AN ANALYSIS OF SURVEY DATA FOR ESTIMATING MARCELLUS SHALE PERCEPTIONS

by

Katrina Zell

B.A. in International Economics, Marshall University, 2007

M.A. in Economics, Miami University, 2008

Submitted to the Graduate Faculty of

the Department of Biostatistics of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Katrina Zell

It was defended on

April 3, 2013

and approved by

Thesis Advisor:
Ada Youk, PhD, Assistant Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Bernard D. Goldstein, MD, Dean and Professor Emeritus
Department of Environmental and Occupational Health
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Jeanine Buchanich, PhD, Assistant Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

AN ANALYSIS OF SURVEY DATA FOR ESTIMATING MARCELLUS SHALE PERCEPTIONS

Katrina Zell, M.S.

University of Pittsburgh, 2013

ABSTRACT

This investigation evaluates different methodologies to determine which is most appropriate to use on survey data. Multivariate ordinary linear regression (OLS), logistic, multinomial logistic (MLR), and ordered logistic regressions (OLR) were conducted to predict support for or opposition to drilling the Marcellus Shale (MS) for natural gas. In all analyses, perceptions of the MS as an economic opportunity and as an environmental and public health threat significantly affected support of drilling, increasing and decreasing, respectively. Women were less supportive of drilling, and having a family-owned natural gas lease increased support of drilling.

The assumptions of OLS were violated, indicating it was a poor choice for these data. The assumptions of logistic regression were met, but literature indicates that dichotomizing an outcome affects the inferences that can be made about the results. The assumptions of the MLR and OLR were violated with the original data, but when collapsing the outcome levels, the assumptions of the MLR were met. The consistently statistically significant predictors had similar estimated odds ratios in both the MLR and the OLR. A larger sample size is necessary in order to get more conclusive results, but it appears that an OLR is the most appropriate methodology to use on a categorical outcome in survey data. The findings of this study can be used to determine the public's attitudes regarding MS activities. These attitudes can be seen as reflective of how the MS is perceived to affect the public's health and which changes need to be made as a way to improve health.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Thank you to Pittsburgh Today and the University of Pittsburgh's University Center for Social and Urban Research for incorporating questions about MS drilling and additional respondents into the survey design, and the university's Graduate School of Public Health for the funding to support this. Thank you to Charles Christen for assistance in survey question development and Renée Zell for technical assistance. Thank you to Jill Kriesky for assistance in preliminary research of Marcellus Shale issues.

Abbreviated article title: Methods for estimating perceptions with survey data

Keywords: Marcellus Shale, survey data, methodological comparisons

1.0 INTRODUCTION

The extraction of natural gas is not a new occurrence, nor is its extraction through unconventional natural gas drilling (UGD), also known as hydraulic fracturing. However, recently there has been much contention surrounding UGD and the methods through which it is employed. Many, such as politicians, members of the natural gas industry, and mineral rights owners, would like to see the UGD process be allowed to grow. The proponents of UGD make claims that drilling will bring in revenue on a variety of fronts, that it is a clean source of energy, and that it can reduce or possibly eliminate our dependence on foreign oil (Chevron, 2011; Cauchon, 2012; Browder, 2012). Those who are against UGD claim that it brings a host of negative health effects, will increase water and air contamination, and that in areas where drilling is occurring, crime rates increase (O'Day and Reece, 2012; Downing, 2012).

An important factor that drives policy-making decisions is the opinions of those who live in the districts of the policy-makers. Just like many other environmental issues, how to proceed with UGD is a contentious issue that has divided many communities. There are arguments that the pace of drilling should slow, increase, and remain the same (Brasier et al., 2011; Jacquet, 2009; Jacquet, 2012). The contributing factors that persuade people to choose one side of an argument or another can be complex. The first part of the focus of this investigation is to determine if there is a difference between two adjacent Pennsylvania counties regarding the opinion on drilling the Marcellus Shale (MS) for natural gas. Secondly, if this difference does exist, an attempt will be made to identify factors that might explain what is contributing to this difference of opinion. In addition, in order to ensure that best methodology is being used, four different types of regression analysis will be employed on the sample. The results of the analyses will be compared and an assessment will be given to determine which methodology is best suited for the survey sample. The results will also be useful in determining how the public perceives their health is being affected by MS activities. Previous studies have found that perception of economic impact, perception of environmental/public health threat, having a natural gas lease, gender, and from where information regarding the drilling activity is coming play key roles in determining support for or opposition to drilling for natural gas (Kriesky et al., 2013; Alter et al., 2010; Jacquet, 2012; Jacquet and Stedman, 2011).

1.1 MARCELLUS SHALE

Marcellus Shale is a sedimentary rock formation that is believed to be over 350 million years old. It is located underneath parts of New York (NY), Pennsylvania (PA), Maryland (MD), Ohio (OH), West Virginia (WV), Virginia (VA), and a very small section of Tennessee (TN). A majority of the shale formation lies underneath PA, WV, NY, and OH. The Energy Information Administration (EIA, 2012) has revised its estimates that the MS may contain up to 49% of the nation's natural gas supply in 2035. The average shale gas well in the MS is believed to be capable of producing 3.1 billion cubic feet (bcf) of gas over a period of approximately 60 years (Andrews, et al. 2009). States such as PA and WV have been proactive in facilitating the development of drilling the MS for natural gas. Other states such as NY and MD have been much more cautious in how drilling the MS is taking place. For example, in 2008, NY placed a four-year moratorium on drilling (Resources for the Future, 2012). The moratorium is designed to allow the environmental impact of UGD to be assessed, and determine what, if any, negative consequences may occur with drilling (McAllister, 2013).

Shale gas is extracted using hydraulic fracturing. The hydraulic fracturing process involves multiple steps (Andrews et al., 2009). First the well is tested to make sure that it can withstand the pressures of the process. A hydrochloric acid solution is then used to clean up the residue from the well casing, then water with a proprietary mix of chemicals is pumped into the well in stages, in an attempt to crack open the shale and release the oil and gas inside. While the Occupational Safety and Hazard Administration (OSHA) requires that all companies keep a list of all chemicals used at each drill site, they are not required to disclose the proportions of each chemicals, as that information can be considered proprietary by the company (OSHA, 2006). This water is subsequently pumped back out, so that it does not block access to the oil and gas in the shale formation.

There are risks of water and air contamination associated with the UGD process. If the well is properly built and sealed, then it prevents the fracturing fluids, gas and drilling fluids from leaking out into the ground water and aquifers and vice versa. There are three principal hydrogeological environments that exist in the MS, all of which may be contaminated by run-off from leaky surface impoundments or poor waste-water management, due to the fact that they are refilled by aquifers that are susceptible to these kinds of contamination (Andrews et al., 2009). As such, disposal of this fracturing water continues to be an issue. In 2008, PA placed a ban on the disposal of hydraulic fracturing fluids into wastewater treatment plants, as they were believed to be causing an increase in total dissolved solids (TDS) in the Monongahela River (PA Environmental Digest, 2009). It was later stated that more of the TDS in the river were due to run-off from abandoned mines. Officials did recognize the need for a way to treat and/or dispose of water from hydraulic fracturing operations, and prohibited the disposal of high-TDS waste-water in PA water sites beginning the first day of 2011 (Andrews et al., 2009).

This study takes advantage of a survey completed by the University Center for Social and Urban Research (UCSUR) to compare attitudes regarding drilling the MS for natural gas of two adjacent counties in Southwest PA with markedly different levels of drilling occurring. UCSUR surveyed residents of a 32 county region centered around Allegheny County (AC), Pennsylvania in a general quality of life survey. Additional funding was provided so that UCSUR could include additional environmental questions (see Table 1) as well as oversample Washington

County (WC), PA, so that it could be compared to neighboring Allegheny County (AC), PA. Washington County is a more rural and highly active drilling area, with 242.5 residents per square mile (US Census Bureau, 2011), and as of December 2012, has 732 unconventional wells dug (PADEP, 2012). In contrast, AC is more urban (1675.6 residents per square mile; US Census Bureau, 2011) and has far less active drilling, with only 22 unconventional wells dug, as of December 2012 (PADEP, 2012). In spite of these differences, AC (730 sq. miles) and WC (857 sq. miles) are relatively similar in size (US Census Bureau, 2011). Because of these differences, we believe that the respondents in WC and AC will have meaningful differing opinions regarding drilling the MS for natural gas. Therefore, in the subsequent regression analyses, county of residence will be the variable of interest.

Table 1: Survey questions considered as predictors

Considering everything, how do you feel about natural gas extraction from the MS?						
	Strongly oppose	Somewhat oppose	Neither oppose nor support	Somewhat support	Strongly support	p-value§
WC	8.6%	16.8%	23.9%	26.6%	24.2%	0.0768
AC	12.5%	16.6%	30.1%	23.0%	17.9%	
First, how closely would you say you have been following the Marcellus Shale issue?						
	Very closely	Somewhat closely	A little bit	Not at all		
WC	30.7%	31.7%	22.4%	15.2%		0.0003
AC	16.6%	34.9%	28.6%	19.9%		
To what extent do you think the MS represents an economic opportunity for this region?						
	Significant opportunity	Moderate opportunity	Slight opportunity	Very little/no opportunity		
WC	47.4%	35.0%	12.0%	5.6%		0.0015
AC	34.5%	36.6%	18.4%	10.5%		
To what extent do you think the MS represents a threat to the environment and public health of the region?						
	Significant threat	Moderate threat	Slight threat	Very little/no threat		
WC	22.4%	35.3%	23.2%	19.2%		0.1090
AC	28.0%	32.4%	26.2%	13.4%		
Have you or anyone in your family signed a lease with a natural gas company for rights to extract natural gas from land that you or someone in your family owns?						
	Yes	No				
WC	29.9%	70.1%				<0.0001
AC	4.3%	95.7%				

Table 1 Continued

Do you feel that state government oversight of the environment should...

	Increase significantly	Increase somewhat	Remain the same	Decrease somewhat	Decrease significantly		
WC	33.5%	32.9%	18.8%	9.8%	4.9%		0.8066
AC	28.9%	34.3%	19.8%	10.3%	6.7%		

Would you say that overall environmental quality in our region is…

	Improving significantly	Improving somewhat	Remaining the same	Getting somewhat worse	Getting significantly worse		
WC	5.1%	23.7%	45.6%	20.6%	5.0%		0.0941
AC	8.9%	29.8%	39.5%	16.9%	4.9%		

How long have you lived at your current residence?

	<1 year	1-3 years	3-5 years	5-10 years	10-20 years	20+ years	
WC	4.7%	13.2%	9.9%	18.0%	16.4%	37.9%	0.0712
AC	9.0%	15.1%	6.8%	19.9%	18.7%	30.5%	

§ The p-value is for the test to see if there is a statistically significant difference between the two counties in the responses to the questions in the table.

1.2 ORDERED OUTCOMES

To determine if there was a difference in opinion regarding the support for or opposition to drilling the MS for natural gas between AC and WC, the survey question “Considering everything, how do you feel about natural gas extraction from the MS region?” was examined. This questions contains five possible responses: “Strongly support”, “somewhat support”, “neither support nor oppose”, “somewhat oppose”, or “strongly oppose”. Given the nature of the outcome, building a model and assessing the effects of the predictors is less straightforward than it would be with a continuous outcome. There are four different methodologies that can be applied to assessing the outcome, ordinary least squares (OLS), (binary) logistic regression (logit), multinomial logistic regression (MLR), and ordered logistic regression (OLR).

1.2.1 Ordinary Least Squares

Before OLR became a popular way to analyze discrete, ordered outcomes, it was common practice to use ordinary linear regression, also called ordinary least squares (OLS). If the dependent variable was ordinal, it was common to assign scores to the levels of the outcome, treat the scaled variable as continuous, and then use OLS on the scaled outcome.

One possible way to assess the support for or opposition to drilling the MS for natural gas is to assume that the outcome has an underlying continuous distribution and that the real line is broken up into a series of intervals that relate to the categorical outcome (Anderson, 1984). If a continuous distribution is assumed, then part of that assumption would be that designation of response to a number (i.e.: 1 = strongly oppose) is not arbitrary and in this dataset, that would indicate that respondents naturally start out being strongly opposed to drilling. Under this assumption, ordinary least squares (OLS) can be utilized, making the interpretation of the results much simpler.

This assumes a standard linear regression model:

$$Y_i = \alpha + ßx_i + \varepsilon_i$$

Where Y_i is the outcome of interest, α is the intercept, x is the vector of covariates, ß is the vector of coefficients, and ε is the error term. Each ß is interpreted as the amount by which Y_i will change, given a one unit increase in ß's respective x_i, all other independent variables held constant. Additionally, estimation of $\widehat{ß}$ by OLS gives a value that minimizes the sum of squared errors. There are four main assumptions when using OLS regression: 1) normality of the errors, 2) constant variance of the error terms (homoscedasticity), 3) independence of the error terms (no serial correlation), and 4) linearity relationship between the outcome and the covariates. Provided these assumptions hold, the OLS estimator of $\widehat{ß}$ yields the best linear unbiased estimator (BLUE), which means that out of all possible estimates, this estimator has the least variance.

1.2.2 Logistic regression

Another method to assess support or opposition of MS drilling is to dichotomize the outcome. In doing this, all respondents who claimed to "neither support nor oppose" drilling the MS for natural gas would be eliminated from the sample. One could argue that these respondents are uninformative due to their indifference and that little information will be lost due to their elimination. This method allows for a more direct comparison of support and opposition, does not require the distributional assumptions that OLS does and has the most straightforward interpretation of all three types of logistic regression. The assumptions of the logistic regression model are: 1) the outcome variable is discrete, 2) the relationship between the independent variables and the log odds of the outcome is linear and, 3) the observations are independent.

The outcome (Y_i) is dichotomous as 0 or 1 and follows a Bernoulli distribution. For these data, if the outcome takes on a value of zero, then the respondent supports drilling. Conversely, if the outcome takes on a value of one, then the respondent opposes drilling. The probability density function of Y_i can be written as:

$$\Pr\{Y_i = y_i\} = \pi_i^{y_i}(1-\pi_i)^{1-y_i}$$

where π is the probability of the outcome equaling one (in this case it is the probability of opposing drilling) and the x_i's are the independent variables:

$$\pi_i = \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}$$

To abide by the linearity assumption, the probability is transformed into log odds:

$$\text{logit}\,\pi_i = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \beta x_i$$

This transformation allows variables with values ranging from negative infinity to positive infinity to be incorporated into a model and end with a result that stays within the (0,1) range. The ß's are interpreted as a one unit increase in x_i results in an increase in the log-odds of Y_i of ß.

1.2.3 Multinomial logistic regression

A third way to analyze the data is to argue that, in spite of its ordinal nature, each respondent's perception on the meaning of each outcome on the Likert scale is subjective and thus different from that of the next respondent's perception. Because the perceptions are so widely varied, one can view the choices to the survey question as categorical, without truly being ordered, and MLR can be applied. This methodology will allow each level of the response to be compared to a baseline category (essentially running multiple binary logits), and may show that some predictors are more or less influential on one level of the outcome versus the baseline, when compared to the other levels versus the baseline.

MLR can be viewed as a way of simplifying the use of binary logistic regression on an outcome with more than two categories. Instead of running *J-1* logistic models for an outcome with *J* levels, and comparing each logit to a designated baseline category, MLR can be used. Each predictor is allowed to have a different coefficient for each different outcome level under MLR. The use of MLR allows the results to be given simultaneously, still using the same pre-specified baseline outcome level. This allows comparisons of the likelihood of the respondents falling into the *J-1* separate categories, prevents analyses from being made using *J-1* different sample sizes, and makes comparisons of the outcome categories to the baseline simpler.

MLR is used when outcome Y_i has three or more categories (i=1,…, J), and each Y_i also follows a Bernoulli distribution. The probability of the i^{th} level of the outcome occurring, π_i, looks similar to that of a logistic regression, except now the denominator changes to account for the fact that the *J-1* levels are being compared a specific baseline level:

$$\pi_i = \frac{\exp(\alpha + \beta x_i)}{1 + \sum_{j=1}^{J} \exp(\alpha + \beta x_j)} \; ; i \neq j$$

To satisfy the linearity assumption of the log odds, the log is taken:

$$\text{logit}\ (\pi_j) = \log\left(\frac{\pi_j}{\pi_k}\right) = \alpha_i + x_i(\beta_j - \beta_k)\ ;\ k \neq j$$

where j is the outcome level against which all comparisons are being made. In order for the multinomial logistic model to be valid, it is assumed that the $\log\left(\frac{\pi_i}{\pi_j}\right)$ is a linear combination of the x_i's. Because the model is linear in the logit, $(\beta_j - \beta_k)$ can be interpreted as a unit increase in x_i that produces an increase in the log-odds of the outcome by $(\beta_j - \beta_k)$, all other predictors held constant. Using MLR, constraints are imposed on the estimates of the equations, such that a $\hat{\beta}$ can be solved for by taking the difference of all other $\hat{\beta}$s of the other (non-redundant) binary logits.

Another crucial assumption of this model is the assumption of independence of irrelevant alternative, which states that the odds of selecting k over j will not be affected if another level of the outcome is changed or eliminated. This can be a difficult assumption to apply to some data, because in reality, the lack of one option very well may change the odds of selecting another option (McFadden, 1974).

1.2.4 Ordered logistic regression

Agresti (1999) states that a cumulative logit model is the most popular way to analyze ordered outcomes. OLR uses the cumulative logit, introduced by McCullagh (1980), which in turn uses cumulative probability:

$$\pi_{ij} = Pr[y_i \leq j] = \sum_{k=1}^{j} Pr[y_i = k]$$

Again the log-odds of the cumulative probability is taken, to satisfy the linearity assumption:

$$logit(\pi_{ij}) = \log[\pi_{ij}/(1 - \pi_{ij})]$$

The OLR, is obtained by modeling the cumulative logit as a linear function of independent variables:

$$logit(\pi_{ij}) = \alpha_j - ßx_i$$

This results in a model where the ß's do not vary for a given level of the outcome. The only difference in the result for a given outcome level is that each level of the outcome has a different intercept, or cut-point. Under this model, the parallel lines assumption is employed, which states that there is a single common slope for each predictor, regardless of the level of the outcome that may be selected. This assumption is often violated and may not necessarily hold for the data (Williams, 2006). Long (2012) also finds that the test of parallel odds is often rejected, but states that the results of OLR and MLR should be compared to better determine if the OLR is an inappropriate model or not. However, the use of this model does allow for increased power and makes the model more parsimonious. Because there are identical slopes for each level of the outcome, outcome levels can be combined. However this can result in a loss of precision in the estimates.

2.0 METHODS

2.1 SURVEY METHODS

When data are collected through a survey, certain steps are necessary in order to ensure that any results from the data are not biased or skewed as a result of the sampling process. Random digit dialing (RDD) is a standard method used to try to randomize the respondents who are sampled. Because it can be very difficult to get a completely random sample, survey weights are applied to the final dataset in order to make the sample more representative of the general population from which the survey sampled. This survey is broken up by phone line and county of residence strata. These strata are sampled independent of one another, so each stratum has its own weighting variable. Failing to use the appropriate analytical techniques for survey data will likely cause underestimation of the standard errors of the predictors. This can potentially lead to variables that appear to be statistically significant, but may, in fact, not be.

2.2 UCSUR SURVEY

The data come from a survey completed by the University of Pittsburgh's University Center for Social and Urban Research (UCSUR) in the fall of 2011. UCSUR had developed a general quality of life (QOL) survey to be administered to residents of the Pittsburgh area and surrounding counties (32 counties in total). Broad ranges of topics were covered, from basic demographic information to questions regarding participation in the arts and health information. The University of Pittsburgh's Graduate School of Public Health secured additional funding so that more environmental questions could be added to the survey, specifically questions regarding perceptions about the MS, as the issue had risen to prominence at that time. Funding was also secured so that WC would be oversampled, in part because UGD is more active there.

Respondents were selected through a random digit dialing process completed by UCSUR's computer assisted telephone interviewing (CATI) lab. Both those with landlines and with cellphones were eligible for selection. There was an overall 11% response rate to the survey, which is consistent with landline and cellphone surveys.

2.2.1 Survey questions:

The questions specifically regarding perceptions and attitudes towards the MS are show below in Table 1. The first question in the table is the outcome, and was taken directly from a mail survey done by Alter et al. (2010). The other questions regarding the MS were developed by a research team for the University of Pittsburgh, so that environmental and public health perspectives could be observed. These and other questions were considered in the analyses. Non-demographic questions that were considered to be potential predictors and made it in to the final model of at least one of the regression types are listed in Table 1.

2.3 STATISTICAL METHODS

To determine if there were differences in attitudes between the two counties, cross-tabulations with chi-squared statistics were calculated to compare the responses of the survey. Next, alternative regression techniques were employed to determine which predictors independently explained support versus opposition to drilling. As county of residence was our predictor of interest, it remained in the models so that we could assess how its influence changed with the inclusion of additional predictors. OLS, logistic, MLR, and OLR analyses will be used. An *a priori* model was selected first, including demographics, a county of residence indicator, and a question related to the respondent's perception of MS activities. Demographics (age, race, gender, and educational attainment) were included in the models, regardless of statistical significance, to keep confounding to a minimum. For each regression method, univariate analyses was performed on all other potential predictors with an α of 0.25 used to determine if the potential predictor would kept for the backwards elimination step of the model building. After reducing the set of potential predictors, for each regression method, a multivariate model was built using a backwards elimination technique using a liberal α=0.10 to keep confounding to a minimum (Vittinghoff et al., 2012). Results were considered statistically significant with a p-value $\leq$0.05 and of borderline statistical significance with a p-value $\leq$0.10.

Checks were performed on all categorical predictors to determine if they should be used as continuous predictors or if they need to be turned into a series of indicators. Variance inflation factors (VIFs) were calculated for an assessment of possible problematic multicollinearity (correlation among the predictors). Standard diagnostics were run on the final model of each regression type. Assessments of overall goodness of fit were performed for all of models. Linearity, homoscedasticity, and normality were checked in the OLS model, as well as checks for outliers. Model specification, influential points, and outliers were checked for in the logistic model. The assumption of independence of irrelevant alternatives was checked in the multinomial logistic model, and the proportional odds assumption was checked in the ordered logistic model. Upon examination of model results and diagnostics, a determination was made as to which type of methodology is most appropriate for these data.

3.0 RESULTS

Cross-tabulations were first examined to determine if there were any significant associations between potential predictors of the outcome and county of residence. As shown in Table 1, there were borderline statistically significant differences in support for and opposition to drilling between the two counties (p=0.0768).

Many more respondents in WC that in AC stated they followed the MS issue very closely (p=0.0003). Additionally, more in WC believed that the MS represents a significant economic opportunity (p=0.0015), although more in WC would also like to see a significant increase in government oversight of the environment (p=0.8066). Conversely, more respondents in AC believed that the MS is a significant environmental or public health threat (p=0.1090). The largest difference between the two areas is that about 25% more respondents in WC have or are related to someone who has signed a natural gas lease (p<0.0001).

3.1.1 Sample demographics by county

In spite of these two counties being geographically similar in size and location, there were some noticeable differences when comparing the demographic variables (Table 2). There was no statistically significant difference overall between the two counties for education level, although almost 8% more respondents in WC have a high school education or less. There were also more women in WC than AC, although this result was borderline statistically significant (p=0.0654). In AC there was a statistically significantly larger African American population compared to WC (p=0.0001).

Table 2: Demographics

	WC	AC	p-value
Age*	50.32	48.63	0.1954
Female	62. 3%	55.4%	0.0654
Black	5.2%	16.1%	0.0001
Education			0.1667
HS or less	37.6%	30.0%	
Some college	25.2%	27.0%	
Bachelors	21.5%	26.5%	
Masters +	15.7%	16.6%	

*mean

3.2 BOTH COUNTIES

3.2.1 Ordinary least squares

The results of ordinary linear regression (OLS) for the null model showed county of residence as an important predictor of support or opposition to drilling the MS (p=0.054). The coefficient for county was positive (ß=0.255), indicating that WC residents are more supportive of drilling the MS for natural gas than are AC residents.

Results of the backwards elimination linear regression modeling are shown in Table 3. County of residence was no longer statistically significant and the coefficient was negative (ß=-0.062), indicating that WC residents were less supportive of drilling than AC residents, controlling for all other variables. None of the demographic predictors in the model were statistically significant with the exception of gender (p=0.001). Marital status (p=0.072) and political party (p=0.094) were borderline statistically significant. As noted above in the statistical methodology, we kept the demographic variables in the model regardless of statistical significance, to control for potential confounding.

Perception of the MS as an environmental or public health threat decreased support for drilling the MS for natural gas as the perceived threat increased (significant threat: ß=-1.579, moderate threat: ß=-0.754, slight threat: ß=-0.178; p<0.0001). Conversely, the more the MS is viewed as an economic opportunity, the greater support the respondent had for drilling it (significant opportunity: ß=1.120, moderate opportunity: ß=0.586, slight opportunity: ß=0.278; p<0.0001). If the respondent, or someone in the respondent's family, had signed a natural gas lease, then support for drilling also increased (ß=0.230; p=0.040). Women, however, were less in favor of drilling than are men (ß=-0.261; p=0.001). Like those who see the MS as an environmental/public health threat, the more state government oversight of the environment a respondent wants, the less supportive (s)he was of drilling (increase significantly: ß=-0.387, increase somewhat: ß=-0.009, remain the same: ß=0.008, decrease somewhat: ß=0.043; p=0.010). With an R^2 of 0.5437, we can see that the predictors in this model explain about 54% of the variation in the outcome.

Table 3: Results of OLS analysis on full sample[§]

Predictor	Coefficient (95% Confidence Interval)	p-value
WC/AC	-0.062 (-0.223, 0.099)	0.450
Follow MS issue	-0.113 (-0.207, -0.020)	0.017
Envir/health threat (baseline is no threat)		<0.0001
Significant threat	-1.579 (-1.913, -1.245)	
Moderate threat	-0.754 (-0.990, -0.518)	
Slight threat	-0.178 (-0.376, 0.020)	
Family lease	0.230 (0.011, 0.450)	0.040
Female	-0.261 (-0.418, -0.104)	0.001
Age	0.003 (-0.003, 0.008)	0.338
Education[*]	0.024	0.544

Table 3 Continued

	(-0.054, 0.103)	
Black	0.113	0.488
	(-0.207, 0.432)	
Econ. opportunity (baseline is no opportunity)		<0.0001
Significant opportunity	1.120	
	(0.668, 1.573)	
Moderate opportunity	0.586	
	(0.167, 1.004)	
Slight opportunity	0.278	
	(-0.157, 0.713)	
Gov't oversight of envir.		0.010
Increase significantly	-0.387	
	(-0.985, 0.211)	
Increase somewhat	-0.009	
	(-0.556, 0.539)	
Remain the same	0.008	
	(-0.512, 0.528)	
Decrease somewhat	0.043	
	(-0.506, 0.592)	
Overall envir. quality	-0.064	0.166
	(-0.154, 0.027)	
Political party		0.094
Democrat	0.092	
	(-0.104, 0.287)	
Republican	0.226	
	(0.022, 0.429)	
Marital status		0.072
Married	-0.193	
	(-0.393, 0.008)	
Widowed	-0.044	
	(-0.391, 0.304)	
Divorced	-0.368	
	(-0.723, -0.013)	
Intercept	3.834	<0.0001
	(2.967, 4.701)	

* 1=HS or less, 2=some college, 3 Bachelors, 4=Graduate school
† 1=Improving significantly, 2=Improving somewhat, 3=Remaining the same, 4=Getting somewhat worse, 5=Getting significantly worse
§ R^2=0.5437

To assess the fit of this model, multiple diagnostics were run. All VIFs were less than 10, indicating that multicollinearity was not an issue with these data. A Shapiro-Wilk test assessing the normality of the data had p-value of <0.0001, indicating that the data were non-normal, however the kernel density plot in Figure 1 showed that the residuals looked fairly normal. A histogram of the outcome is shown in Figure 2. Because of the categorical nature of the outcome, it was difficult to discern if the histogram showed that the data were non-normal.

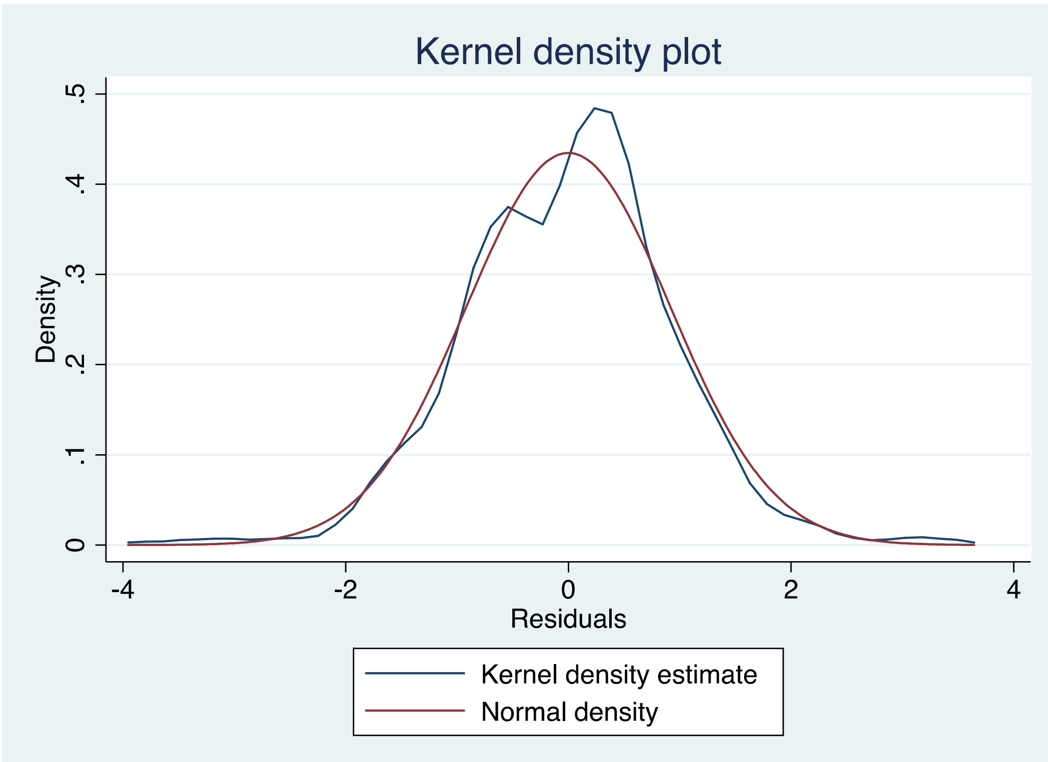

Figure 1: Kernel density plot

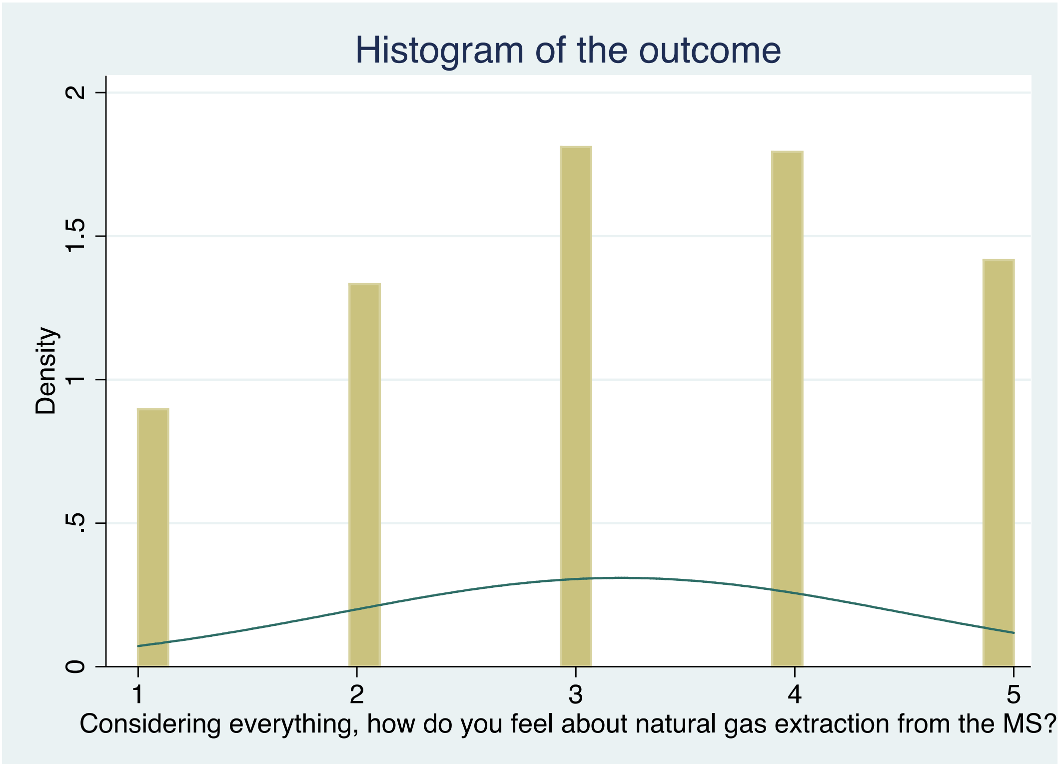

Figure 2: Histogram of outcome

The Breusch-Pagan test for homoscedasticity on non-normal data was statistically significant with p=0.0144, indicating that the errors are heteroscedastic. Figure 3 shows a plot the residuals versus the fitted values, where there is a clear pattern, further asserting that there was a problem with the errors.

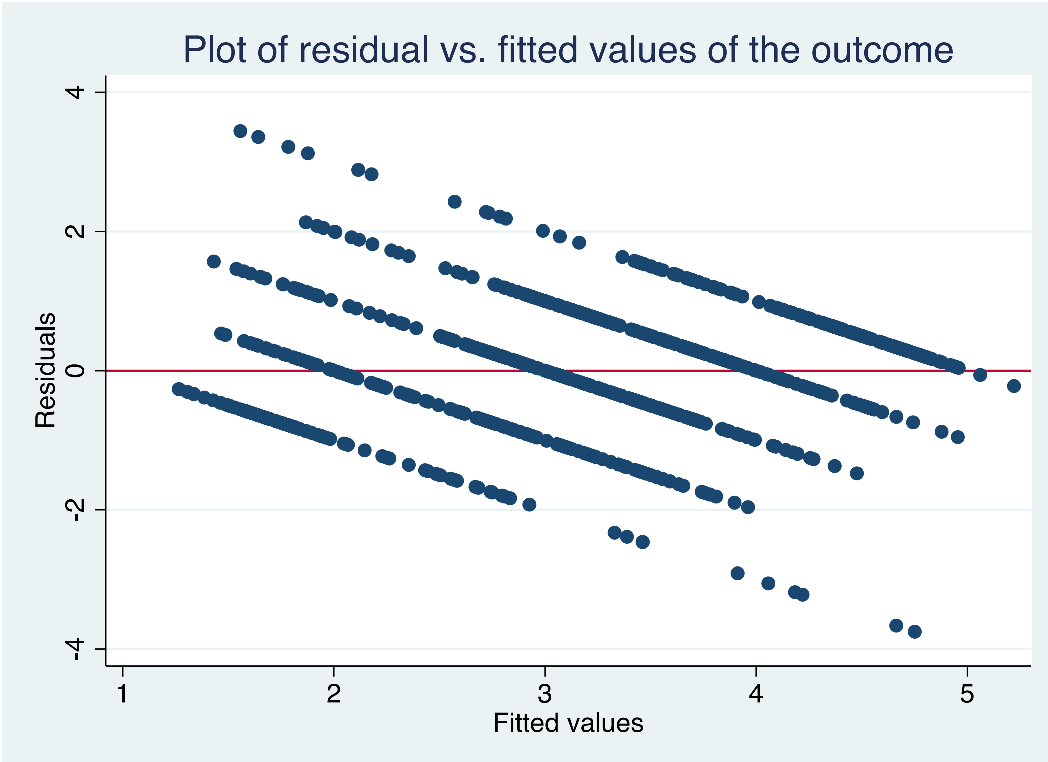

Figure 3: Residuals vs. fitted values

3.2.2 Logistic regression

For the logistic regression models, the levels of the outcome were collapsed to create a dichotomous dependent variable. To create this dichotomous variable, those who said they "strongly oppose" and "somewhat oppose" drilling the MS for natural gas were given a value of one (n=377), those who answered "strongly support" and "somewhat support" were given a value of zero (n=506), and those who chose "neither support nor oppose" were omitted from the same (n=341). While those omitted represent approximately 28% of the sample, they can be viewed as respondents who are not useful in predicting support or opposition, as they are essentially claiming to be un-opinionated about the topic. The results of the null model for the logistic regression showed that county of residence was borderline statistically significant (p=0.101). With an odds ratio of 0.684, being a WC resident decreased the odds of opposing drilling by about 32%. County of residence is the factor of interest in predicting support for or opposition to drilling the MS for natural gas, and was included in the predictive model so it can be observed how it changes when controlling for additional factors.

The statistical significance and directionality of the predictors in the logistic model were very similar to that of the OLS model. In both models, once the null model is adjusted for the additional predictors, the effect of county of residence does not retain borderline statistical significance (p=0.586), and with an odds ratio of 1.193, does little to change support for or opposition to drilling (Table 4). Again, as with the OLS results, being a resident of WC actually seemed to slightly increase opposition to drilling the MS for natural gas, when all else is held constant.

Table 4: Results of logistic regression on the full sample

Predictor	Odds Ratio (95% Confidence Interval)	p-value
WC/AC	1.193 (0.632, 2.252)	0.586
Follow MS issue	1.003 (0.665, 1.513)	0.989
Envir/health threat (baseline is no threat)		<0.0001
Significant threat	42.303 (11.779, 151.927)	
Moderate threat	8.665 (2.833, 26.502)	
Slight threat	1.054 (0.277, 4.004)	
Family lease	0.632 (0.243, 1.639)	0.344
Female	2.240 (1.168, 4.297)	0.015
Age	1.001 (0.984, 1.017)	0.938
Education*	0.833 (0.601, 1.155)	0.273
Black	1.426 (0.459, 4.432)	0.539
Econ. opportunity (baseline is no opportunity)		<0.0001
Significant opportunity	0.052	

Table 4 Continued

	(0.010, 0.264)	
Moderate opportunity	0.169	
	(0.034, 0.830)	
Slight opportunity	0.345	
	(0.057, 20.075)	
Gov't oversight of envir.		0.037
Increase significantly	2.484	
	(0.337, 18.291)	
Increase somewhat	1.158	
	(0.170, 7.889)	
Remain the same	0.482	
	(0.064, 3.600)	
Decrease somewhat	0.836	
	(0.098, 7.103)	
Political party		0.055
Democrat	0.478	
	(0.225, 1.018)	
Republican	0.360	
	(0.148, 0.877)	
Intercept	0.736	0.848
	(0.032, 17.146)	

* 1=HS or less, 2=some college, 3 Bachelors, 4=Graduate school

As with the OLS model, the more a respondent perceived the MS to be an environmental or public health threat, the more opposed (s)he was to drilling (significant threat: OR=42.303, moderate threat: OR=8.665, slight threat: OR=1.054; $p<0.0001$). The wide confidence intervals associated with the different threat levels were due to the small cell counts for support or oppose. Also similar was that the more the MS is perceived to be an economic opportunity, the more supportive the respondent was of drilling (significant opportunity: OR=0.052, moderate opportunity: OR=0.169, slight opportunity: OR=0.345; $p<0.0001$) and that being female decreased support for drilling (OR=2.240; $p=0.015$). As with the OLS model, the effects of the other demographic variables were not statistically significant. Unlike in the OLS model, the predictor "The overall environmental quality of the region is . . ." was not included in the logistic model. Opinion on how the state government should handle oversight of the environment was included again, and was statistically significant this time ($p=0.037$). Similar to the perception of environmental/public health threat, the more government oversight of the environment a respondent wanted, the more opposed (s)he is to drilling (increase significantly: OR=2.484, increase somewhat: OR=1.158, remain the same: OR=0.482, decrease somewhat: OR=0.836; p=0.037). Belonging to both the Democratic and the Republican political parties increased support of drilling (Democrat: OR=0.478, Republican: OR=0.360; p=0.055).

An assessment of the overall fit of the model was performed using Hosmer and Lemeshow's goodness of fit test. The null hypothesis of a lack of fit to the model was not rejected, indicating that the overall fit of the model was good. A linktest analysis performed in STATA revealed that meaningful predictors were selected for the model and that there was no specification error in the model. Collinearity diagnostics showed that all VIFs were less than 10, indicating multicollinearity was not an issue.

As can be seen from the figures (4-7), there were issues with these data. Figure 4 showed that there were 40 outliers, including three that were possibly problematic, as they were well outside the absolute value of two. The scatter plot of the deviance residuals (Figure 5) also showed that there were multiple observations that were not well explained by the model fit in Table 3. The scatter plot of the Pregibon leverage in Figure 6 showed that there were 69 high leverage points. However, the plot of the DFBETAs (Figure 7) showed that deleting any of the observations would not affect the estimates of the parameters.

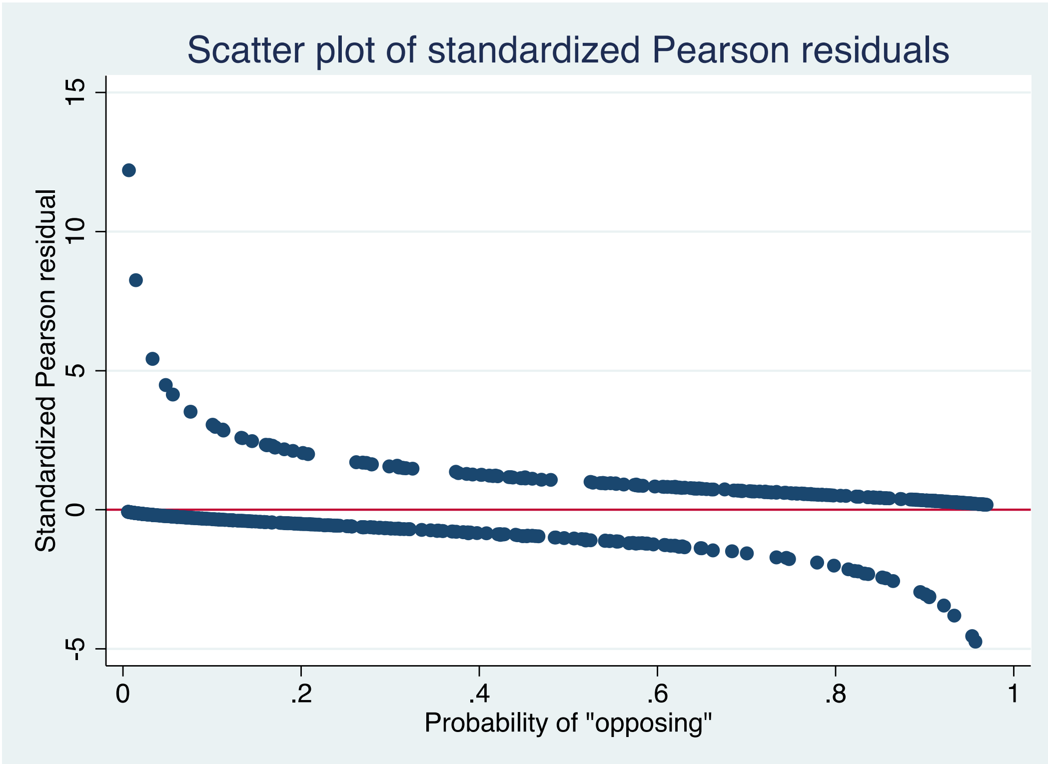

Figure 4: Scatter plot of Pearson residuals

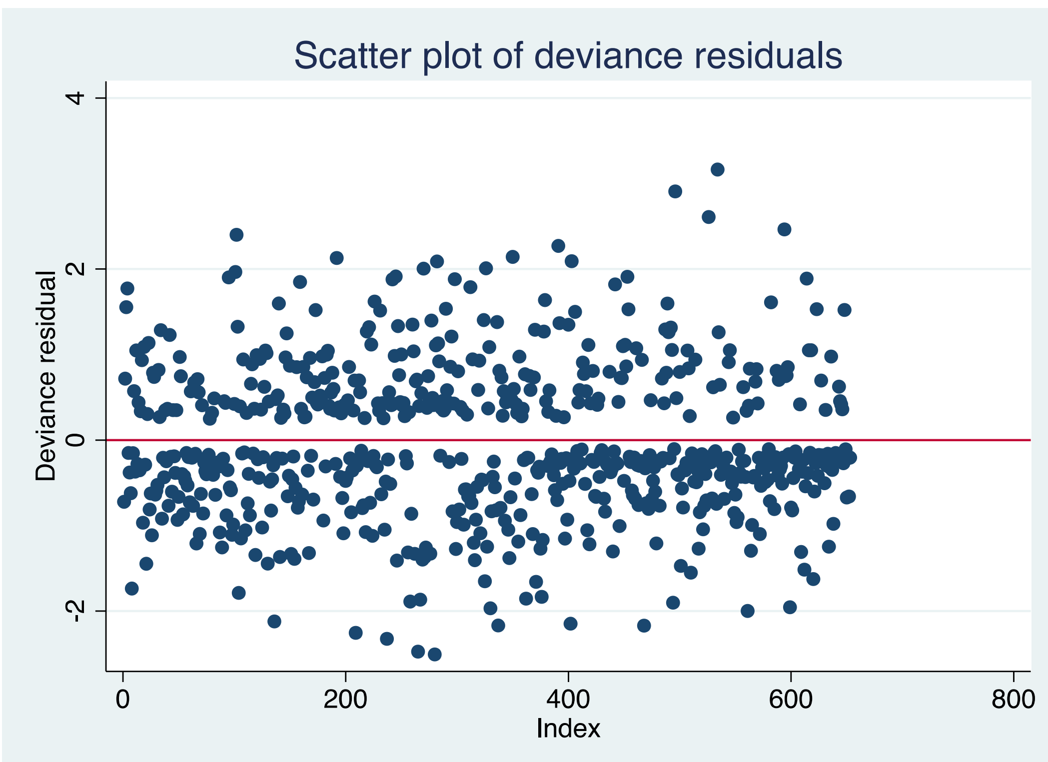

Figure 5: Scatter plot of deviance residuals

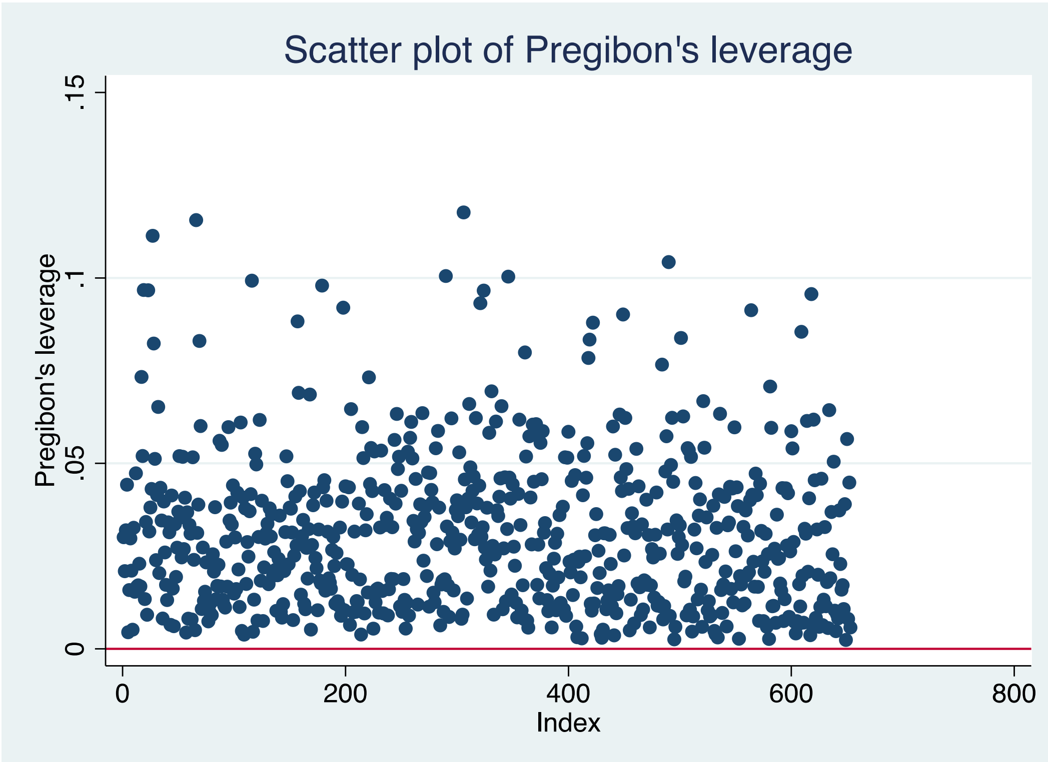

Figure 6: Scatter plot of Pregibon leverage

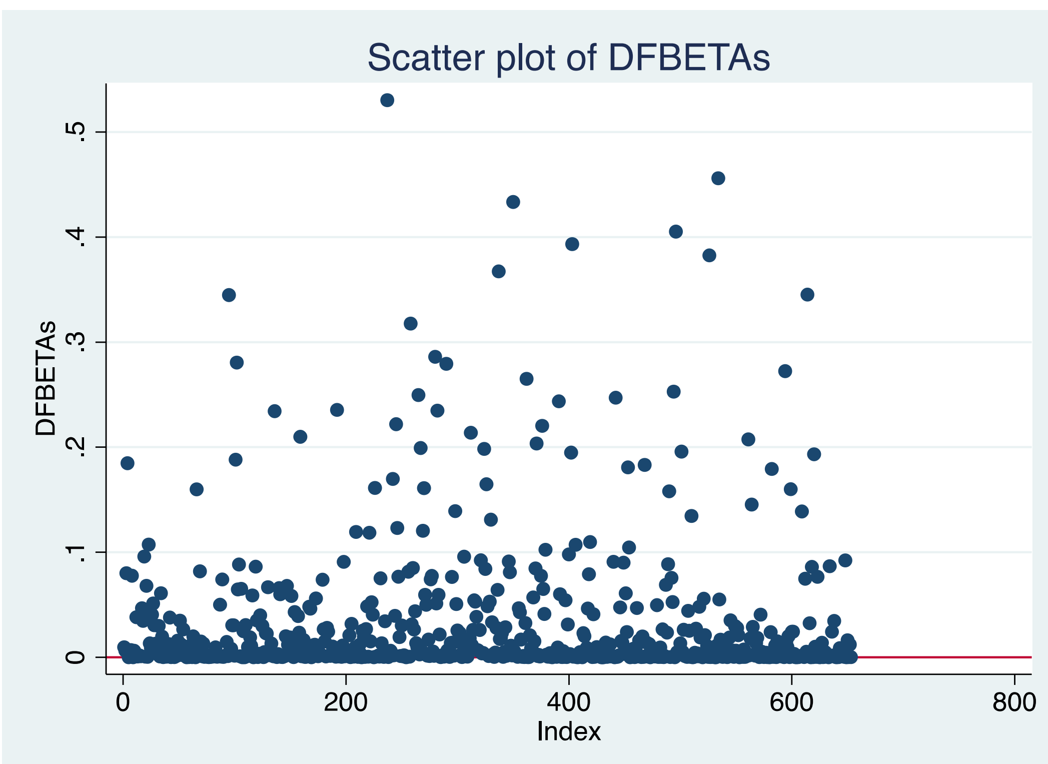

Figure 7: Scatter plot of DFBETAs

3.2.3 Multinomial logistic regression

Univariate analysis of the geographic indicator showed a not statistically significant p-value (ORs: strongly oppose = 0.612, somewhat oppose = 1.060, somewhat support = 1.215; strongly support = 1.283; global p=0.2735). In the multinomial model, when controlling for all other predictors, the effect of county of residence was not statistically significant (ORs: strongly oppose = 0.996, somewhat oppose = 1.030, somewhat support = 0.980; strongly support = 0.688; global p=0.847), and generally uninformative, as most of the odds ratios were close to one. However, being a WC resident resulted in a 31.2% decrease in the odds of strongly supporting drilling the MS for natural gas (Table 5) compared to having no opinion.

Table 5: Results of MLR on the full sample

	Odds ratio					
Predictor	Strongly oppose	Somewhat oppose	Neither oppose nor support	Somewhat support	Strongly support	Global p-value
	(95% Confidence Intervals)					
WC/AC	0.996	1.030	(baseline)	0.980	0.688	0.847
	(0.451, 2.202)	(0.542, 1.955)		(0.540, 1.779)	(0.328, 1.446)	
Follow MS issue						<0.0001
Very closely	9.486	14.862		11.170	43.022	
	(1.546, 58.195)	(3.726, 59.282)		(3.076, 40.563)	(7.209, 256.753)	
Somewhat closely	6.505	5.949		6.138	32.480	
	(1.311, 32.288)	(1.797, 19.689)		(2.376, 15.854)	(7.146, 147.631)	
A little	1.577	5.007		5.267	2.977	
	(0.306, 8.134)	(1.639, 15.302)		(2.251, 12.323)	(0.699, 12.682)	
Envir/health threat (baseline is no threat)						<0.0001
Significant threat	1.983	106.540		0.093	0.027	
	(0.424, 9.280)	(15.179, 747.771)		(0.025, 0.350)	(0.006, 0.116)	
Moderate threat	0.137	26.266		0.222	0.016	
	(0.025, 0.751)	(4.159, 165.872)		(0.078, 0.634)	(0.005, 0.053)	
Slight threat	0.052	7.726		0.410	0.184	
	(0.005, 0.586)	(1.018, 58.610)		(0.143, 1.171)	(0.060, 0.559)	
Family lease	0.983	0.956		1.530	2.735	0.168
	(0.246, 3.921)	(0.351, 2.601)		(0.618, 3.791)	(1.095, 6.828)	
Female	1.117	0.766		0.507	0.281	0.004

Table 5 Continued

	(0.478, 2.611)	(0.386, 1.521)	(0.281, 0.915)	(0.138, 0.569)	
Age	1.021	1.014	1.010	1.016	0.623
	(0.991, 1.052)	(0.990, 1.039)	(0.990, 1.031)	(0.988, 1.046)	
Education*	1.024	0.693	1.000	0.884	0.092
	(0.716, 1.464)	(0.499, 0.963)	(0.753, 1.327)	(0.611, 1.279)	
Black	0.734	1.758	1.589	1.082	0.403
	(0.204, 2.640)	(0.649, 4.767)	(0.647, 3.903)	(0.322, 3.633)	
Econ. opportunity (baseline is no opportunity)					<0.0001
Significant opportunity	0.109	0.388	4.786	2.453	
	(0.019, 0.630)	(0.074, 2.022)	(0.893, 25.659)	(0.289, 20.818)	
Moderate opportunity	0.127	0.608	3.319	0.294	
	(0.023, 0.685)	(0.124, 2.980)	(0.633, 17.393)	(0.0035, 2.496)	
Slight opportunity	0.138	0.250	0.975	0.061	
	(0.026, 0.734)	(0.047, 1.318)	(0.177, 5.382)	(0.005, 0.775)	
Overall envir. quality					0.001
Improving significantly	1.027	0.902	0.890	23.013	
	(0.117, 9.019)	(0.136, 5.993)	(0.146, 5.445)	(2.777, 190.740)	
Improving somewhat	0.587	0.852	0.744	13.117	
	(0.133, 2.593)	(0.193, 3.774)	(0.160, 3.468)	(2.105, 81.747)	
Remaining the same	0.314	1.306	0.576	5.488	
	(0.076, 1.303)	(0.313, 5.441)	(0.129, 2.563)	(0.931, 32.360)	
Getting somewhat worse	1.003	0.891	1.107	6.126	
	(0.245, 4.101)	(0.212, 3.737)	(0.228, 5.368)	(0.901, 41.667)	
Marital status					0.141
Married	0.415	0.707	0.466	0.266	
	(0.158, 1.093)	(0.320, 1.562)	(0.222, 0.978)	(0.097, 0.727)	
Widowed	0.270	0.346	0.513	0.369	
	(0.053, 1.371)	(0.090, 1.321)	(0.165, 1.595)	(0.071, 1.912)	
Divorced	1.648	1.093	0.962	0.349	
	(0.411, 6.612)	(0.329, 3.629)	(0.308, 3.000)	(0.070, 1.727)	
Length of time at residence†	0.738	0.909	0.776	0.833	0.076
	(0.558, 0.977)	(0.725, 1.141)	(0.638, 0.945)	(0.642, 1.080)	
Intercept	4.494	0.023	1.681	0.732	0.301
	(0.235, 86.095)	(0.001, 0.458)	(0.122, 23.245)	(0.028, 18.938)	

* 1=HS or less, 2=some college, 3 Bachelors, 4=Graduate school
† 1=Less than 1 year, 2=1-3 years, 3=3-5 years, 4=5-10 years, 5=10-20 years, 6=20+ years

Similar to the OLS and logit models, perception of environmental/public health threat and perception of economic opportunity were statistically significant and had similar directionality. Gender was statistically significant, with women again being more opposed to drilling then were men in three of the four logits (strongly oppose: OR=1.117, somewhat support: OR=0.507, strongly support: OR=0.281; global p=0.004), with the exception of “somewhat oppose” vs. “neither oppose nor support”. In this instance being female decreased the odds of opposing drilling by about 23%. The MLR model also showed that the more respondents believed the overall environmental quality of the region was improving, the stronger their support was for drilling the MS for natural gas (for example, with the selection “improving significantly” the odds ratios were: strongly oppose = 1.027, somewhat oppose = 0.902, somewhat support = 0.890, strongly support = 23.013; global p=0.001).

If respondents opposed drilling, the more they followed the MS issue, the more strongly they opposed. In the same vein, if respondents supported, the more they followed the MS issue, the more strongly they supported (global p<0.001). So respondents who supported drilling, supported more strongly if they follow MS, and the respondents that opposed, opposed more strongly if they follow MS. The same kind of trend occurred regarding length of time at current residence. Respondents that opposed drilling opposed more strongly the longer they had lived at their current residence. Respondents that supported drilling supported more strongly the longer they had lived at their current resident (global p=0.076).

The assumption of independence of irrelevant alternatives (IIA) was tested using the Small-Hsiao test of IIA. We rejected the null hypothesis that the odds of outcome J vs. outcome K are independent of the other alternatives (all p-values <0.001), which means that the estimated odds may change depending on which outcome was selected as the baseline outcome. In this model, due to using a five-level outcome and multiple predictors with three or more categories, the potential for small cell counts was quite high. To see if small cell counts were affecting the results, the multinomial model was then re-run using collapsed levels of the outcome. The outcome was turned into a three-level variable (1=oppose, 2=neither support nor oppose drilling, 3=support). This model (Table 6) yielded very similar results to those in Table 5. The results of the IIA test for the collapsed outcome indicated that the estimated odds would not change if the baseline category were changed (p=0.986 and p=0.940). Multicollinearity is not an issue with either multinomial model, as the largest calculated VIF is 5.72.

Table 6: Results of MLR with collapsed outcome on full sample

Predictor	Odds ratio Oppose	Neither oppose nor support	Support	Global p-value§
	(95% Confidence Intervals)			
WC/AC	0.971	(baseline)	0.850	0.836
	(0.540, 1.746)		(0.482, 1.496)	
Follow MS issue				0.0002
Very closely	13.105		12.591	
	(3.723, 46.136)		(3.629, 43.688)	
Somewhat closely	6.086		7.215	
	(2.069, 17.903)		(2.771, 18.790)	
A little	3.907		4.511	
	(1.428, 10.691)		(1.852, 10.993)	
Envir/health threat (baseline is no threat)				<0.0001
Significant threat	8.547		0.083	
	(2.190, 33.354)		(0.025, 0.279)	
Moderate threat	1.370		0.115	
	(0.383, 4.899)		(0.042, 0.316)	
Slight threat	0.335		0.238	
	(0.075, 1.492)		(0.084, 0.671)	
Family lease	0.962		1.627	0.378
	(0.376, 2.462)		(0.693, 3.823)	
Female	0.930		0.443	0.004
	(0.496, 1.745)		(0.257, 0.763)	
Age	1.001		0.995	0.690
	(0.984, 1.019)		(0.979, 1.011)	
Education*	0.787		0.958	0.212
	(0.594, 1.041)		(0.738, 1.243)	
Black	1.720		2.233	0.223
	(0.622, 4.762)		(0.893, 5.588)	
Econ. opportunity (baseline is no opportunity)				<0.0001
Significant opportunity	0.238		3.188	
	(0.054, 1.060)		(0.709, 14.338)	
Moderate opportunity	0.388		1.622	
	(0.096, 1.573)		(0.373, 7.055)	
Slight opportunity	0.220		0.423	
	(0.051, 0.942)		(0.089, 2.019)	
Political party				0.056

Table 6 Continued

Democrat	0.557 (0.281, 1.102)	0.807 (0.429, 1.519)	
Republican	0.681 (0.265, 1.751)	1.773 (0.729, 3.968)	
Intercept	0.889 (0.098, 8.063)	2.496 (0.343, 18.175)	0.544

* 1=HS or less, 2=some college, 3 Bachelors, 4=Graduate school
§ P-value is for the global test of statistical significance.

3.2.4 Ordered logistic regression

As with the OLS and logit model, univariate analysis of the geographic indicator showed that WC residents were more supportive of drilling (OR=1.380) than were AC residents (p=0.066). Controlling for additional predictors, Table 7 shows that being a WC resident resulted in a decrease in the odds of supporting drilling (OR=0.817, p=0.273).

Table 7: Results of OLR analysis on full sample

Predictor	Odds Ratio (95% Confidence Interval)	p-value
WC/AC	0.817 (0.568, 1.174)	0.273
Follow MS issue		0.014
Very closely	2.288 (1.069, 4.987)	
Somewhat closely	1.831 (0.965, 3.475)	
A little	1.010 (0.577, 1.770)	
Envir/health threat (baseline is no threat)		<0.0001
Significant threat	0.022 (0.009, 0.050)	
Moderate threat	0.126 (0.069, 0.230)	
Slight threat	0.436 (0.250, 0.758)	
Family lease	1.968 (1.177, 3.289)	0.010

Table 7 Continued

Female	0.581	0.003
	(0.407, 0.829)	
Age	0.997	0.614
	(0.987, 1.008)	
Education*	1.018	0.836
	(0.863, 1.200)	
Black	1.351	0.543
	(0.657, 2.779)	
Econ. opportunity (baseline is no opportunity)		<0.0001
Significant opportunity	16.950	
	(5.355, 53.654)	
Moderate opportunity	4.982	
	(1.706, 14.551)	
Slight opportunity	2.652	
	(0.875, 8.037)	
Gov't oversight of envir.		0.011
Increase significantly	0.297	
	(0.076, 1.155)	
Increase somewhat	0.671	
	(0.183, 2.456)	
Remain the same	0.724	
	(0.200, 2.619)	
Decrease somewhat	0.769	
	(0.198, 2.988)	
Overall envir. quality†	0.818	0.063
	(0.662, 1.011)	
Years at residence		0.308
Less than one	0.391	
	(0.160, 0.954)	
1-3	1.010	
	(0.574, 1.778)	
3-5	0.987	
	(0.464, 2.102)	
5-10	0.828	
	(0.507, 1.353)	
10-20	0.750	
	(0.453, 1.240)	

* 1=HS or less, 2=some college, 3 Bachelors, 4=Graduate school
† 1=Improving significantly, 2=Improving somewhat, 3=Remaining the same, 4=Getting somewhat worse, 5=Getting significantly worse

As shown in Table 7, the more the MS was perceived to be an environmental/public health threat, the less supportive the respondents were of drilling (significant threat: OR=0.022, moderate threat: OR=0.126, slight threat: OR=0.436; p<0.0001). Also as before, the more the MS was perceived to be an economic opportunity, the more supportive respondents were towards drilling the MS for natural gas (significant opportunity: OR=16.950, moderate opportunity: OR=4.982, slight opportunity: OR=2.652; p<0.0001). Women were still found to be less supportive of drilling in the OLR model. Those who signed a natural gas lease, or had a family member who signed a lease, were more supportive of drilling (OR=1.968; p=0.010).

As with the results of the logit, respondents who wanted to see an increase in government oversight of the environment were more opposed to drilling (increase significantly: OR=0.297, increase somewhat: OR=0.671, remain the same: OR=0.724, decrease somewhat: OR=0.769; p=0.011). The better the overall quality of the environment was perceived to be by the respondent, the more opposed the respondent was to drilling (OR=0.818; p=0.063). Those who have lived in the same area for a short period of time (less than one year) (OR=0.391) and those who have lived in the same area for an extended period of time (five to ten years: OR=0.828, ten to 20 years: OR=0.750) were also more opposed to drilling than those have lived in the area for a moderate amount of time (one to three years: OR=1.010, three to five year: OR=0.987, p=0.07).

To test the assumption of proportional odds, the Brant Test was performed. The null hypothesis that the estimated coefficients were the same across all the outcomes was rejected (p<0.0001). This could mean either that OLR was not appropriate for these data or that two or more of the levels should be collapsed due to small cell counts. The model was then re-run using collapsed categories of the outcome were 1=oppose, 2=neither support nor oppose drilling, 3=support. This model (data not shown) yielded very similar results to those in Table 7. While “years at residence” and “overall environmental quality” were omitted from the model with the collapsed outcomes, all other results were similar in both magnitude and statistical significance. When the parallel slopes assumption was tested on this model, there were still variables that violated this assumption. Again, there is no issue with multicollinearity among the predictors.

3.3 WASHINGTON COUNTY

The same backwards elimination procedure was used to analyze the sub-population of WC. A MLR methodology was used, with the condensed outcome for the sub-population analysis, as the IIA assumption was not violated when used on the full sample with the collapsed outcome. The use of the methodology resulted in many of the same predictors that were in the MLR used on the full sample ending up in the final model, in the same form.

Similar to the results of the MLR with collapsed outcomes on the full sample, a series of indicators was used to analyze the effects of following the MS issue, perception of the MS as an environmental/public health threat, and perception of the MS as an economic opportunity. As with the full sample, if they opposed drilling, the more closely respondents follow the MS issue, the more they opposed (Table 8, ORs: very closely = 18.619, somewhat closely = 7.436, a little = 4.203; global p-value = 0.045). Conversely, if a respondent supported drilling, the more closely (s)he follows the MS issue, the more (s)he supportsed drilling (ORs: very closely = 20.569, somewhat closely = 6.454, a little = 5.637; global p-value = 0.045). Also similar to the results of the full model, the more a respondent perceived the MS to be an environmental/public health threat, the more (s)he opposed drilling the MS for natural gas (Oppose: significant threat OR=3.350, moderate threat OR=0.551, slight threat OR=0.182; $p<0.0001$). The more a respondent perceived the MS to be an economic opportunity, the more supportive (s)he was of drilling (Support: significant opportunity OR=1.977, moderate opportunity OR=0.723, slight opportunity OR=0.195; $p<0.0001$). Like the results using the full sample, the WC-only sample also showed that the effect of having a family-held natural gas lease was not statistically significant (Support OR=0.852, Oppose OR=1.251; p=0.484).

Table 8: Results of MLR regression analysis with collapsed outcome on WC sample

	Odds Ratio			
	Oppose	Neither support nor oppose	Support	Global p-value§
	(95% Confidence Interval)			
Follow MS issue				0.045
Very closely	18.619		20.569	
	(3.049, 113.701)		(2.788, 151.743)	
Somewhat closely	7.436		6.454	
	(1.765, 31.331)		(1.353, 30.792)	
A little	4.203		5.637	
	(1.010, 17.499)		(1.336, 23.777)	
Envir/health threat (baseline is no threat)				<0.0001
Significant threat	3.350		0.045	
	(0.238, 47.208)		(0.004, 0.474)	
Moderate threat	0.551		0.060	
	(0.044, 6.858)		(0.008, 0.482)	
Slight threat	0.182		0.225	
	(0.012, 2.678)		(0.028, 1.806)	
Family lease	0.729		1.251	0.484
	(0.252, 2.108)		(0.494, 3.170)	
Female	0.852		0.743	0.741
	(0.366, 1.982)		(0.350, 1.578)	
Age	0.993		1.007	0.380
	(0.970, 1.016)		(0.983, 1.031)	
Education*	0.592		0.744	0.047
	(0.391, 0.896)		(0.509, 1.088)	
Black	18.963		14.616	0.100
	(1.281, 280.748)		(0.745, 286.779)	
Econ. opportunity (baseline is no opportunity)				<0.0001
Significant opportunity	0.103		1.977	
	(0.005, 2.134)		(0.145, 26.936)	
Moderate opportunity	0.176		0.723	
	(0.010, 3.205)		(0.055, 9.463)	
Slight opportunity	0.055		0.195	
	(0.003, 1.148)		(0.013, 2.982)	
Political party				0.164
Democrat	0.432		0.296	
	(0.165, 1.134)		(0.109, 0.804)	
Republican	0.545		0.590	
	(0.175, 1.704)		(0.210, 1.659)	

Table 8 Continued

Intercept	15.044 (0.278, 813.642)	9.508 (0.226, 400.216)	0.078

* 1=HS or less, 2=some college, 3 Bachelors, 4=Graduate school
§ P-value is for the global test of statistical significance.

Unlike the results on the full sample, there was no statistically significant gender effect (p=0.741). There was, however, a statistically significant education effect. While the odds ratio for having a HS education or less was slightly smaller for opposition relative to neither supporting nor opposing drilling, for both categories, a higher attainment of education resulted in greater odds of both support and opposition, relative to the baseline (Oppose: 0.592, Support: 0.744; global p=0.047). The choice of membership to a political party was not statistically significant in this model (global p=0.164). Being a Democrat reduced the odds of opposition relative to neither supporting or opposing (OR=0.432) and reduced the odds of supporting relative to neither supporting nor opposing (OR=0.296).

Unlike the results of the diagnostic tests performed on the full sample with the condensed outcome, the results from the WC-only analysis showed that these data do violate the independence of irrelevant alternatives assumption (p<0.0001 for both comparisons).

3.4 ALLEGHENY COUNTY

As with the WC-only sample, the same backwards elimination procedure was used to analyze the sub-population of AC. A backwards elimination MLR methodology was used with the condensed outcome for the sub-population analysis, as the IIA assumption was not violated when used on the full sample. With the AC-only sample, the final model was similar to that of the final model using the full sample, but not quite as similar as the results when using the WC-only sample.

As with the results of the MLR with collapsed outcomes on the full sample and the results of the WC-only sample, a series of indicators were used to analyze the effects of following the MS issue, perception of the MS as an environmental/public health threat, and perception of the MS as an economic opportunity. Also like the full sample and the WC-only sample, if they opposed drilling, the more closely respondents followed the MS issue, the more strongly they opposed (Table 9, ORs: very closely = 12.052, somewhat closely = 5.988, a little = 4.153; global p-value = 0.002). If a respondent supported drilling, the more closely (s)he followed the MS issue, the more (s)he supported drilling (ORs: very closely = 14.775, somewhat closely = 12.238, a little = 6.181; global p-value = 0.002). Also similar to the results of the both MLR models using the collapsed outcome, the more a respondent perceived the MS to be an environmental/public health threat, the more (s)he opposed drilling the MS for natural gas (Oppose: significant threat OR=22.211, moderate threat OR=4.701, slight threat OR=1.005; global p<0.0001). The more a respondent perceived the MS to be an economic opportunity, the more supportive (s)he was of drilling (Support: significant opportunity OR=4.584, moderate opportunity OR=2.321, slight opportunity OR=0.442; global p=0.0001). The results of using the full sample also showed that the effect of having a family-held natural gas lease was not statistically significant (Support OR=0.902, Oppose OR=1.406; global p=0.843). Like the results using the MLR with collapsed outcomes on the full sample, women were more opposed to drilling than are men (Oppose OR = 0.710, Support OR = 0.277; global p=0.005).

Table 9: Results of MLR regression analysis with collapsed outcome on AC sample

	Odds Ratio (95% Confidence Interval)			
	Oppose	Neither oppose nor support	Support	Global p-value§
Follow MS issue		(baseline)		0.002
Very closely	12.052 (2.221, 65.398)		14.775 (2.745, 79.519)	
Somewhat closely	5.988 (1.377, 26.034)		12.238 (3.511, 42.651)	
A little	4.153 (1.052, 166.394)		6.181 (1.811, 21.098)	
Envir/health threat (baseline is no threat)				<0.0001
Significant threat	22.211 (3.745, 131.722)		0.062 (0.013, 0.284)	
Moderate threat	4.701 (0.851, 25.967)		0.151 (0.046, 0.499)	
Slight threat	1.005		0.291	

Table 9 Continued

	(0.135, 7.494)	(0.084, 1.007)	
Family lease	0.902	1.406	0.843
	(0.137, 5.954)	(0.307, 6.449)	
Female	0.710	0.277	0.005
	(0.302, 1.669)	(0.124, 0.620)	
Age	1.031	1.009	0.203
	(0.995, 1.067)	(0.980, 1.039)	
Education*	0.907	1.044	0.796
	(0.592, 1.90)	(0.719, 1.515)	
Black	1.378	1.580	0.703
	(0.473, 4.014)	(0.525, 4.758)	
Econ. opportunity (baseline is no opportunity)			0.0001
Significant opportunity	0.258	4.584	
	(0.047, 1.422)	(0.980, 21.433)	
Moderate opportunity	0.375	2.321	
	(0.074, 1.896)	(0.487, 11.066)	
Slight opportunity	0.254	0.442	
	(0.055, 1.168)	(0.081, 2.411)	
Marital status			0.166
Married	1.069	0.549	
	(0.418, 2.735)	(0.209, 1.442)	
Widowed	0.540	0.362	
	(0.097, 3.018)	(0.076, 1.726)	
Divorced	4.135	2.306	
	(1.116, 15.314)	0.593, 8.976)	
Years at residence			0.032
Less than one	72.475	28.783	
	(5.852, 897.652)	(2.234, 370.779)	
1-3	1.526	1.990	
	(0.287, 8.120)	(0.582, 6.800)	
3-5	0.620	1.811	
	(0.129, 2.977)	(0.425, 7.720)	
5-10	0.955	0.856	
	(0.300, 3.039)	(0.256, 2.862)	
10-20	1.598	0.704	
	(0.518, 4.930)	(0.252, 1.968)	
Intercept	0.026	0.838	
	(0.001, 0.479)	(0.073, 9.548)	0.023

* 1=HS or less, 2=some college, 3 Bachelors, 4=Graduate school
§ P-value is for the global test of statistical significance.

Unlike the other two models using the collapsed levels of the outcome, this model included marital status and years at current residence as predictors. When looking at support relative to neither opposing nor supporting, the years spent at current residence appeared to have a linear decreasing trend (ORs: less than one year = 28.783, 1-3 years = 1.990, 3-5 years = 1.811, 5-10 years = 0.856, 10-20 years = 0.704), while the odds ratios for opposition relative to neither opposing nor supporting seemed to be more parabolic-shaped (ORs: less than one year = 72.475, 1-3 years = 1.526, 3-5 years = 0.620, 5-10 years = 0.955, 10-20 years = 1.598) (global $p=0.032$).

Again, unlike the results of the diagnostic tests performed on the full sample with the condensed outcome, the results from the AC-only analysis showed that these data violated the independence of irrelevant alternatives assumption ($p<0.0001$ for both comparisons).

4.0 CONCLUSIONS

4.1 DISCUSSION

We have looked at four different methods of analyzing an ordered categorical outcome. For years it was standard practice to scale an ordinal outcome, treat it as continuous and build an ordinary linear regression model (Clogg and Shihadeh, 1994). The assumption of normality appeared to have been violated by these data. However, conflicting visual analysis and statistical tests made it difficult to state with certainty. It is most likely that the significant result of the Shapiro-Wilk test was telling us the data are non-normal because the outcome was not actually continuous. All diagnostic findings were based off of analyses carried out on the unweighted survey data. As Hinkins et al. (2009) find, survey weights are important for assessing linear regression diagnostic results of survey data. While the data may plot as normal in STATA, if plotted in R, using the package specifically developed for linear regression survey data by Lumley (2009), we may see that the data are non-normal. One of the biggest issues is that the assumption of homoscedasticity was violated. The Breusch-Pagan test of homoscedasticity was statistically significant, which indicated the error term does not have a constant variance. A violation of this assumption means OLS is not the appropriate methodology to use for these data. Long (1997) states that OLS is not appropriate to use on a categorical outcome, due to heteroscedastic errors, which give misleading results.

Over the years, however, a number of issues have been discovered with the practice of using OLS with a categorical outcome. Studies comparing the effects of using OLS on a categorical outcome have yielded fairly consistent results. Taylor et al. (2006) found that using OLS on a categorical outcome resulted in a substantial loss of power, while using both an OLR and an ordered probit did a slightly better job at protecting against a loss of power. Agresti (1990) pointed out that when using OLS on a categorical outcome, the estimated responses are not restricted to the range of the outcome, allowing for a predicted response outside the range of possibilities. This possibility could be realized using the results the OLS model (Table 3). The intercept for the OLS model is 3.834, and the coefficient for "significant opportunity" as a response to the question "To what extent do you think the MS represents an economic opportunity for this region?" is 1.120. If a respondent were to select "significant opportunity as his/her response to the question, claim to be a Republican (ß=0.226) and have baseline levels for all other predictors, that would yield an estimated response value of 5.18. That response value falls outside the range of possible answers.

A categorical outcome may be collapsed into a binary outcome in order to make analysis of the results easier or because of small cell counts. In this instance, changing to a dichotomous outcome reduced the sample size by 341 respondents. The overall fit of the model was found be good, and there was no finding of model misspecification (neither through missing predictors not included nor through unnecessary predictors included). While there were multiple high leverage points and a few influential points, it was found that the deletion of any of those observations would not affect the predicted estimates. Caution should be used when reviewing the results of the model diagnostics, as they could only be run on unweighted data, and this might have resulted in artificially inflated results (Chambers and Skinner, 2003). Strömberg (1996) states that dichotomizing ordered outcomes can affect the inference drawn about the data, as the estimates may change even if the true effect of the outcome remains the same. Murad et al. (2003) also found that collapsing the levels of the outcome can produce overly conservative Wald tests, and does not improve asymptotic approximation of the test.

When analyzing the data as categorical, using all five outcomes, there were issues with the assumptions involved in the models. The assumption of independence of irrelevant alternatives was violated when using the MLR on the data, and the assumption of proportional odds was violated when using the OLR. A possible way to deal with these assumption issues is to collapse the levels of the outcome. This could potentially result in a loss of power, but Taylor et al. (2006) found that power loss when collapsing a five-level outcome is minimal.

After the levels were collapsed so that the choices were "oppose," "neither oppose nor support," and "support," both the OLR and MLR models were re-run. The proportional odds assumption was still violated when using the OLR, but the IIA assumption was met for the MLR. As the MLR used on the data with the collapsed outcome levels appeared to be the best model, it was then used on the WC-only and AC-only sub-populations. With both sub-populations the IIA assumption was again violated. This, however, could have occurred due to the reduced sample sizes being used (Cheng and Long, 2007). Small cell counts could have contributed to the expected variations in the estimates, depending on the baseline outcome selected.

The results of the test of the proportional odds assumption need to be considered when trying to determine if OLR is the best fit to estimate the model. Sobel (1997) points out that using an OLR when this assumption is violated can lead to estimates that do not accurately reflect how the outcome is determined by the predictors, and that incorrect estimates lead to incorrect inferences about the data. However, Long (2012) noted that the proportional odds assumption is often violated, and that a comparison of the estimates from the OLR and MLR should be made before dismissing the OLR as inappropriate. Because the IIA assumption for MLR and the proportional odds assumption for OLR are both frequently violated, a comparison of the results from these two models needed to be made.

As seen in Table 10, some of the estimated ORs in the MLR and OLR were similar. For example, the effect of county of residence in the two models were similar in that they were all close to one and statistically not significant. Additionally, effect of perception of environmental/public health threat, family lease-holding and gender were similar. Also seen in Table 10 were differences in the results between the two methods. Perhaps the most notable difference was in the estimated effects of perception of the MS as an economic opportunity. The results of the OLR indicated that if a respondent felt that the MS is any kind of economic opportunity, (s)he would increase his/her support for drilling (significant opportunity = 16.950, moderate opportunity = 4.982, slight opportunity = 2.652; $p<0.0001$). The results of the MLR showed that perception of the MS as an economic opportunity increased support relative to neither supporting nor opposing drilling, too, but with a lesser effect (Strongly support ORs: significant threat = 2.453, moderate threat = 0.294, slight threat = 0.061; global $p<0.0001$). The

MLR also showed perception of the MS as an economic opportunity decreased the odds of opposing drilling relative to neither supporting nor opposing drilling (Strongly oppose ORs: significant threat = 0.109, moderate threat = 0.127, slight threat = 0.138; global p<0.0001).

Table 10: Comparison of results of MLR and OLR

Predictor	Multinomial logit OR					Ordered logit OR
	Strongly oppose	Somewhat oppose	Neither oppose nor support	Somewhat support	Strongly support	
	(95% Confidence Intervals)					
WC/AC	0.996	1.030	(baseline)	0.980	0.688	0.817
	(0.451, 2.202)	(0.542, 1.955)		(0.540, 1.779)	(0.328, 1.446)	(0.568, 1.174)
Follow MS issue						
Very closely	9.486***	14.862***		11.170***	43.022***	2.288*
	(1.546, 58.195)	(3.726, 59.282)		(3.076, 40.563)	(7.209, 256.753)	(1.069, 4.987)
Somewhat closely	6.505***	5.949***		6.138***	32.481***	1.831*
	(1.311, 32.288)	(1.797, 19.689)		(2.376, 15.854)	(7.146, 147.631)	(0.965, 3.475)
A little	1.577***	5.007***		5.267***	2.978***	1.010*
	(0.306, 8.134)	(1.639, 15.302)		(2.251, 12.323)	(0.699, 12.682)	(0.577, 1.770)
Envir/health threat (baseline is no threat)						
Significant threat	1.983***	106.540***		0.093***	0.027***	0.022***
	(0.424, 9.280)	(15.179, 747.771)		(0.025, 0.350)	(0.006, 0.116)	(0.009, 0.050)
Moderate threat	0.137***	26.267***		0.222***	0.016***	0.126***
	(0.025, 0.751)	(4.159, 165.872)		(0.078, 0.634)	(0.005, 0.053)	(0.069, 0.230)
Slight threat	0.052***	7.726***		0.410***	0.184***	0.436***
	(0.005, 0.586)	(1.018, 58.610)		(0.143, 1.171)	(0.060, 0.559)	(0.250, 0.758)
Family lease	0.983	0.956		1.530	2.735	1.968**
	(0.246, 3.921)	(0.351, 2.601)		(0.618, 3.791)	(1.095, 6.828)	(1.177, 3.289)
Female	1.117**	0.766**		0.507**	0.281**	0.581**
	(0.478, 2.611)	(0.386, 1.521)		(0.281, 0.915)	(0.138, 0.569)	(0.407, 0.829)
Age	1.021	1.014		1.010	1.016	0.997
	(0.991, 1.052)	(0.990, 1.039)		(0.990, 1.031)	(0.988, 1.046)	(0.987, 1.008)
Education§	1.024	0.693		1.000	0.884	1.018
	(0.716, 1.464)	(0.499, 0.963)		(0.753, 1.327)	(0.611, 1.279)	(0.863, 1.200)
Black	0.734	1.758		1.589	1.082	1.351

Table 10 Continued

	(0.204, 2.640)	(0.649, 4.767)	(0.647, 3.903)	(0.322, 3.633)	(0.657, 2.779)
Econ. opportunity (baseline is no opportunity)					
Significant opportunity	0.109***	0.388***	4.786***	2.453***	16.950***
	(0.019, 0.630)	(0.074, 2.022)	(0.893, 25.659)	(0.289, 20.818)	(5.355, 53.654)
Moderate opportunity	0.127***	0.608***	3.319***	0.294***	4.982***
	(0.023, 0.685)	(0.124, 2.980)	(0.633, 17.393)	(0.0035, 2.496)	(1.706, 14.551)
Slight opportunity	0.138***	0.250***	0.975***	0.061***	2.652***
	(0.026, 0.734)	(0.047, 1.318)	(0.177, 5.382)	(0.005, 0.775)	(0.875, 8.037)
Gov't oversight of envir. should . .					
Increase significantly					0.297*
					(0.076, 1.155)
Increase somewhat					0.671*
					(0.183, 2.456)
Remain the same					0.724*
					(0.200, 2.619)
Decrease somewhat					0.769*
					(0.198, 2.988)
Overall envir. quality					0.818
					(0.662, 1.011)
Improving significantly	1.027***	0.902***	0.890***	23.0132***	
	(0.117, 9.019)	(0.136, 5.993)	(0.146, 5.445)	(2.777, 190.740)	
Improving somewhat	0.587***	0.852***	0.744***	13.117***	
	(0.133, 2.593)	(0.193, 3.774)	(0.160, 3.468)	(2.105, 81.747)	
Remaining the same	0.314***	1.306***	0.576***	5.488***	
	(0.076, 1.303)	(0.313, 5.441)	(0.129, 2.563)	(0.931, 32.360)	
Getting somewhat worse	1.003***	0.891***	1.107***	6.126***	
	(0.245, 4.101)	(0.212, 3.737)	(0.228, 5.368)	(0.901, 41.667)	
Marital status					
Married	0.415	0.707	0.466	0.266	
	(0.158, 1.093)	(0.320, 1.562)	(0.222, 0.978)	(0.097, 0.727)	
Widowed	0.270	0.346	0.513	0.369	
	(0.053, 1.371)	(0.090, 1.321)	(0.165, 1.595)	(0.071, 1.912)	
Divorced	1.648	1.093	0.962	0.349	
	(0.411, 6.612)	(0.329, 3.629)	(0.308, 3.000)	(0.070, 1.727)	
Years at residence	0.738	0.909	0.776	0.833	
	(0.558, 0.977)	(0.725, 1.141)	(0.638, 0.945)	(0.642, 1.080)	
Less than one					0.391
					(0.160, 0.954)
1-3					1.010

Table 10 Continued

					(0.574, 1.778)
3-5					0.987
					(0.464, 2.102)
5-10					0.828
					(0.507, 1.353)
10-20					0.750
					(0.453, 1.240)
Intercept	4.494	0.023	1.681	0.732	
	(0.235, 86.095)	(0.001, 0.458)	(0.122, 23.245)	(0.028, 18.938)	

§ 1=HS or less, 2=some college, 3 Bachelors, 4=Graduate school
* significant at 0.05
** significant at 0.01
*** significant ≤0.001

There were some differences in the variables included in the two models as well. In the OLR model perception of state government oversight of the environment was included, as well as years at residence, while in the MLR model perception of overall environmental quality and marital status were included. Marital status and perception of overall environmental quality were not statistically significant in their respective models. Most of the estimated ORs in these predictors had quite wide confidence intervals and/or confidence intervals that contain a value of one. This would indicate that the results need to be cautiously interpreted, and perhaps that there were too few observations in the cells to get a more accurate and narrower estimate.

Akkus and Ozkoc (2012) compared the results of using a MLR and OLR on survey data regarding politics in the European Union. They found that, while the IIA assumption was violated for the MLR model, the proportional odds assumption was met with the OLR. The data used for their study had many more observations (n=56,752) than did this study, and the descriptive statistics showed that the smallest cell count for any of the variables segmented by outcome level was 541. When these numbers are compared to the numbers in this study, it would seem reasonable to wonder if the results of the Akkus and Ozkoc (2012) study found that OLR was more appropriate because of the much larger sample used, and to wonder if the proportional odds assumption would hold with this study if it had a similar sample size.

4.2 LIMITATIONS

Assessments of the regression assumptions were limited by the fact that survey weights cannot be applied when running model diagnostics. Rao and Thomas pointed out that when performing logistic regression on survey data such as the data used in this paper, testing of model fit is not permitted (Chambers and Skinner, 2003). Because there have been no studies investigating the effects of diagnostic analyses performed on unweighted survey data in a logistic regression setting, how the results may have varied is unknown. As such, all determinations made regarding the different methodologies should be interpreted cautiously.

By many standards, the sample size (n=1301, unweighted) may be considered large. However, when the number of predictors included in the model, survey weighting, and the fact that the outcome contained five categories are taken into account, it does not seem quite as large. Hosmer and Lemeshow (2004) recommend a minimum of ten observations per independent variable when performing logistic regression. That is for binary logistic regression, however, and there is no real formula for multinomial logistic regression. If the same standard is used, and it multiplied by the number of levels of the outcome being compared (*J-1*), then for this model we should use a minimum of 40 observations per independent variable. With 29 independent variables in the MLR, that means a minimum of 1,160 observations would be recommended. As there were 940 observations used in the MLR model, we would be short of meeting the suggested minimum by 220 observations. There were 30 predictors in the OLR model. By using the same logic for the OLR model the minimum recommended number of observations would be 1,200. There were 858 observations used in the OLR, resulting in an even larger gap (342 observations) in the suggested number of observations to be used. An attempt to use a more parsimonious model may have resulted in confounding issues, so the obvious solution for further study would be to use a larger sample size. Unfortunately for this study, it is not feasible to increase the sample size due to time and financial constraints of both the surveyors and the respondents. Perhaps, if the survey were to be assessed on another population, the responses could be changed so that an obvious order may be observed, eliminating (hopefully) the need to choose between OLR and MLR. Inclusion of more continuous predictors in the survey would also help reduce the number of observations needed.

Design of the survey questions could be altered so that the scope of the questions is narrower and less subjective to each respondent. Instead of asking respondents to choose on a scale that ranged from “strongly oppose” to “strongly support,” they could be asked a modified version of the question, “Considering everything, how do you feel about natural gas extraction from the MS?” The new survey question could be phrased as, “Considering everything, how much do you support natural gas extraction from the MS?” The responses could be kept on a five-category Likert scale, but changed to be set up as a “grouped continuous” variable (McCullagh, 1980), 1 = “0-20% support” (formerly “strongly opposed”), 2 = “21-40% support” (formerly “somewhat opposed”), 3 = 41-60% support (formerly “neither support nor oppose”), 4 = 61-80% support (formerly “somewhat support”), and 5 = “80-100% support” (formerly “strongly support”). To avoid potential confusion from respondents who do not support drilling or have no opinion on drilling, interviewers could be instructed to explain that a very low percent is equivalent to stating opposition to drilling and a middle-range percent can be thought of as more undecided than for or against.

4.3 FURTHER STUDIES

While the results of this investigation were not conclusive, it did highlight a number of areas which merit further investigation. There have been numerous studies that have tried to determine which methodology is most appropriate to use for a categorical outcome. However, there have been very few that have attempted to make the distinction when using survey data. Of the research that has used survey data, little to no mention has been made as to how survey weighting can affect the estimates, or even if survey weighting was used. Even rarer is a discussion of how to determine if models for survey-weighted data meet the requirements necessary to use the data.

There has been much discussion as to whether or not tests for IIA and parallel slopes should be used as absolute determinants of model fit. These tests are not available in the current software packages for use with survey data. Any model diagnostics performed on regression models that utilize survey data must be done without survey-weighting the data, as current statistical software packages (STATA, SAS, SPSS, etc.) do not permit survey weights to be applied when checking assumptions. It is known that standard errors are underestimated when survey weights are not used on survey data, so this implies that other errors are being introduced when survey weights are not used when checking model assumptions. Given this, future methodological research should include model diagnostic testing for models that incorporate survey weighting. With hospitals, governments and employers using quality of life surveys to make assessments of performance, it would seem that these types of diagnostics would be important contributions to the field of survey research.

4.4 CONCLUDING REMARKS

Regardless of the type of regression methodology used, there were some consistencies in the estimated results. With the inclusion of additional predictors the effect of county of residence was not statistically significant in any of the estimated models. This indicated that, while on its own, county of residence might appear to play a role in determining support or opposition to drilling, including additional predictors explained the perceived difference. However, regardless of whether the MLR or the OLR is used, for the sample, with all other predictors held constant, being a WC resident decreases support of drilling. This result is contrary as to what was expected. It was believed that those who stood a better chance of benefitting from drilling (through an increase in local business and employment, etc.) would be more supportive. In spite of WC being a better position to benefit due to higher levels of drilling activity, residents appear to be slightly less supportive than are AC residents. This result may be occurring due to surveying respondents who have been around MS drilling activity for an extended period of time and are now seeing more the perceived negative effects. Other studies, such as Anderson and Theodori (2009) found that environmental concerns, specifically regarding potential water contamination, were big issues with community leaders. Housing shortages, an increase in crime, an increase in noise pollution, and an increase in odor pollution have also been raised as concerns among residents in areas were hydraulic fracturing were occurring (Anderson and Theodori, 2009; Subra, 2009; Blevins et al., 2004).

In all the estimated models using the full sample, gender, perception of the MS as an environmental/public health threat and perception of the MS as an economic opportunity were always statistically significant. However, as can be seen when comparing the results in Tables 3-7, those were the only real similarities in estimated effects and statistical significance across the models. While additional variables were included in the model, using an alpha of 0.10, to control for potential confounding, most of them ended up not being very influential in determining support for or opposition to drilling.

The findings from the survey in this paper are different from those in the previous study involving these data and those of other similar studies using mail surveys. The biggest difference found is that, controlling for all other predictors, the residents in the county with higher levels of drilling activity (WC) are actually less likely to support drilling the MS for natural gas than residents in the county with lower levels of drilling activity (AC). In Kriesky et al. (2013), they found that, using the same data, being a resident of WC makes the respondent more supportive of drilling than being a resident of AC. While the same methodology was employed in the Kriesky et al. (2013) paper, different predictors were used; the Kriesky et al. (2013) paper focused more on where respondents were getting their information about the MS. This paper considered overall environmental quality, desire for government intervention in the environment, marital status, and political party as possible predictors. Missing observations in the different predictors contributed to different overall sample sizes between these two sets of results. It also should be noted that in the Kriesky et al. (2013) paper, while they found that being a WC resident increases support of drilling this result was also not statistically significant and was very close to one (OR=1.085, p=0.656).

However, similar to the findings of Kriesky et al. (2013), perception of the MS as an economic opportunity and having a family owned lease increase support of drilling the MS for natural gas, while perception of the MS as an environmental/public health threat and being female decrease support of drilling. Jacquet (2012) also found that women and those who claim to be more concerned about potential environmental impacts of UGD are more opposed to drilling. Concern about the potential environmental effects of UGD and women being more concerned about environmental issues are common results in investigations such as this one (Alter et al., 2010; Brasier et al., 2011; Anderson and Theodori, 2009; Theodori, 2009). Additional surveys also found economic and environmental health perspectives played a significant role in determining level of support of or opposition to drilling. Similar to the findings in this study, respondents who believe that UGD will provide an economic benefit to their region are more supportive of drilling, and those who believe UGD will cause environmental or health issues are less supportive of drilling (Theodori, 2009; Alter et al., 2010; Jacquet, 2012). Those who believe that they will benefit from signing a mineral rights lease are also frequently much more supportive of drilling (Jacquet, 2012; Jacquet, 2005; Alter et al, 2010). Issues such as water and air pollution, increased crime rates and safety issues have frequently been listed as reasons why respondents believe that UGD will negatively affect them (Anderson and Theodori, 2009; Witter et al., 2008, Jacquet, 2005; Brasier et al., 2011). These findings are all consistent with the estimated impact of the predictors used in this study.

The OLR shows that, once other factors are controlled for, those in WC are less supportive of drilling the MS for natural gas (OR=0.817, p=0.273). In spite of a lack of a clear delineation between the levels of the outcome, using OLR to analyze survey data with this kind of categorical outcome is the best option. The most influential predictors in both the MLR and the OLR had similar influence on determining the outcome, and it is this investigator's belief that, given a large enough sample size, the estimates would be even closer, with smaller confidence intervals, consistent with Akkus and Ozkoc's (2012) findings.

Given that the estimates of most of the highly significant predictors were similar when comparing the OLR and MLR, it would appear that the OLR is the best model to use. The differences that result when modeling are most likely explained by differences in cell counts due to more and fewer equations being estimated (Cheng and Long, 2007). It is entirely possible that if a larger sample was used for this current study, the parallel slopes assumption would not be violated, and more conclusive evidence would be available for using OLR.

BIBLIOGRAPHY

Agresti, A., 1990. Categorical data analysis. New York, NY, Wiley-Interscience Publication.

Agresti, A., 1999. Modelling ordered categorical data: recent advances and future challenges. Statistics in Medicine, 18(17-18), 2191-2207.

Akkus, O., Ozkoc, H., 2012. A comparison of the models over the data on the interest level in politics in Turkey and countries that are members of the European Union: Multinomial or ordered logit model? Research Journal of Applied Sciences, Engineering and Technology, 4(19): 3646-3657.

Alter, T., Brasier, K., McLaughlin, D., Willits, F., Ooms, T., Tracewski, S., 2010. A baseline socioeconomic analysis for the Marcellus Shale development in Pennsylvania. Pennsylvania: Appalachian Regional Commission. Available from: http://www.institutepa.org/PDF/Marcellus/MarcellusShaleStudy08312010.pdf.

Anderson, J.A., 1984. Regression and ordered categorical variables. Journal of the Royal Statistical Society, Series B (Methodological), 46(1), 1-30.

Anderson, B. J., Theodori, G. L., 2009. Local leaders' perceptions of energy development in the Barnett Shale. Southern Rural Sociology, 24(1), 113–129.

Andrews, A., Folger, P., Humphries, M., Copeland, C., Tiemann, M., Meltz, R., Brougher, C., 2009. Unconventional gas shales: Development, technology, and policy issues. Congressional Research Service, CRS Report for Congress. Available from: http://www.fas.org/sgp/crs/misc/R40894.pdf.

Blevins, A., Jensen, K., Coburn, M., Utz, S., 2004. Social and economic impact assessment of Sublette and Sweetwater Counties. Prepared for Shell Exploration and Production Company by the University of Wyoming Department of Sociology. Wyoming: University of Wyoming. Available from: ⟨http://www. sublettewyo.com/DocumentCenter/Home/View/368⟩.

Brasier, K.J., Filteau, M.R., Jacquet, J., Stedman, R.C., Kelsey, T.W., Goetz, S.J., 2011. Residents' perceptions of community and environmental impacts from development of natural gas in the Marcellus Shale: A comparison of Pennsylvania and New York cases. Journal of Rural Social Sciences, 26(1), 32-61.

Browder, C., 2012. Perdue open to 'fracking' in NC. wral.com. Available from: http://www.wral.com/news/state/nccapitol/story/10856881/.

Cauchon, D., 2012. States eye trusts for new oil, gas revenue. USA Today. Available from: http://www.usatoday.com/story/news/nation/2012/11/19/states-oil-gas-revenue/1715851/.

Chambers, R. L., & Skinner, C. J. (Eds.), 2003. Analysis of survey data, Wiley Series in Survey Methodology, Vol. 338.

Cheng, S., Long, J. S., 2007. Testing for IIA in the multinomial logit model. Sociological methods and research, 35(4), 583-600.

Chevron, 2011. Natural gas from shale: Unlocking energy from shale gas formations. Available from: ⟨http://www.chevron.com/deliveringenergy/naturalgas/shale gas/?utm_campaign=Sha le&utm_medium=cpc&utm_source=Google&utm_term=Hydraulic_Fracturing⟩.

Clogg, C. C., Shihadeh, E. S., 1994. Statistical models for ordinal variables. Thousand Oaks, CA, SAGE Publications.

Downing, B., 2012. Americans against fracking formed by 100 groups. Akron Beacon Journal Online. Available from: ⟨http://www.ohio.com/blogs/drilling/ohio-uti ca-shale-1.291290/americans-against-fracking-formed-by-100-groups-1. 356203⟩.

Hinkins, S., Mulrow, E., Scheuren, F., 2009. Visualization of Complex Survey Data: Regression Diagnostics. Available from American Statistical Society, Section on survey research methods: http://www.amstat.org/Sections/Srms/Proceedings/y2009/Files/303956.pdf.

Hosmer, D. W., and Lemeshow, S., 2004. Applied logistic regression (Vol. 354). Wiley-Interscience.

Jacquet, J., 2005. Index crimes, arrests, and incidents in Sublette county 1995–2004. Sublette community partnership, Wyoming. Available from: ⟨http://www.sublettewyo.com/DocumentCenter/Home/View/351⟩.

Jacquet, J., 2009. Energy Boomtowns and Natural Gas: Implications for Marcellus Shale Local Governments and Rural Communities. NERCRD Rural Development Paper no. 43. State College, PA: Northeast Regional Center for Rural Development.

Jacquet, J., 2012. Environmental Concern and Industry Experience: Landowner Attitudes Toward Natural Gas and Wind Farm Development in Northern Pennsylvania, in: Jacquet J. (first ed.), Landowner Attitudes and Perceptions of Impact from Wind and Natural Gas Development in Northern Pennsylvania: Implications for Energy Landscapes in Rural America. Cornell University, New York.

Jacquet, J., Stedman, R., 2011. Natural gas landowner coalitions in New York State: emerging benefits of collective natural resource management. Journal of Rural Social Sciences. 26(1), 62.

Kriesky, J. K., Goldstein, B. D., Zell, K. M., Beach, S., 2013. Differing Opinions about Natural Gas Drilling in Two Adjacent Counties with Different Levels of Drilling Activity. Energy Policy, http://dx.doi.org/10.1016/j.enpol.2013.03.005i.

Long, J. S., 1997. Regression models for categorical and limited dependent variables. Thousand Oaks, CA, Sage Press.

Long, J. S., 2012. Regression models (forthcoming). Best and Wolf, Sage Publications.

Lumley, T., 2009. Survey: analysis of complex survey samples. R package version 3.16.

McAllister, E., 2013. NY fracking decision faces further delay on health study. Thomas Reuters. Available from: http://newsandinsight.thomsonreuters.com/Legal/News/2013/02_-_February/NY_fracking_decision_faces_further_delay_on_health_study/.

McCullagh, P., 1980. Regression models for ordinal data (with discussion). Journal of Royal Statistical Society, Series B (Methodological), 42(2), 109-142.

McFadden, D., 1974. The measurement of urban travel demand. Journal of public economics, Vol. 3(4), 303-328.

Murad, H., Fleischman, A., Sadetzki, S., Geyer, O., Freedman, L. S., 2003. Small samples and ordered logistic regression: Does it help to collapse categories of outcomes? The American Statistician, 57(3), 155-160.

O'Day, S., Reece, J. L., 2012. Top environmental concerns in fracking. Oil and gas monitor. Available from: ⟨http://www.oilgasmonitor.com/top-environmental- concerns-fracking/1557/⟩.

OSHA, 2006. Toxic and Hazardous Substances, Standard Number: 1910.1000 TABLE Z-1, Title: Table Z-1 Limits for Air Contaminants. Available from: http://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=standards&p_id=9992.

PA Environmental Digest, 2009. DEP Detects total dissolved solids over standards in Monongahela River. Available from: http://www.paenvironmentdigest.com/newsletter/default.asp?NewsletterArticleID=13795

PADEP, 2012. DEP Office of Oil and Gas management permits issued report. Harrisburg, PA. Available from: http://www.depreportingservices.state.pa.us/ReportServer/Pages/ReportViewer.aspx?/Oil_Gas/Permits_Issued_Detail.

Resources for the Future, 2012. A Review of Shale Gas Regulations by State. Centers for Energy Economics and Policy. Available from: http://www.rff.org/centers/energy_economics_and_policy/Pages/Shale_Maps.aspx.

Sobel, M., 1997. Modeling symmetry, asymmetry, and change in ordered scales with midpoints using adjacent category logit models for discrete data. Sociological Methods and Research, 26(2), 213-232.

Strömberg, U., 1996. Collapsing ordered outcomes categories: A note of concern. American Journal of Epidemiology, 144(4), 421-424.Taylor, A., West, S., Aiken, L., 2006. Loss of power in logistic, ordinal logistic, and probit regression when an outcome variable is coarsely categorized. Educational and Psychological Measurement, 66(2), 228-239.

Subra, W., 2009. Health survey results of current and former DISH/Clark, Texas residents. Earthworks' _Oil and Gas Accountability Project. New Iberia, LA. Available from: ⟨http://www.earthworksaction.org/library/detail/health_survey_results_of_current_and_former_dish_clark_texas_residents/⟩.

Taylor, A., West, S., Aiken, L., 2006. Loss of power in logistic, ordinal logistic, and probit regression when an outcome variable is coarsely categorized. Educational and Psychological Measurement, 66(2), 228-239.

Theodori, G.L., 2009. Paradoxical perceptions of problems associated with unconventional natural gas development. Southern Rural Sociology, 24(3), 97.

US Census Bureau, 2011. State & County QuickFacts. Available from: http://quickfacts.census.gov/qfd/index.html.

Vittinghoff, E., Glidden, D. V., Shiboski, S. C., & McCulloch, C. E. (2011). Regression methods in biostatistics: linear, logistic, survival, and repeated measures models. Springer.

Williams, R., 2006. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. Stata Journal Vol. 6(1): 58-82.

Witter, R., Stinson, K., Sackett, H., Putter, S., Kinney, G., Teitelbaum, D., Newman, L., 2008. Potential exposure-related human health effects of oil and gas development: A white paper. Available: ⟨http://docs.nrdc.org/health/files/hea_08091702a.pdf⟩.