-
414 project.org/web/packages/dotwhisker/index.html

Annexure 1:

Wealth Index Analysis

National Household Needs Assessment

Socio-Economic Status

June 22 2022

Contents

List of Tables

List of Figures

1 Initial Household SES Variables

1.1 Variables selected for first PCA estimation

The variables selected for generating the wealth index is derived from household's durable asset ownership as well as household utilities and facilities. Table 1 provides the categories and some examples of the variables used for computing the initial wealth index.

Table 1: Variables included in wealth index

Productive Assets	**Non-productive Assets**	**Household Facilities**
Livestock (bulls)	Table	Land ownership
Livestock (cattle)	Mattress	House ownership
Chicken	Radio	Main light source
Other poultry	TV (black, color)	Main water source
Goats	Room divider, cupboard	Main toilet facility
Sheep	Satellite dish	Floor material
Pigs	Fridge, freezer	Wall material
Grass cutter	Phone, camera	Roof material
Rabbit	Bicycle, motor	Cooking space
Horse, donkey, cart	Personal car	
Tractor	Commercial vehicle	
	Electric appliances	

1.2 Recoding Variables

1.2.1 Variables coded as 0

- Missing
- Don't know
- Don't want to answer
- Other (in some instances)

Example 1 In the example in Table 2, the first question (left-hand side sub-table) is "Does the HH currently own any livestock, herds, other farm animals?". Several participants respond with 'i don't want to know' and 'i don't want to answer'. This shows up in the subsequent questions as missing. For example in the subsequent question of 'how many bulls does your household own' (right-hand side sub-table).

Table 2: Responses to household ownership of livestock

	Frequency	Share
I Don't Know	11	0.15
I Don't Want to Answer	2	0.03
No	5354	72.71
Yes	1996	27.11
Total	7363	100.00

	Frequency	Share
1	16	0.22
2	13	0.18
3	7	0.10
4	2	0.03
5	3	0.04
7	2	0.03
8	1	0.01
20+	1	0.01
.	7318	99.39
Total	7363	100.00

I recode the missing observations from the previous table as zeros in Table 3.

Table 3: Summary statistics of variables used for asset index

	Frequency	Share
0	7318	99.39
1	16	0.22
2	13	0.18
3	7	0.10
4	2	0.03
5	3	0.04
7	2	0.03
8	1	0.01
20	1	0.01
Total	7363	100.00

I recode the response to the question on land ownership; but subsequently drop this variable from the analysis as the question was randomized.

1.2.2 Categorical variables

- Recode categorical variables as binary

Table 4: Responses to household ownership of land

	Frequency	Share
I Don't Know	3	0.04
Acres	124	1.68
Hectares	4	0.05
Plot	405	5.50
.	6827	92.72
Total	7363	100.00

	Frequency	Share
0	6830	92.76
1	533	7.24
Total	7363	100.00

Example 2

- Recode categorical variables as new categorical variables

Example 3 The options for house ownership were: 1, Sole Ownership | 2, Joint Ownership | 3, Family/relation's house | 4, House provided rent free for unlimited time | 5, Perching | 6, Renting | -8, Other | -99, I don't know | -88, I don't want to answer.

I recoded to (-99=0) (-88=0) (-8=1) (3/5=1) (6=2) (1/2=3). 'Other' was also coded as '1' because all the options provided fell under government housing or mission housing.

Please see the do file for more details on the recoded variables.

1.3 Analysis

- Review and revise variables included in principal component analysis (PCA)
- First principal component is the wealth index
- Choice of using index as continuous or categorical (i.e. wealth quintiles)

1.4 Initial SES Index

Table 5 shows the distribution of the socio-economic status across the study households. This index is from the initial exploration with all potential variables that. The sample of households for which there is an SES index is a mere 3,622, just about half of the national sample. This brings up the question of how to revise the selected variables. I address this in the PCA post estimation section.

Table 5: Socio-economic status

	Frequency	Share
least poor	724	19.99
less poor	724	19.99
poor	725	20.02
more poor	724	19.99
very poor	725	20.02
Total	3622	100.00

Table 6 presents the summary statistics of the variables used.

Table 6: Summary statistics of variables used for asset index

	Mean	Standard deviation	Min	Max
bull	0.02	0.33	0	20
cat	0.17	2.03	0	100
chick	3.34	9.32	0	100
o_pltry	0.32	2.92	0	100
gt	1.04	3.41	0	50
shp	0.57	2.98	0	100
pg	0.13	1.39	0	50
rbt	0.03	0.61	0	30
grsct	0.00	0.20	0	16
hors	0.01	0.15	0	5
tv_bl	0.04	0.27	0	9
tv_c	0.63	0.63	0	7
cam	0.02	0.22	0	8
was	0.02	0.17	0	8
gen	0.01	0.19	0	9
clck	0.21	0.48	0	10
mat	1.31	1.11	0	10
div	0.36	0.66	0	9
rad	0.57	0.56	0	10
tvd	0.49	0.61	0	9
tabl	1.33	1.13	0	10
phn	1.63	1.27	0	10
frd	0.34	0.56	0	10
frz	0.08	0.31	0	4
bic	0.25	0.58	0	10
cmv	0.05	0.25	0	6
car	0.05	0.27	0	5
trt	0.01	0.17	0	10
fn	0.83	1.02	0	10
ac	0.02	0.23	0	7
cmp	0.10	0.36	0	9
irn	0.43	0.58	0	9
bat	0.01	0.18	0	10
sm	0.13	0.44	0	10
str	0.12	0.40	0	10
ktl	0.08	0.33	0	9
ckr	0.01	0.17	0	10
tstr	0.02	0.23	0	10
r_ckr	0.07	0.27	0	4
mwv	0.04	0.24	0	9
stv	0.01	0.17	0	10
lnd	0.07	0.26	0	1
hown	2.97	1.88	0	6
kit	0.53	0.57	0	2
lgt	0.81	0.39	0	1
flr	0.87	0.34	0	1
wll	0.40	0.49	0	1
rof	0.46	0.50	0	1
wat	2.22	0.77	0	3
toi	0.94	0.61	0	2

1.5 Initial PCA Post-estimation

Table 7 presents the output of standard methods for studying correlation matrices to assess whether the variables have strong linear relations with each other. These methods can be seen as pre-estimation rather than as post-estimation methods. The first column contains the output for the inspection of the squared multiple correlation (the regression R^2) of each variable on all other variables. The SMC measures help identify variables that cannot be explained well from the other variables, and can help reevaluate whether they should be included in the analysis. Some of the SMC are small but may not warrant exclusion.

The variables with the smallest SMC and under consideration for exclusion are

- animals (grass cutter, pig, rabbit)
- camera
- battery (inverter)
- stove top pressure cooker and
- kitchen space.

The second column of Table 7 contains the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy, which compares the correlations and the partial correlations between variables. If the partial correlations are relatively high compared to the correlations, the KMO measure is small, and a low-dimensional representation of the data is not possible.

Using the Kaiser (1974) characterization of KMO values,

- 0.00 to 0.49 unacceptable
- 0.50 to 0.59 miserable
- 0.60 to 0.69 mediocre
- 0.70 to 0.79 middling
- 0.80 to 0.89 meritorious
- 0.90 to 1.00 marvelous

we can say that our KMO value, 0.90 is marvelous.

Table 7: Squared Multiple Correlation and Kaiser-Meyer-Olkin

Variable Name	**Variable**	**SMC**	KMO
Livestock (bulls)	bull	0.15	0.55
Livestock (cattle)	cat	0.16	0.61
Chicken	chick	0.28	0.75
Poultry (other)	o_pltry	0.18	0.71
Goats	gt	0.25	0.75
Sheep	shp	0.31	0.69
Pigs	pg	0.10	0.62
Rabbit	rbt	0.03	0.62
Grass cutter	grsct	0.02	0.50
Horse/donkey/cart	hors	0.15	0.60
TV (black and white)	tv_bl	0.16	0.36
Color TV	tv_c	0.58	0.91
Camera	cam	0.06	0.93
Washing machine	was	0.19	0.95
Generator	gen	0.42	0.90
Clock	clck	0.26	0.96
Mattress	mat	0.51	0.93
Cupboard/wardrobe/room divider	div	0.41	0.96
Radio	rad	0.14	0.91
Satellite dish	tvd	0.55	0.92
Table	tabl	0.31	0.95
Phone	phn	0.42	0.93
Fridge	frd	0.44	0.96
Freezer	frz	0.27	0.95
Bicycle	bic	0.17	0.80
Commercial vehicle	cmv	0.22	0.92
Personal car	car	0.38	0.89
Tractor	trt	0.42	0.80
Fan	fn	0.60	0.96
Air conditioning	ac	0.25	0.89
Computer (desktop/laptop)	cmp	0.33	0.95
Electric iron	irn	0.50	0.96
Battery (inverter)	bat	0.03	0.82
Sewing machine	sm	0.11	0.84
Store/shop/kiosk	str	0.18	0.93
Electric kettle	ktl	0.40	0.95
Electric pressure cooker	ckr	0.64	0.77
Electric toaster	tstr	0.51	0.91
Rice cooker	r_ckr	0.30	0.94
Microwave oven	mwv	0.29	0.93
Stovetop pressure cooker	stv	0.03	0.90
Land ownership	lnd	0.06	0.86
House ownership	hown	0.10	0.87
Kitchen space	kit	0.08	0.86
Main light source	lgt	0.31	0.92
Floor material	flr	0.34	0.82
Wall material	wll	0.42	0.86
Roofing material	rof	0.22	0.84
Main water source	wat	0.29	0.94
Toilet	toi	0.35	0.94
Overall			**0.90**

The scree plot in Figure 1 displays the number of the principal component versus its corresponding eigenvalue. We can use the scree plot to select the number of components to use based on the size of the eigenvalues. The ideal pattern is a steep curve, followed by a bend, and then a straight line. It's recommended to use the components in the steep curve before the first point that starts the line trend. The first component is used to predict the socio-economic status for our sample. This is because the first accounts for a large proportion while the remaining principal components account for a very small proportion of the variability (close to zero) and are probably unimportant.

Figure 1: Scree plot of initial PCA

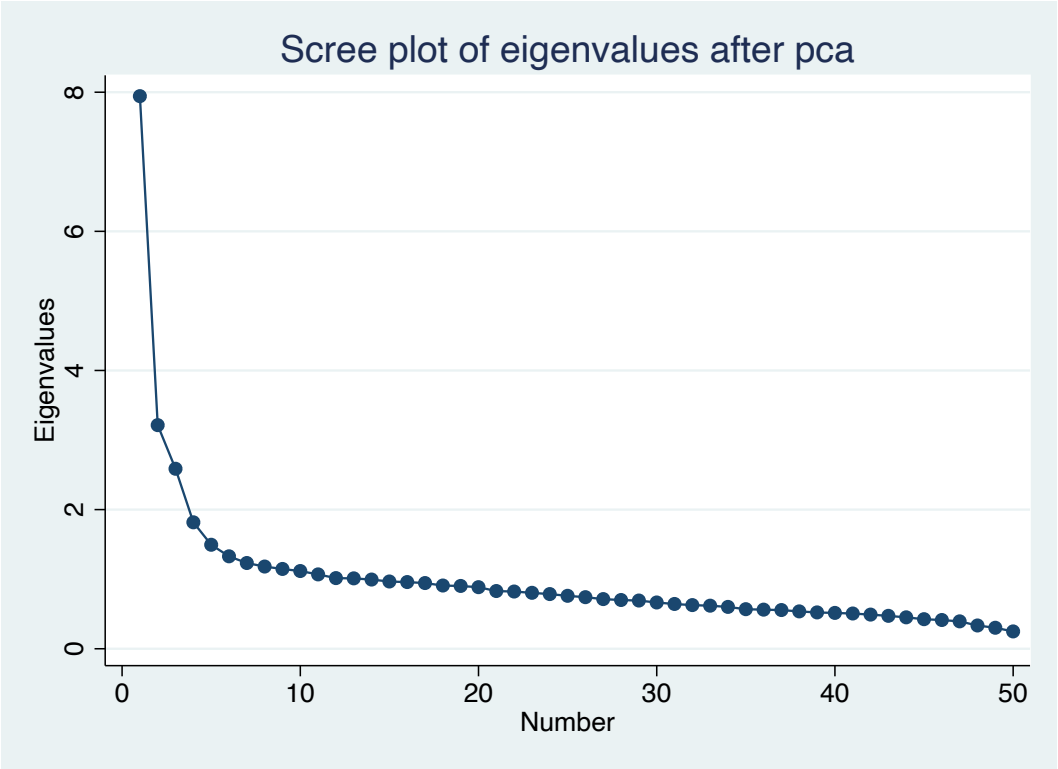
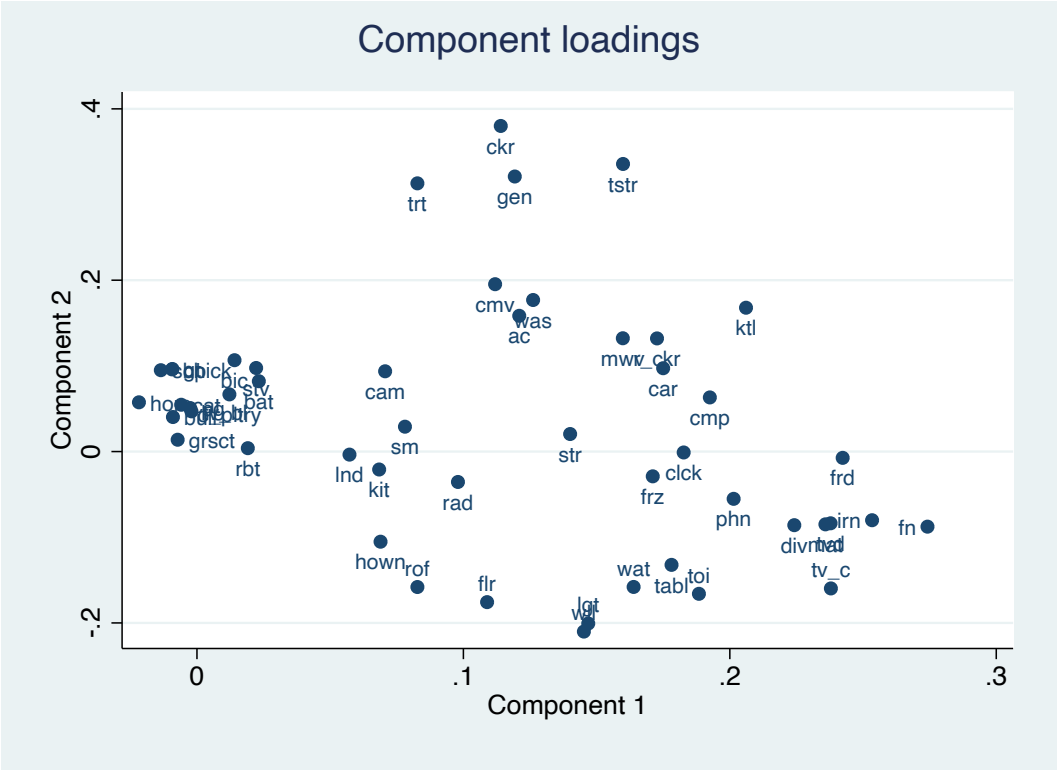


Figure 2 displays the loading plot and is used to identify which variables have the largest effect on each component. Loadings can range from -1 to 1. Loadings close to -1 or 1 indicate that the variable strongly influences the component. Loadings close to 0 indicate that the variable has a weak influence on the component.

Figure 2: Loading plot of initial PCA



Based on the loading plot, the variables under consideration for exclusion are

- animals (grass cutter, chicken, rabbit, other poultry, horse)
- battery (inverter)
- stove top pressure cooker
- bicycle

2 Revised PCA and SES Index

From the initial PCA post estimation, I exclude variables with the smallest SMC and lowest effect on the first component loadings. I also exclude materials used to construct wall, roof and floor of house, as they was randomly assigned to a subset of the sample - this was the cause of the SES sample size being available for just about 3,622 households. I re-run the PCA and subsequently exclude land ownership and house ownership as they have a close to zero effect on the first component loadings.

Figure 8 presents the revised SES index. The sample size here is larger, and the SES is available for 7,306 households.

Table 8: Socio-economic status

	Frequency	Share
least poor	1461	20.00
less poor	1461	20.00
poor	1461	20.00
more poor	1461	20.00
very poor	1462	20.01
Total	7306	100.00

2.1 Revised PCA Post-estimation

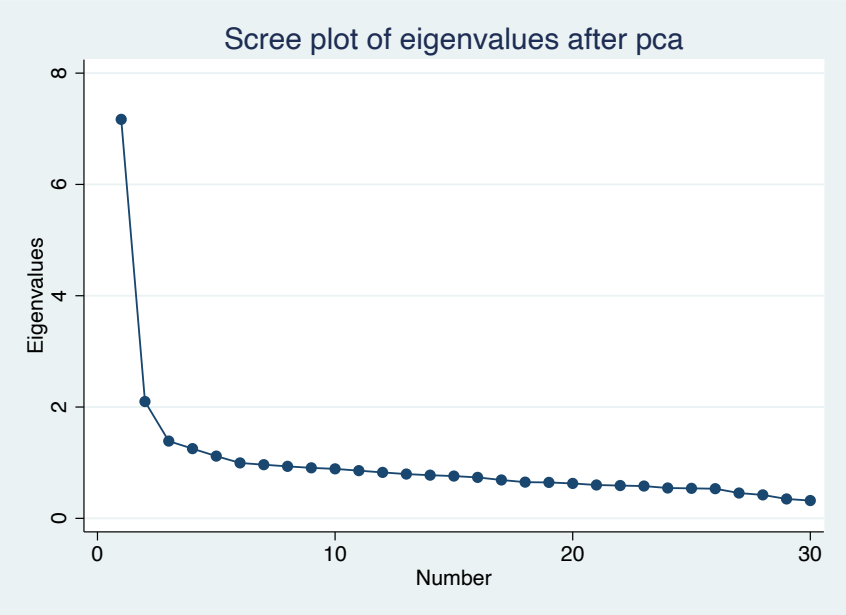
Table 9 presents the SMC and KMO for the revised PCA post estimation. The KMO of 0.93 is marvelous and higher than the initial KMO.

Table 9: Revise Squared Multiple Correlation and Kaiser-Meyer-Olkin

Variable Name	**Variable**	**SMC**	KMO
Color TV	tv_c	0.55	0.92
Camera	cam	0.04	0.92
Motorcycle	was	0.12	0.93
Generator	gen	0.22	0.83
Clock	clck	0.25	0.96
Mattress	mat	0.48	0.94
Cupboard/wardrobe/room divider	div	0.41	0.97
Radio	rad	0.11	0.92
Satellite dish	tvd	0.51	0.92
Table	tabl	0.30	0.95
Phone	phn	0.39	0.94
Fridge	frd	0.43	0.96
Freezer	frz	0.21	0.96
Commercial vehicle	cmv	0.10	0.90
Personal car	car	0.27	0.93
Fan	fn	0.60	0.96
Air conditioning	ac	0.19	0.88
Computer (desktop/laptop)	cmp	0.31	0.95
Electric iron	irn	0.47	0.96
Sewing machine	sm	0.09	0.86
Store/shop/kiosk	str	0.13	0.94
Electric kettle	ktl	0.25	0.95
Electric pressure cooker	ckr	0.18	0.75
Electric toaster	tstr	0.13	0.87
Rice cooker	r_ckr	0.27	0.93
Microwave oven	mwv	0.29	0.92
Kitchen space	kit	0.06	0.92
Main light source	lgt	0.24	0.92
Main water source	wat	0.25	0.93
Toilet	toi	0.33	0.94
Overall			**0.93**

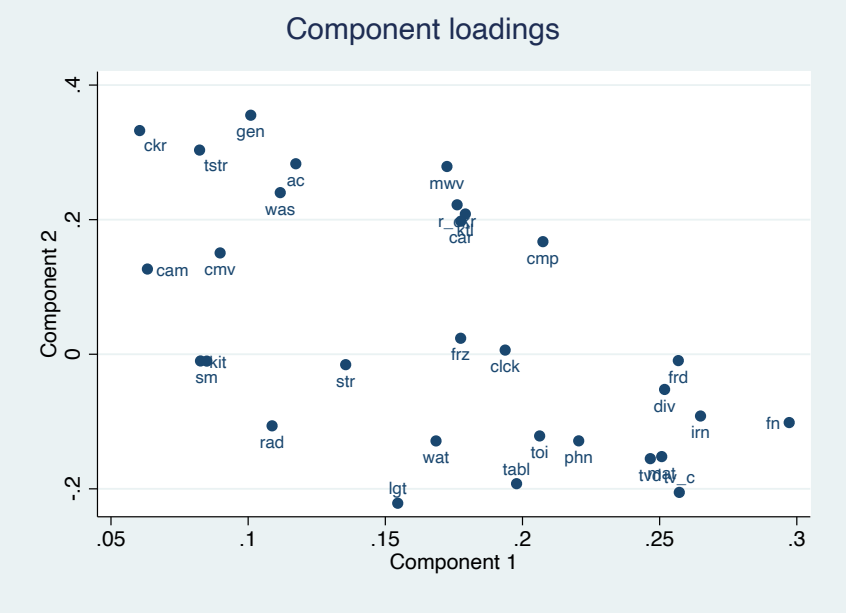
The scree plot indicates that the first component is sufficient to compute the socio-economic status index for our sample.

Figure 3: Scree plot of revised PCA



The loading plot suggests that the selected variables have a stronger influence on the first component, compared to the initial selection.

Figure 4: Loading plot of revised PCA



3 SES by Rural-Urban, Region and Stove Type

3.1 SES by Rural-urban households and by region

Table 10: Socio-economic status by rural-urban households

	Frequency	Share
rural		
least poor	322	4.41
less poor	505	6.91
poor	729	9.98
more poor	988	13.52
very poor	1275	17.45
Total	3819	52.27
urban		
least poor	1139	15.59
less poor	956	13.09
poor	732	10.02
more poor	473	6.47
very poor	187	2.56
Total	3487	47.73
Total		
least poor	1461	20.00
less poor	1461	20.00
poor	1461	20.00
more poor	1461	20.00
very poor	1462	20.01
Total	7306	100.00

As expected, Greater Accra has the highest percentage of 'least poor' households and the lowest percentage of 'very poor' households.

Table 11: Socio-economic status by region (percent)

Region	least poor	less poor	poor	more poor	very poor	Total
Ahafo	4.24	5.13	4.79	6.84	3.49	4.90
Ashanti	14.92	14.03	9.86	7.26	3.90	9.99
Bono	6.43	6.23	5.34	4.72	2.87	5.12
Bono East	3.56	3.29	5.13	5.54	6.98	4.90
Central	5.89	7.94	11.36	11.02	6.16	8.47
Eastern	9.51	10.20	10.95	10.47	4.17	9.06
Greater Accra	25.19	14.58	5.82	2.81	0.55	9.79
North East	0.68	1.23	4.45	5.68	12.11	4.83
Northern	2.33	3.97	5.07	4.18	14.02	5.91
Oti	3.01	4.24	6.64	6.43	4.31	4.93
Savannah	1.85	2.26	4.52	6.78	9.17	4.91
Upper East	1.98	3.29	3.29	5.89	12.79	5.45
Upper West	2.53	3.15	3.29	5.13	9.78	4.78
Volta	4.59	5.27	6.57	6.78	4.86	5.61
Western	8.28	10.06	8.01	5.34	1.16	6.57
Western North	5.00	5.13	4.93	5.13	3.69	4.78
Total	100.00	100.00	100.00	100.00	100.00	100.00
	(1461)	(1461)	(1461)	(1461)	(1462)	(7306)

3.2 SES by stove type

As expected the least poor household largely use LPG while the most poor household typically use the 3-stone cook stoves or open fires as primary cookstoves.

Figure 5: Primary stove by SES

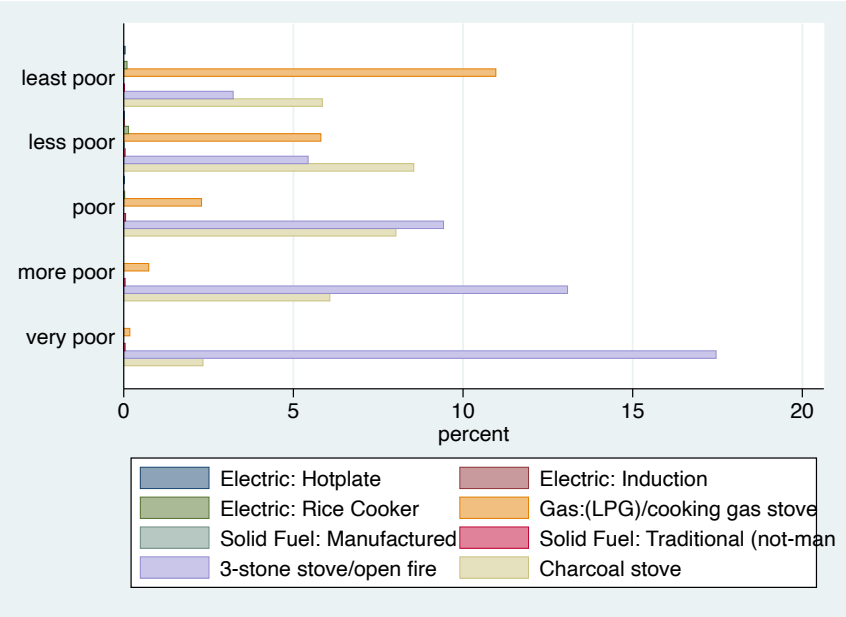


Table 12: Socio-economic status by primary stove

Primary Stove	least poor	less poor	poor	more poor	very poor	Total
Electric: Hotplate	0.21	0.07	0.07	0.00	0.00	0.07
Electric: Induction	0.00	0.07	0.00	0.00	0.00	0.01
Electric: Rice Cooker	0.48	0.69	0.14	0.00	0.00	0.26
Gas:(LPG)/cooking gas stove	54.30	29.03	11.56	3.70	0.90	19.99
Solid Fuel: Manufactured	0.00	0.07	0.00	0.00	0.00	0.01
Solid Fuel: Traditional (not-manufactured)	0.07	0.21	0.28	0.21	0.21	0.19
3-stone stove/open fire	15.97	27.15	47.51	65.62	87.22	48.62
Charcoal stove	28.97	42.71	40.43	30.47	11.67	30.83
Total	100.00	100.00	100.00	100.00	100.00	100.00
	(1453)	(1440)	(1427)	(1434)	(1440)	(7194)

In figure 6 all socio-economic groups use charcoal as secondary stove. However the least poor tend to have charcoal stoves as secondary stoves compared to other SES groups.

Figure 6: Secondary stove by SES

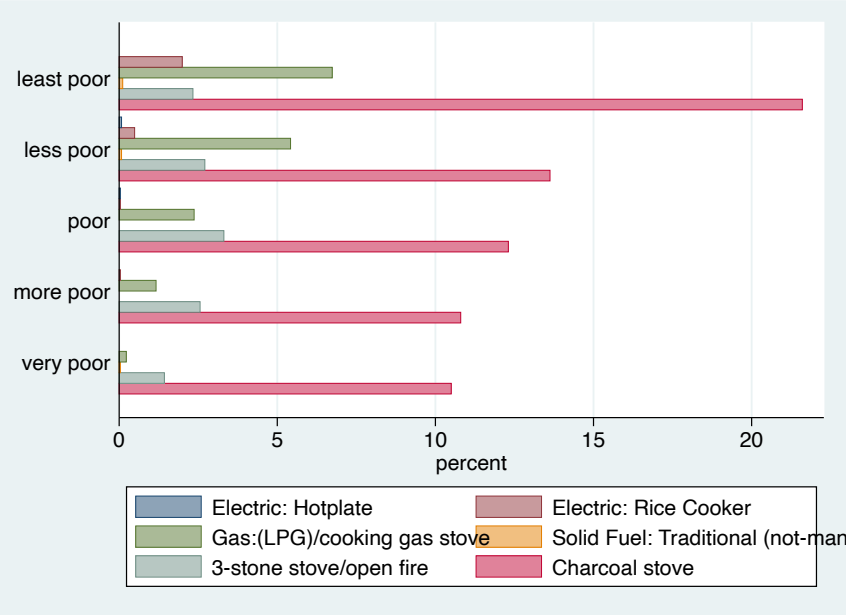


Table 13: Socio-economic status by secondary stove

Sec Stove	least poor	less poor	poor	more poor	very poor	Total
Electric: Hotplate	0.00	0.34	0.21	0.00	0.00	0.11
Electric: Rice Cooker	6.08	2.18	0.21	0.26	0.00	2.56
Gas:(LPG)/cooking gas stove	20.55	24.20	13.12	8.01	1.85	15.92
Solid Fuel: Traditional (not-manufactured)	0.34	0.34	0.00	0.00	0.31	0.23
3-stone stove/open fire	7.12	12.10	18.33	17.57	11.73	12.34
Charcoal stove	65.90	60.84	68.12	74.16	86.11	68.84
Total	100.00	100.00	100.00	100.00	100.00	100.00
	(871)	(595)	(480)	(387)	(324)	(2657)

4 Nomalizing the wealth scores

The range of wealth index score (p1) was -3.8 to +43.9 with median of -0.5. As negative wealth index scores are intuitively difficult to comprehend, we planned a modified score list such that the lowest score is +1. So, we added +4.8 (+3.8 +1) to p1 values for transformation. Thus, we created a modified wealth index score (mod_sc) with range of +1 to +48.7 and median value of +4.3