BMI Small Area Estimates: Using a Generalized Linear Mixed Multilevel Model

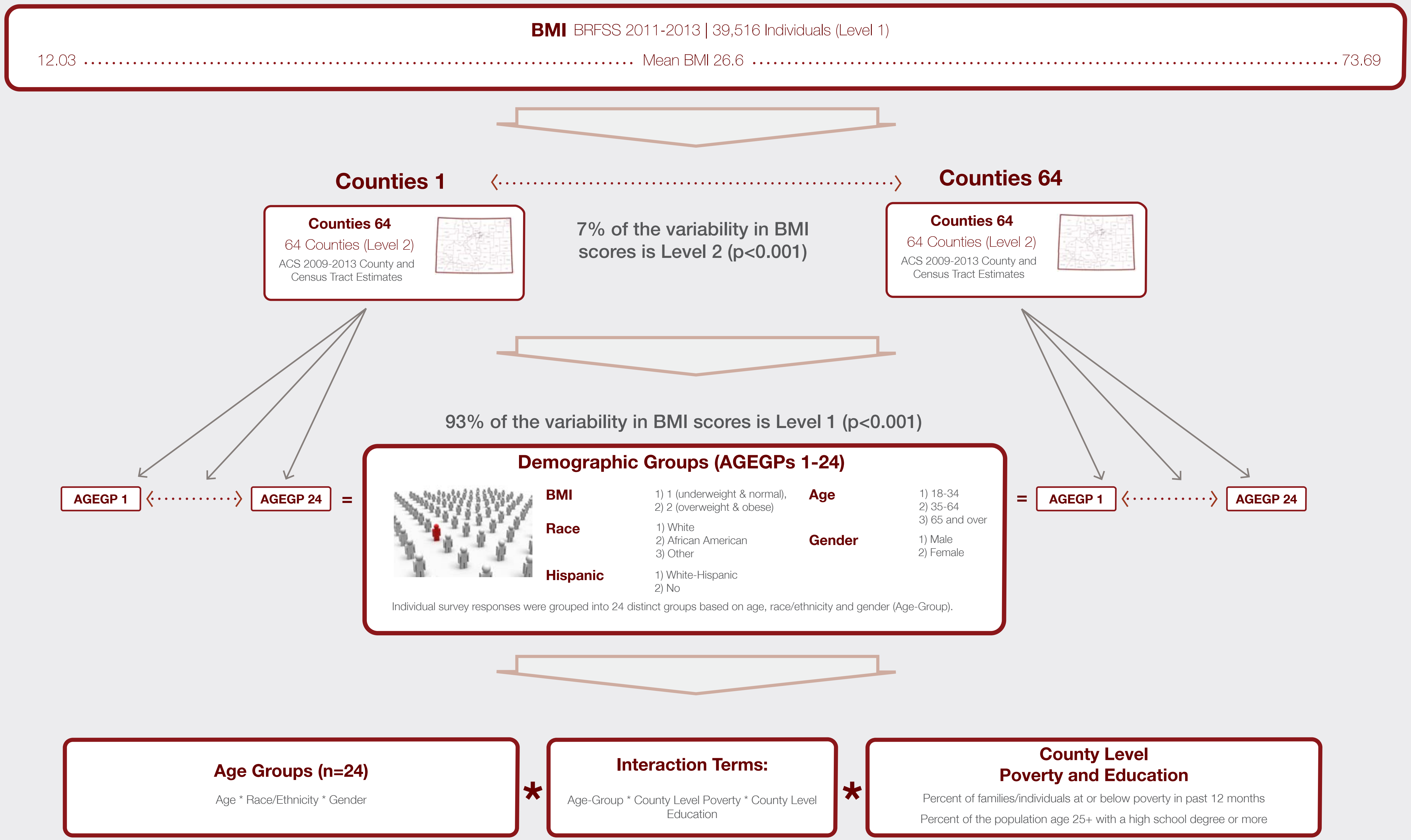
Current health surveillance systems struggle to generate health outcome estimates at geographies smaller than the state level. Some states, such as Colorado, have expanded sampling to develop reliable county level health estimates. However even within counties, there is considerable variability that may occur and a county level estimate may not provide enough detail. Smaller geographies, such as census tracts, are often needed to understand the degree of a problem and hone in on specific populations.

Small area models are statistical models used to generate health outcome estimates at a geography smaller than possible with traditional surveillance methods. In examining BMI outcomes (overweight/obese), we fit a multilevel model using individual Behavioral Risk Factor Surveillance System (BRFSS) data in addition to socio-demographic and contextual information from the U.S. Census (ACS). Individuals' results are nested within geographic boundaries (counties) where both individual characteristics (demographic) as well as location characteristics are used to model the probability of being overweight/obese. We can begin to account for the variability occurring between groups and locations by incorporating random effects into the model.

The multilevel model we use is a generalized linear mixed multilevel model. We model individual level BRFSS weighted survey responses 2011-2013 (n=36,719) grouped within counties (n=64) and demographic groups (n=24). The outcome variable Overweight and/or Obese (Yes/No) was based on self reported height and weight from individual survey responses. With SAS 9.3 we run PROC GLIMMIX to calculate an odds ratio and predicted probability for each demographic group (age*race*sex) for each county. Using 2009-2013 American Community Survey 5-Year Estimates for census tracts stratified by age, race and gender; we use the county demographic group predicted probabilities to calculate the estimated number of individuals who are overweight/obese (this calculation is based on the assumption that age group # in county # will have the same outcome throughout the census tracts within that county).

The model was estimated using the LaPlace estimation based on examples from previously documented SAE. We evaluate model fit using a likelihood ratio test (~chi-square difference) comparing values in -2Log Likelihood values. We also evaluate differences in AIC and BIC values between models. The predicted probabilities are estimated from covariate data from all the counties, not just from a single county. The use of all available data to model BMI leads to an increase in the effective sample size for a given area allowing for estimates for geographies with limited survey data available.

Model Setup



BRFSS Overweight/Obese Status (Yes or No) = Sex+Age+Race/Ethnicity (Individual Level) + Education (County Level) + Poverty (County Level) + Age-Group * County Level Poverty * County Level Education (Interaction) * Random Effect (Individual and County Level)

Figure 1

Colorado Department of Public Health and Enviornment
Center for Health and Environmental Data

Process steps using a multilevel regression model for small area BMI BRFSS estimates

COLORADO Department of Public Health & Environment

Updated: June 2015

Model Framework

The first model we fit is a null model that has no independent variables, only a random effect for the intercept. This model allows us to obtain estimates for the variance for residuals and intercept when only clustering by county is considered.

Eq. 1 $\eta_{ij} = \beta_{0j} + \beta_{1j} X_{ij}$

The next model (Eq. 1) is a level-1 model with one individual level predictor. η_{ij} represents the log odds of being overweight/obese for individual *i* in county *j*. β_{0j} is the average log odds of being overweight/obese in county *j*. X_{ij} is the individual level variable for individual *i* in county *j*. β_{1j} is the slope for X_{ij}. This slope describes the relationship between the individual level variable, demographics, and the outcome variable, overweight/obese.

*Generalized multilevel models assume no error at level-1, so in order to calculate ICC we assume the dichotomous outcome comes from an unknown continuous latent variable with a level-1 residual that follows a logistic distribution with a mean 0 and variance 3.29.

Eq. 2 $\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j} \qquad \beta_{1j} = \gamma_{10}$

Equation 2 expands on the previous model by adding one county or level-2 predictor variable, γ_{00}. γ_{00} is the log odds of being overweight/obese in an average county. W_j is a county level predictor for county *j* (county level poverty, income, education, urban-rural). γ_{01} is the slope for W_j. u_{0j} is the level-2 error term or random variable associated with county *j*. γ_{10} is the average effect of the individual level predictor.

Eq. 3 $\eta_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} W_j + u_{0j}$

Equation 3 is a combination of Equation 1 & 2, where Equation 2 terms are substituted into Equation 1. As follows, the log odds of being overweight/obese for individual *i* in county *j* is calculated by: the log odds of being overweight/obese of an average individual in an average county, the effect of the individual demographic characteristics, county level predictors and county level error.
From this equation, additional error terms and interactions are added into the final model based on a theoretical framework, previous work and model fit statistics.

Model Variables

Variable	Label	Type	Notes	Source
Overweight/Obese	_RFBMI5	Dependent	1) Underweight or normal weight 2) Overweight or obese	BRFSS 2011-2013
County	COUNTY	+ Random	County of residence	BRFSS 2011-2013
Age, Gender, Race/Ethnicity	AGEGP	+ Random	BRFSS Variables: AGE, _MRACE1, SEX	BRFSS 2011-2013
Education	EDU	Fixed	County Level Education * Based on Natural Breaks 1) % Pop. w/ High School or more >94% 2) % Pop. w/ High School or more 94% - 89% 3) % Pop. w/ High School or more 89% - 85% 4) % Pop. w/ High School or more <85%	ACS 2009-2013
Poverty	POVERTY	Fixed	County Level Poverty * Based on US Census poverty designations 1) % Families and Individuals at or below Poverty <10% 2) % Families and Individuals at or below Poverty 10% - 20% 3) % Families and Individuals at or below Poverty 20% - 25% 4) % Families and Individuals at or below Poverty >25%	ACS 2009-2013
AGEGP*EDU*POVERTY	n/a	Interaction	AGEGP*EDU*POVERTY	BRFSS 2011-2013
BRFSS Weight	CO_COMBOWT	Survey Weight	BRFSS County weighting variable	BRFSS 2011-2013

Table 1

What is it we want to answer?

Determine the extent to which socio-demographics can be used to predict BMI in Colorado. Our primary interest is in understanding overweight/obese rates by census tract and the influence of individual level socio-demographic characteristics and county level characteristics on the chance of being overweight/obese.

1) What are the odds of being overweight/obese for the average county in Colorado?

2) Does the percent overweight/obese vary across counties? How much of the variance in BMI (underweight/normal - overweight/obese) is attributable to individuals and to counties?

3) What is the relationship between individual socio-demographics and being overweight/obese?

4) Are there county level variables associated with an individual's likelihood of being overweight/obese?

5) Develop census tract level estimates of the percent of the population that is overweight and/or obese.

1) We calculate an estimate for the log odds of being overweight/obese in a typical county in Colorado at 0.2873 (odds=1.3328 and probability=0.5713).

2 & 3) Using the Covariance Parameter Estimate table, we can calculate the intraclass correlation coefficient (ICC) to determine how much of the total variation in the probability of being overweight/obese is accounted for by counties.

ICC=(0.2372/(0.2372+3.29))=0.0672 or 6.72%

The ICC indicates that 7% of the variability in overweight/obesity is accounted for by county (level-2) while 93% of the variability is accounted for by individuals (level-1). The 7% of variability between counties is a statistically significant amount of variability in the log odds of being overweight/obese between counties (est:0.2372; z=5.64, p<.0001).

*Generalized multilevel models assume no error at level-1, so in order to calculate ICC we assume the dichotomous outcome comes from an unknown continuous latent variable with a level-1 residual that follows a logistic distribution with a mean 0 and variance 3.29.

4) County Level Educational Attainment, County Level Poverty

5) See Map 01

Model Building

Individuals (Level 1) nested Counties (Level 2)

The following diagram generally outlines the process by which our multilevel model was specified. Through each progressive model, model fit was measured using a likelihood ratio test looking at -2LL values between models. AICc and BIC values are also assessed for a reduction in value between models

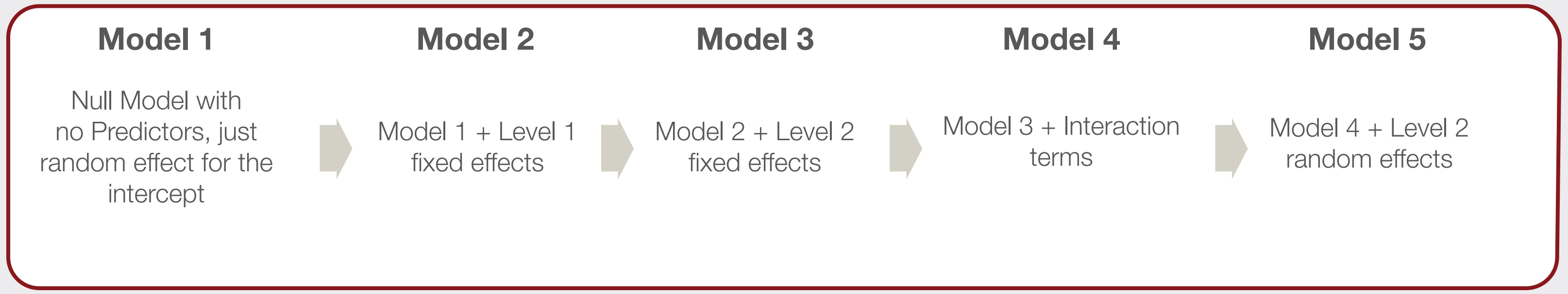
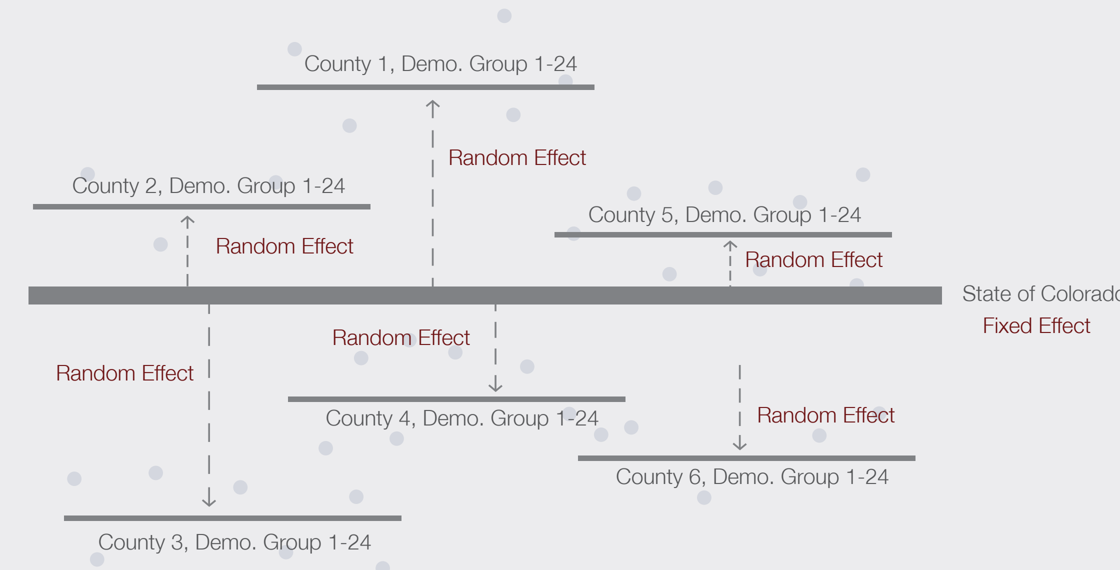


Figure 2



The following figure is a conceptualiztion of the relationships in a multilevel model *Figure 3*

Estimates from a 2-Level Generalized Linear Multilevel Model Predicating the Probability of being Overweight/Obese in Colorado (n=36,719)

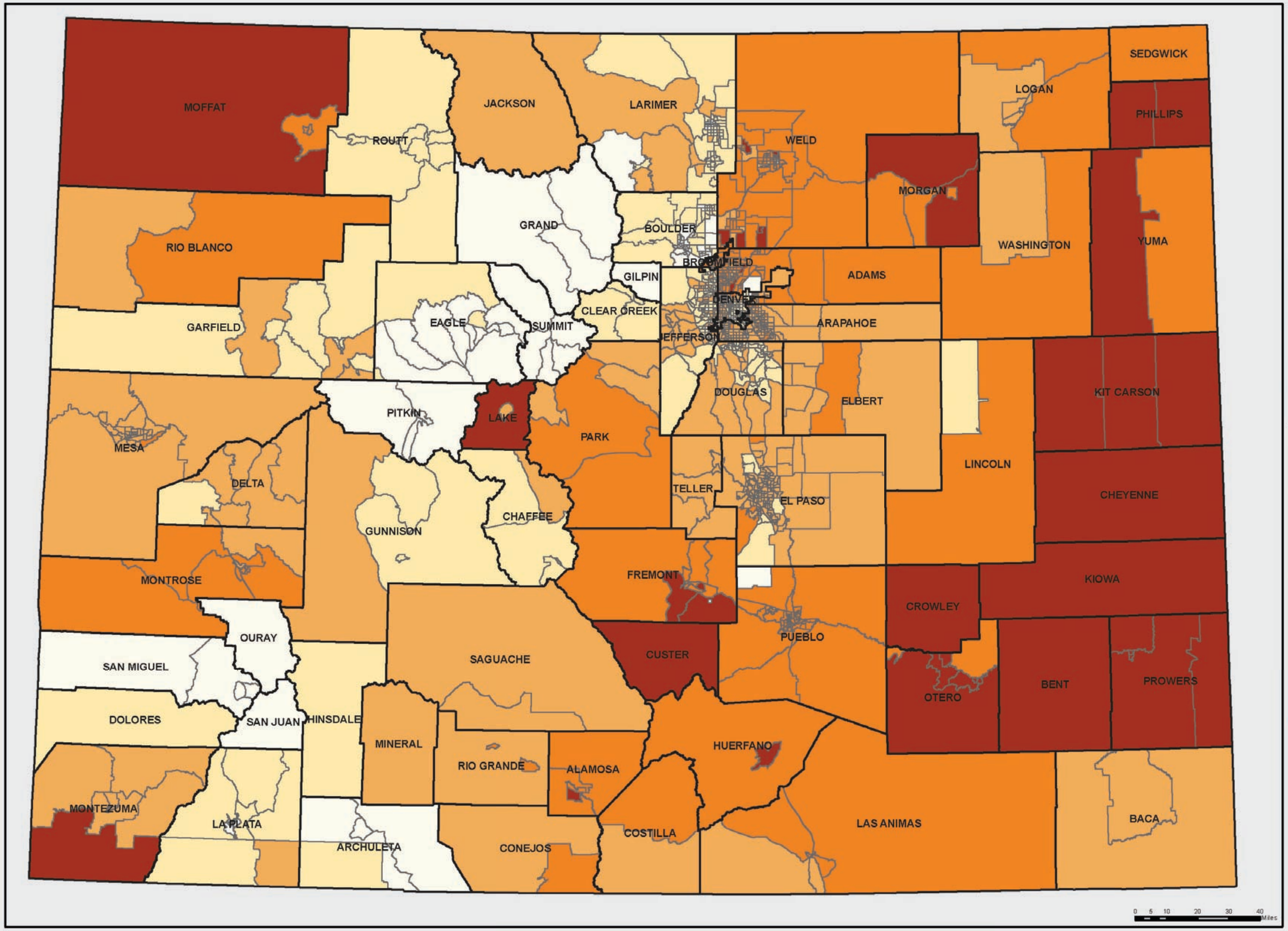
		Model 1	Model 2	Model 3	Model 4	Model 5
Fixed Effects						
Intercept*	Level 1	0.2873** (0.06)	0.2051** (0.06)	-0.4292** (0.001)	0.06891** (0.67)	-0.6550 (1.5186)
agegp*	Level 1		0.01144** (0.00)	0.01142** (0.00)	-0.00896** (0.00)	
edu*	Level 1			0.05862 (0.36)	0.1598** (0.00)	0.2133 (0.4113)
poverty*	Level 1			0.06567** (0.00)	-0.1529** (0.09)	-0.2971 (0.6857)
agegp * edu * poverty	Cross Level Interaction				0.004301** (0.00)	
Random Effects						
Intercept*		0.2372** (0.04)	0.2259** (0.04)	0.1498** (0.00)	0.1785** (0.000)	0.7016** (0.2931)
agegp						9.0637** (0.6030)
Model Fit						
-2LL		5,061,430	4,875,232	4,875,207	4,871,642	4,477,827
AIC		5,061,434	4,875,238	4,875,217	4,871,654	4,477,931
BIC		5,061,438	4,875,244	4,875,228	4,871,667	4,477,043

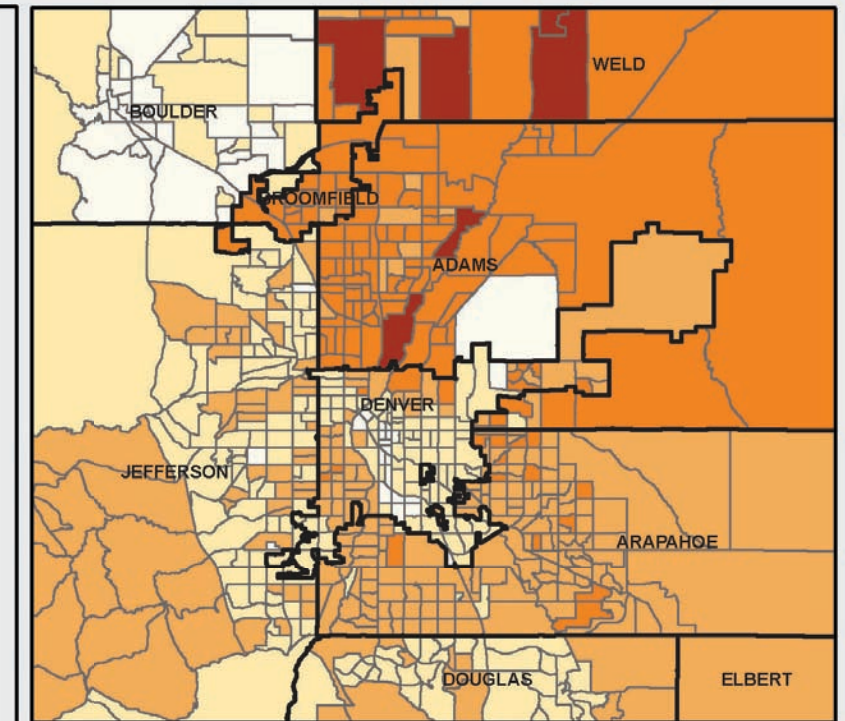
* logit, **p<0.05, ICC=0.07

We can assess model fit though a likelihood ratio test (chi-square difference test) comparing the difference in -2LL values between two nested models. We also look at AIC and BIC

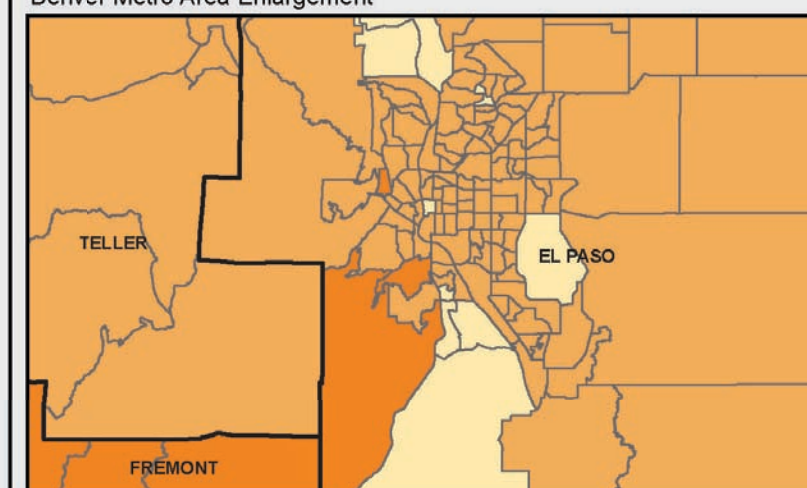
Estimation Method = Laplace

Table 2





Denver Metro Area Enlargement



El Paso County Enlargement

0 1.5 3 6 9 12 Miles

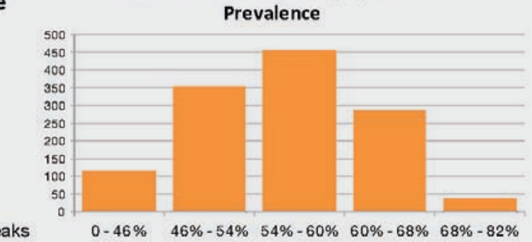
Legend

BMI >25.0: Small Area Estimate
***Percent Overweight/Obese**

- 0 - 46%
- 46% - 54%
- 54% - 60%
- 60% - 68%
- 68% - 82%

* Catagories based on natural breaks

Distribution of Overweight/Obese Prevalence



State of Colorado by Census Tract

Colorado Overweight and Obese by Census Tract: Percent of the Population Age 18+ with a BMI Greater than 25.0 (2011-2013)

Estimates are model based small area estimates based on BRFSS (2011-2013) and American Community Survery (2009-2013) data

Map 01

Colorado Department of Public Health and Enviornment
Center for Health and Environmental Data

Process steps using a multilevel regression model for small area BMI BRFSS estimates

COLORADO
Department of Public Health & Environment

Updated: June, 2015