Health & Place 30 (2014) 154–164

Contents lists available at ScienceDirect

Health & Place

journal homepage: www.elsevier.com/locate/healthplace

Does poor health predict moving, move quality, and desire to move?: A study examining neighborhood selection in US adolescents and adults

Erin C. Dunn [a,b,c], Ashley Winning [d], Natalya Zaika [e], S.V. Subramanian [d,*]

[a] Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA
[b] Department of Psychiatry, Massachusetts General Hospital/Harvard Medical School, Boston, MA, USA
[c] Stanley Center for Psychiatric Research, The Broad Institute of Harvard and MIT, Cambridge, MA, USA
[d] Department of Social and Behavioral Sciences, Harvard School of Public Health, 677 Huntington Avenue, Boston 02115, MA, USA
[e] Eliot-Pearson Department of Child Development, Tufts University, Medford, MA, USA

ARTICLE INFO

Article history:
Received 21 February 2014
Received in revised form
28 July 2014
Accepted 19 August 2014

Keywords:
Desire to move
Health
Move quality
Moving
Neighborhood selection

ABSTRACT

To date, research has rarely considered the role of health in shaping characteristics of the neighborhood, including mobility patterns. We explored whether individual health status shapes and constrains where individuals live. Using the National Longitudinal Study of Adolescent Health data, we examined whether 16 health indicators predicted moving, move quality, and desire to move. 3.8% of adolescents ($n=490$) reported a move in the past year. In the unadjusted models, 10 health indicators were associated with moving; the magnitude of association for these health indicators was similar to socio-demographic characteristics. 7 of these health-moving associations persisted after adjusting for covariates. Health was also associated with moving quality, with a greater number of past year health problems in the child being associated with moving to a lower income neighborhood and parent disability or poor health being associated with moving to a higher income neighborhood. Almost every poor health status indicator was associated with a greater desire to move. Findings suggest that health status influences moving, and a reciprocal framework is more appropriate for examining health-neighborhood linkages.

1. Introduction

For almost a century, scholars have examined neighborhoods as a determinant of individual well-being (March et al., 2008; Kawachi and Berkman, 2003; Faris and Dunham, 1939; Silver et al., 2002; Jencks and Mayer, 1990; Mayer and Jencks, 1989). To date, evidence has accumulated to suggest that multiple dimensions of the neighborhood are associated with mental and physical health outcomes across the life course, even after adjusting for individual-level attributes (Ahern and Galea, 2011; Galea et al., 2005, 2007; Mair et al., 2008; Theall et al., 2013; Leventhal and Brooks-Gunn, 2000). The hypotheses underpinning these “neighborhood effects” studies have been that economic (e.g., levels of poverty in the neighborhood), social (e.g., perceptions of safety, social control, social cohesion), and physical features (e.g., quality and maintenance of property) of neighborhoods contribute to patterns of health and illness (Diez Roux and Mair, 2010; Brenner et al., 2013; Browning and Cagney, 2003). These findings have been detected among studies using different study designs and multiple measures of the neighborhood environment. Although prior studies have generally found that the effect of neighborhoods on health is small, these studies do suggest that there may be identifiable and malleable predictors at the neighborhood level that could be intervened upon to shift the distribution of health problems in the population.

One of the most fundamental concerns with interpreting prior results from neighborhood effects research, particularly in non-experimental or observational studies, relates to selection effects, or the sorting of individuals by neighborhoods. Critics have argued that observed neighborhood-health associations may not be due to neighborhoods *causing* people to become ill, but rather can be explained by people *selecting* neighborhoods based on their health status, behaviors, or other structural factors. In other words, health status and health behaviors may shape the composition, or characteristics of individuals, in a given neighborhood. As a result, a major concern of prior studies is that reported neighborhood effects on health may not be real, but rather an artifact of the failure to statistically control for variables that cause individuals to select neighborhoods on the basis of their previous health status or conditions (Subramanian et al., 2007).

There is a rich history of work illustrating that a range of personal factors, such as race/ethnicity, socioeconomic status, and marital

* Corresponding author.
E-mail address: SVSUBRAM@hsph.harvard.edu (S.V. Subramanian).

http://dx.doi.org/10.1016/j.healthplace.2014.08.007
1353-8292/

status, as well as neighborhood or structural features, such as levels of discrimination, housing prices, and distribution of resources, have shaped and constrained where individuals live. For example, local access to high paying jobs (Wilson, 2011), housing quality (Epstein, 2003), housing supply (Woldoff and Ovadia, 2008), and policies related to housing discrimination (Teater, 2009) have all been shown to restrict residential possibilities and promote segregation particularly among minorities and low-income families. As a result, there are concentrations of people within specific communities who share similar features with respect to race and socioeconomic status (Ioannides and Zabel, 2008).

In this study, we explore whether individual health status is one of the characteristics that may further shape and constrain where individuals live. Beyond simply providing new knowledge to better understand potential neighborhood selection effects, or conceptualizing prior health status as a nuisance variable when studying neighborhood effects, we think examination of the role of health on neighborhood mobility is useful in its own right for several reasons. To date, very little research has considered the "health to neighborhood" relationship, or the role of health on shaping characteristics of the neighborhood, including mobility patterns. Instead, the majority of prior studies have focused solely on the role of neighborhoods on health. As the relationship between neighborhoods and health is likely bidirectional (Dunn et al., 2014), with neighborhoods shaping health and health status determining the composition and characteristics of neighborhoods, we think greater exploration of both directions of this bidirectional relationship would be informative for understanding the role of social determinants and health and identifying opportunities for intervention. For example, a greater understanding of the "health to neighborhood" relationship could increase knowledge of the consequences of health on mobility and may help identify potential intervention targets to optimize neighborhoods and reduce risk for illness.

Empirical studies examining how residents' health status influences their choice of neighborhood environment are rare and findings are sometimes mixed (van Lenthe et al., 2007; Verheij et al., 1998); however, extant studies suggest that well-being may play a role in choosing a neighborhood or in residential mobility. At least three different sources of prior studies are consistent with such a hypothesis. The first source of evidence comes from quasi-experimental and experimental studies of residential mobility, including the Gautreaux Program, which resulted from a court order to reduce discrimination in the location of new housing projects (Rosenbaum, 1995), and the Moving to Opportunity (MTO) experiment (Sampson, 2008), which was a large randomized control trial examining the impact of housing vouchers. In both studies, researchers found that certain individuals were more likely than others to be motivated to move. Regardless of whether they were assigned to the experimental or control group (Rosenbaum, 1995; Ludwig et al., 2008), the results of the MTO project showed that two-thirds of the control group moved between waves of data collection (Feins and Shroder, 2005). Thus, even in experimental studies, individuals *choose* their residence.

Second, observational studies examining aspects of *residential mobility* (e.g., frequency in change of residence) by health status have found more mobility among adults with depression (Sampson and Sharkey, 2008), and children with poor health (Busacker and Kasehagen, 2012), even after adjusting for sociodemographic factors. Overall, the literature describing the relationship between mobility and health has focused largely on mental health compared to other health problems (e.g., Jelleyman and Spencer, 2008).

Third, the strongest evidence investigating the relationship between health status and locale comes largely from studies of changes in geographical location on a large scale (e.g., moves to new states or countries) that examine associations between health and moving using cross-sectional methods (e.g., Piro et al., 2007). Most of the work on migrations has been conducted in areas of political, and massive social unrest that led to mass emigration (e.g., Erlanger, 2011). Studies of large scale migrations have found differential characteristics of migrants by health status when age is accounted for, with young migrants having better health than non-migrants and older migrants having poorer health than non-migrants (Bentham, 1988; Findlay, 1988; Halliday and Kimmitt, 2008; Larson et al., 2004). Movements of particular age groups in or out of a neighborhood contribute to the observed concentration of positive health status within affluent areas and the concentration of negative health status among deprived areas (Connolly et al., 2007). Thus, regardless of the direction of the health status, it appears that health is influencing the choice to move. Moreover, studies have also found that migrants are selective in their choice of destination, with health status influencing the decision of where to move (Larson et al., 2004). Migrants who move from lower to higher socioeconomic neighborhoods have also demonstrated better health (prior to moving) than those who move from higher to lower socioeconomic locations (Norman et al., 2005); this particularly refers to individuals with chronic health conditions (van Lenthe et al., 2007; Cox et al., 2007).

Outside of this body of research, knowledge is limited concerning whether people with specific health problems are more likely to live in certain residential environments, especially high poverty neighborhoods. Increasing knowledge in this area is crucial to understanding not just the magnitude of potential neighborhood selection and how to interpret prior and future "neighborhood effects" studies, but is also an important dimension in unpacking the complex relationship between health and neighborhoods. To address these gaps, this study used nationally-representative data to examine whether parent- or child-reported health problems were associated with moving to a new neighborhood. In the current study, we examined the independent contribution of health and socioeconomic factors to neighborhood choice, as defined by moving, move quality, and desire to move. We examined whether individuals with health problems were more likely than their healthy counterparts to report (a) moving to a new neighborhood, (b) moving to a neighborhood with greater levels of poverty, and (c) a desire to move from their current neighborhood.

2. Methods

2.1. Sample and procedures

Data came from the National Longitudinal Study of Adolescent Health (AddHealth), a United States nationally-representative longitudinal survey of adolescents (Harris, 2013). AddHealth recruited a school-based sample of adolescents in grades 7 through 12 and has followed respondents for a total of four waves into young adulthood. At Wave 1, 20,745 adolescents participated in a detailed in-home interview. In addition, 17,670 parental caregivers completed a Wave 1 survey. At Wave 2, which was completed approximately one year after Wave 1, 14,738 in-home Wave 1 respondents were interviewed.

2.2. Measures

2.2.1. Predictors: health status

Adolescent-reported measures. The Wave 1 survey included items on different dimensions of health status and health-related

Table 1
Descriptive statistics on health measures and covariates at wave 1 in the total sample (N=12,793) and by move status.

Child-reported items	**Move status**					
	Total sample (N=12,793)		**Non-mover (N=12,303)**		**Mover (N=490)**	
	N	**%**	***N***	**%**	***N***	**%**
General health status						
Good/very good/excellent	11,915	93.18	11,466	93.23	449	91.82
Fair/poor	872	6.82	832	6.77	40	8.18
Health-related absences from school						
None/just a few	12093	94.62	11648	94.77	445	90.82
Weekly to every day	687	5.38	643	5.23	44	8.98
Worst injury						
Minor	10,816	84.75	10413	84.84	403	82.41
Serious	1946	15.25	1860	15.16	86	17.59
Physical disability						
No	12468	97.54	11996	97.58	472	96.52
Yes	315	2.46	298	2.42	17	3.48
Alcohol use						
Never	7036	55.56	6804	55.85	232	48.13
A little	4562	36.02	4375	35.91	187	38.80
A lot	1066	8.42	1003	8.23	63	13.07
Marijuana use						
No	10,875	86.49	10,480	86.63	395	82.98
Yes	1699	13.51	1618	13.37	81	17.02
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Depressive symptoms	0.59	0.40	0.58	0.39	0.68	0.43
# of Health symptoms	3.14	2.96	3.12	2.95	3.67	3.17
BMI	22.34	4.42	22.34	4.43	22.3	4.17
Parent-reported items	***N***	**%**	***N***	**%**	***N***	**%**
# of Health problems in child						
0	6144	48.88	5917	48.89	227	48.61
1	4328	34.43	4183	34.56	145	31.05
2 or more	2098	16.69	2003	16.55	95	20.34
# of Health problems in parents						
0	2759	22.20	2669	22.29	90	19.69
1–2	6572	52.87	6326	52.84	246	53.83
3 or more	3099	24.93	2978	24.87	121	26.48
Parent disability						
No	11,954	93.86	11,515	94.01	439	90.14
Yes	782	6.14	734	5.99	48	9.86
Parent general health status						
Good/very good/excellent	10,881	85.32	10,504	85.64	377	77.25
Fair/poor	1872	14.68	1761	14.36	111	22.75
Parent alcohol use						
Never	5689	44.69	5455	44.55	234	48.25
A little	5375	42.22	5197	42.44	178	36.70
A lot	1666	13.09	1593	13.01	73	15.05
Parent smoking						
No	9076	71.27	8788	71.74	288	59.38
Yes	3658	28.73	3461	28.26	197	40.62
Child mental or physical disability						
No	10,455	83.08	10,089	83.26	366	78.21
Yes	2130	16.92	2028	16.74	102	21.79
Covariates	***N***	**%**	***N***	**%**	***N***	**%**
Sex						
Female	6553	51.22	6254	50.83	299	61.02
Male	6240	48.78	6049	49.17	191	38.98
Race						
White	6851	53.58	6614	53.79	237	48.37
Black	2532	19.80	2434	19.79	98	20.00
Hispanic	1940	15.17	1847	15.02	93	18.98
Other	1464	11.45	1402	11.40	62	12.65
Parent welfare receipt						
No	11,124	88.61	10,754	88.98	370	79.06

Table 1 (*continued*)

Child-reported items	Move status					
	Total sample (N=12,793)		Non-mover (N=12,303)		Mover (N=490)	
	N	%	*N*	%	*N*	%
Yes	1430	11.39	1332	11.02	98	20.94
Parent highest education						
College or more	4462	36.67	4351	37.12	111	24.83
Less than college	7707	63.33	7371	62.88	336	75.17
Parent highest occupation						
Professional/manager	4260	36.16	4143	36.43	117	28.75
Technical/office worker/sales	3119	26.48	3027	26.62	92	22.60
Service/construction/military	4401	37.36	4203	36.96	198	48.65
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Child age (in years)	15.74	1.58	15.73	1.57	15.91	1.66

Cell entries are numbers and percentages for categorical variables (or means and standard deviations for continuous variables). Movers refer to participants who reported having moved to a new census tract between Wave 1 and Wave 2. Columns may not add to 100% due to rounding. Columns may not add to the total sample size due to missing values.

behaviors. *General health status* was ascertained by the item: in general, how is your health (0=good/very good/excellent health; 1=fair/poor health)? Frequency of *health related absences from school* were derived from two questions: in the past month, how often did a health or emotional problem cause you to miss: (1) a day of school, or (2) a social or recreational activity (0=none/just a few; 1=about once a week to every day)? *Injury-related health* was captured by the degree of seriousness of the worst injury during the past year (0=very minor or minor; 1=serious to extremely serious). *Physical disability* was captured with the item: do you have difficulty using your hands, arms, legs, or feet because of a permanent physical condition (0=no; 1=yes)? Frequency of *alcohol use* was ascertained by the item: during the past 12 months, on how many days did you drink alcohol (0=never; 1=a little, defined as 1–2 days a year to 2–3 days per month; 2=a lot, defined as 1–2 days a week to every day)? *Drug use* was ascertained by the item: during the past 30 days, how many times did you use marijuana (0=never; 1=at least one time)? *Depressive symptoms* (coded as a scale) were assessed by the 19-item version of the Center for Epidemiologic Studies of Depression Scale (CES-D). *Number of health symptoms* (coded as a count of the conditions experienced once a week or more) was derived from questions about the frequency of experiencing 20 health symptoms over the past 12 months (e.g., headaches, stomachaches, chest pains, dizziness, poor appetite). *BMI* (coded continuously) was calculated by dividing weight (in pounds) by height-squared (in inches).

Parent-reported measures. The Wave 1 parent survey included a collection of measures on different dimensions of health status and health related behaviors of both the adolescent and parent reporter (usually the mother or female head of household). *Total number of health problems of the child* was assessed by questions about the presence of six current health problems: obesity, migraines, allergies, asthma, alcoholism, and diabetes (0=0 health problems; 1=1 health problem; 2=2 or more health problems). Health problems of the biological mother and father were assessed by the same six items and summed to create a variable capturing *total number of health problems of the parents* (0=0 health problems; 1=1–2 health problem; 2=3 or more health problems). *Parent disability* was ascertained by the item: are you disabled (0=no; 1=yes)? *Parent general health status* was ascertained by the item: how is your general physical health (0=good/very good/excellent health; 1=fair/poor health)? *Parental alcohol* use derived from the item: how often do you drink alcohol (0=never; 1=a little, defined as 1–2 days a year to 2–3 days per month; 2=a lot, defined as 1–2 days a week to every day)? *Parental smoking* status was determined by the item: do you smoke (0=no; 1=yes)? *Child physical or mental disability* was assessed by two questions about physical disability (concerning difficulty using hands or arms and feet or legs) and by three questions on mental disability (concerning mental retardation, learning disabilities, and special education) (0=no disability; 1=any disability).

2.2.2. *Outcomes: move status, move quality, and desire to move*

We created three measures to describe participants' move status, move quality, and desire to move. First, we constructed a binary indicator denoting whether or not the adolescent had moved Census tracts between Waves 1 and 2 (0=did not move; 1=moved). Second, among those who moved, we created a continuous and a binary measure of *move quality*. Specifically, for each mover, we used Census-level data to determine the percentage of residents in the tract who lived in poverty. Census tract poverty estimates were derived for both the past neighborhood, from which the participant moved from, and the new neighborhood, to which the participant moved. With these measures, we created a continuous score of move quality by subtracting Wave 1 neighborhood poverty from Wave 2 poverty. Thus, a negative beta coefficient can be interpreted as a move to a lower poverty neighborhood and a positive beta coefficient can be interpreted as a move to a higher poverty neighborhood. We also constructed a dichotomous measure of move quality based on whether the new census tract had a lower or a higher percentage of poverty compared to the old one. Finally, *desire to move* was derived from the parent-reported item: how much would you like to move away from this neighborhood (0=no desire to move and 1=some or very much desire)?

2.2.3. *Covariates*

All adjusted models contained controls for *sex* (0=female and 1=male), *age* (continuous), adolescent self-reported race/ethnicity (0=white; 1=Black; 2=Hispanic; and 3=other) and, as measures of *socioeconomic status*, parental education (highest level attained by either mother or father; 0=college degree or higher; and 1=less than college), parental occupation (highest level attained by either mother or father; 0=professional or manager; 1=technical, office, or sales; 2=service, construction, factory, transportation, security, military, etc.), and parental receipt of public assistance, such as welfare (0=no and 1=yes). All covariates were taken from Wave 1.

Table 2
Unadjusted and adjusted logistic regression analysis modeling association between health measures and move status (mover vs. non-mover) in the total sample ($N=12{,}793$).

Child-reported items	Model 1			Model 2		
	OR	95% CI	*p*-Value	OR	95% CI	*p*-Value
General health status						
Good/very good/excellent	ref	ref		ref	ref	
Fair/poor	1.23	0.91–1.65	0.17	1.20	0.85–1.70	0.29
Health-related absences from school						
None/just a few	ref	ref		ref	ref	
Weekly to every day	1.79	1.39–2.31	< 0.0001	1.85	1.37–2.50	< 0.0001
Worst injury						
Minor	ref	ref		ref	ref	
Serious	1.20	0.96–1.49	0.12	1.24	0.95–1.63	0.12
Physical disability						
No	ref	ref		ref	ref	
Yes	1.45	0.85–2.47	0.17	1.25	0.69–2.25	0.46
Alcohol use						
Never	ref	ref		ref	ref	
A little	1.25	0.98–1.60	0.07	1.23	0.94–1.60	0.13
A lot	1.84	1.31–2.60	0.0005	1.61	1.06–2.43	0.02
Marijuana use						
No	ref	ref		ref	ref	
Yes	1.33	1.01–1.75	0.04	1.23	0.91–1.66	0.18
Depressive symptoms	1.73	1.35–2.22	< 0.0001	1.58	1.20–2.07	0.001
# of Health symptoms	1.06	1.02–1.10	0.001	1.06	1.02–1.10	0.002
BMI	1.00	0.98–1.02	0.86	0.99	0.97–1.02	0.48
Parent-reported items						
# of Health problems in child						
0	ref	ref		ref	ref	
1	0.90	0.72–1.14	0.39	0.96	0.75–1.23	0.73
2 or more	1.24	0.97–1.58	0.09	1.11	0.83–1.49	0.48
# of Health problems in parents						
0	ref	ref		ref	ref	
1–2	1.15	0.90–1.49	0.27	1.17	0.87–1.57	0.31
3 or more	1.42	0.93–1.56	0.16	1.08	0.78–1.49	0.65
Parent disability						
No	ref	ref		ref	ref	
Yes	1.72	1.24–2.37	0.001	1.46	0.93–2.30	0.10
Parent general health status						
Good/very good/excellent	ref	ref		ref	ref	
Fair/poor	1.76	1.38–2.23	< 0.0001	1.40	1.05–1.85	0.02
Parent alcohol use						
Never	ref	ref		ref	ref	
A little	0.80	0.65–0.98	0.03	0.91	0.71–1.17	0.47
A lot	1.07	0.81–1.42	0.65	1.30	0.93–1.83	0.13
Parent smoking						
No	ref	ref		ref	ref	
Yes	1.74	1.41–2.14	< 0.0001	1.62	1.27–2.06	0.0001
Child mental or physical disability						
No	ref	ref		ref	ref	
Yes	1.39	1.09–1.77	0.008	1.41	1.06–1.90	0.02
Covariates						
Sex						
Female	ref	ref		ref	ref	
Male	0.66	0.54–0.81	< 0.0001	0.70	0.56–0.87	0.001
Race						
White	ref	ref		ref	ref	
Black	1.12	0.80–1.57	0.50	1.00	0.68–1.39	0.89
Hispanic	1.41	0.92–2.14	0.11	1.22	0.75–1.99	0.42
Other	1.23	0.85–1.78	0.26	1.15	0.76–1.73	0.52
Parent welfare receipt						
No	ref	ref		ref	ref	
Yes	2.14	1.58–2.90	< 0.0001	1.76	1.20–2.58	0.004
Parent highest education						
College or more	ref	ref		ref	ref	
Less than College	1.78	1.38–2.32	< 0.0001	1.44	1.08–1.91	0.01

Table 2 (*continued*)

Child-reported items	Model 1			Model 2		
	OR	95% CI	*p*-Value	OR	95% CI	*p*-Value
Parent highest occupation						
Professional/manager	ref	ref		ref	ref	
Technical/office worker/sales	1.08	0.80–1.44	0.62	0.90	0.67–1.22	0.49
Service/construction/military	1.67	1.29–2.16	< 0.0001	1.22	0.92–1.63	0.17
Child age (in years)	1.07	0.98–1.17	0.11	1.04	0.95–1.14	0.38

Cell entries are odds ratios (exponentiated beta coefficients), 95% confidence intervals, and p-values. Model 1 examined the unadjusted association between child- or parent- reported health status measures on move status (movers vs. non-movers). Model 2 examined the effect of child- or parent- reported predictors on move status, adjusting for the following covariates: child sex, race, and age; and parent education, occupation, and welfare receipt. In both models, each predictor variable was entered into the model individually; thus, no predictors were entered simultaneously.

2.2.4. *Statistical analysis*

We fit a series of multiple logistic regression models examining the association between health status with move status (mover vs. not a mover), quality of move (moved to a census tract of lower poverty vs. moved to census tract with higher poverty), and desire to move (some or very much desire vs. no desire). We also used a set of linear regression models to examine the association between health status and our continuous measure of move quality (i.e., change in level of neighborhood poverty). Across all models, each health indicator was examined individually to avoid problems with model fit due to multi-collinearity; thus, no predictors were entered simultaneously. All analyses were conducted in SAS version 9.2 using an analytic sample of 12,793 participants, who completed surveys in Waves 1 and 2, whose parents/caregivers completed a survey in Wave 1, and who had complete data on all relevant study variables. Missing on health indicators and covariates was minimal, ranging from 0% (for sex) to 7.9% (for parent highest occupation). Apart from parent occupation and education, all variables had less than 3% missing data.

3. Results

Respondents were generally in good health, though a large proportion reported having some health problems in the past year or engaging in risky behaviors. As shown in Table 1, which presents descriptive statistics for all predictors and covariates in the total sample and by move status, adolescents in the sample self-reported being in good health (93.18% reported good, very good or excellent health). However, 16.69% had more than one health problem, 15.25% had a serious injury, 8.42% drank alcohol a lot, 13.51% used marijuana in the past month, and 16.92% had a mental or physical disability. Similarly, the majority of parents (85.32%) reported good, very good, or excellent health, though, 24.93% reported 3 or more health problems (in either parent), 13.09% drank alcohol a lot, and 28.73% were smokers.

Moving was a rare occurrence in AddHealth, with 3.8% of respondents ($n=490$) reporting a move in the approximate one year period between Wave 1 and Wave 2. In testing the association between each of the 16 health status measures and the odds of moving, we found slightly more than half (10 health indicators) were associated with move status (Model 1, Table 2). For many, the magnitude of the association between health status and moving was comparable to the association between socio-demographic characteristics and moving. For example, the odds of moving was 1.84 among youth who used alcohol a lot, compared to 1.78 for parents who had less than a college degree and 2.14 for parental receipt of public assistance. After adjusting for covariates, seven out of 10 of these associations remained significant (Model 2, Table 2). Specifically, children who reported more health related absences from school (OR=1.85; $p<0.001$) and frequent alcohol use (OR=1.61; $p<0.02$) were more likely to move. A significant and positive association was also found between depressive symptoms (OR=1.58; $p=0.001$) and number of past year poor health symptoms and move status (OR=1.06; $p=0.002$). With respect to parent-level measures, parents who reported fair or poor health were more likely than those with good, very good, or excellent health to report moving (OR=1.40; $p=0.02$). In addition, parents who smoked (OR=1.62; $p<0.0001$) or reported that their child had a physical or mental disability were also more likely to move (OR=1.41; $p=0.02$). Even after imposing a Bonferroni adjustment for the number of health-specific tests conducted (adjusted p-value=0.003), we found that four (out of 16) health indicators (i.e., child health related absences, depressive symptoms, number of past year health symptoms, and parent smoking) remained associated with moving, with effect sizes ranging in magnitude from small (OR=1.06) to modest (OR=1.85). The magnitude of the association between health related absences and moving was greater than the association between all demographic and socioeconomic characteristics and moving, including receipt of public assistance (OR=1.76) and low educational attainment (OR=1.44).

To evaluate further the magnitude of these differences, we calculated predicted probabilities of moving using beta estimates from these logistic regression models. In these analyses, which focused on the referent group (White females of average age and high socioeconomic status), the probability of moving was nearly double among those with many child health related absences (4.7%) compared to those with few absences (2.6%) and almost double among those with a parent who smoked (3.9% vs. 2.5% among non-smoking parents). Moreover, the absolute difference in the probability of moving was 5% when comparing those with 0 past year health symptoms to those with 10 symptoms or those with low versus high depressive symptoms.

For those who moved ($N=490$), the level of poverty in their old neighborhood compared to their new neighborhood was typically similar. For example, on average, the difference in poverty between the old versus new neighborhood was less than 1% (mean=0.09%; sd=12.41%). However, for some people there was a considerable difference between their old and new neighborhood; (minimum absolute difference=0.06%; maximum absolute difference=48.70%). More than half of the mover sample (55.10%; $n=270$) reported a greater than 5% change in poverty (in either direction; i.e., higher or lower) between their old neighborhood and new neighborhood; 33.88% ($n=166$) had between or equal to a 1% and 5% change; and the remaining 11.02% ($n=54$) had less than a 1% change in neighborhood poverty level.

Table 3
Adjusted linear regression analysis modeling association between health measures and change in percent poverty among movers' former and new neighborhoods ($n = 490$).

Child-Reported Items	**Model 1**			**Model 2**		
	Beta	**SE**	***p*-Value**	**Beta**	**SE**	***p*-Value**
General health status						
Good/very good/excellent	ref	ref		ref	ref	
Fair/poor	0.026	0.020	0.21	0.035	0.022	0.11
Health-related absences from school						
None/just a few	ref	ref		ref	ref	
Weekly to every day	0.014	0.020	0.49	0.027	0.021	0.19
Worst injury						
Minor	ref	ref		ref	ref	
Serious	−0.012	0.015	0.43	−0.021	0.016	0.21
Physical disability						
No	ref	ref		ref	ref	
Yes	0.052	0.031	0.087	0.045	0.035	0.21
Alcohol use						
Never	ref	ref		ref	ref	
A little	0.002	0.012	0.85	0.006	0.014	0.66
A lot	0.003	0.018	0.87	−0.01	0.021	0.50
Marijuana use						
No	ref	ref		ref	ref	
Yes	0.005	0.015	0.72	0.007	0.017	0.69
Depressive symptoms	0.016	0.013	0.22	0.015	0.014	0.31
# of Health symptoms	0.004	0.002	0.019	0.005	0.002	0.007
BMI	−0.001	0.001	0.65	−0.001	0.001	0.57
Parent-reported items						
# of Health problems in Child						
0	ref	ref		ref	ref	
1	0.016	0.013	0.22	0.010	0.014	0.49
2 or more	−0.003	0.012	0.85	0.000	0.017	0.98
# of Health problems in parents						
0	ref	ref		ref	ref	
1–2	0.022	0.015	0.16	0.012	0.016	0.45
3 or more	0.006	0.017	0.75	0.002	0.019	0.93
Parent disability						
No	ref	ref		ref	ref	
Yes	−0.030	0.019	0.11	−0.058	0.024	0.01
Parent general health status						
Good/very good/excellent	ref	ref		ref	ref	
Fair/poor	−0.057	0.013	< 0.0001	−0.053	0.016	0.001
Parent alcohol use						
Never	ref	ref		ref	ref	
A little	−0.006	0.012	0.64	−0.010	0.014	0.45
A lot	0.007	0.017	0.67	0.002	0.018	0.93
Parent smoking						
No	ref	ref		ref	ref	
Yes	0.016	0.011	0.17	0.025	0.013	0.06
Child mental or physical disability						
No	ref	ref		ref	ref	
Yes	0.013	0.014	0.36	0.010	0.016	0.53
Covariates						
Sex						
Female	ref	ref		ref	ref	
Male	−0.017	0.011	0.14	−0.016	0.012	0.21
Race						
White	ref	ref		ref	ref	
Black	−0.008	0.015	0.60	−0.036	0.017	0.03
Hispanic	0.009	0.015	0.56	0.007	0.017	0.69
Other	0.010	0.018	0.58	−0.016	0.020	0.40
Parent welfare receipt						
No	ref	ref		ref	ref	
Yes	−0.014	0.014	0.31	−0.031	0.017	0.08
Parent highest education						
College or more	ref	ref		ref	ref	
Less than college	−0.016	0.014	0.23	−0.015	0.015	0.31

Table 3 (*continued*)

Child-Reported Items	Model 1			Model 2		
	Beta	SE	*p*-Value	Beta	SE	*p*-Value
Parent highest occupation						
Professional/manager	ref	ref		ref	ref	
Technical/office worker/sales	−0.017	0.017	0.30	−0.020	0.015	0.20
Service/construction/military	−0.011	0.014	0.43	−0.030	0.018	0.09
Child Age (in years)	−0.004	0.003	0.25	−0.002	0.004	0.54

Cell entries are parameter estimates (beta coefficients), standard errors, and *p*-values. Model 1 examined the unadjusted effect of child- or parent- reported health status measures on change in percent poverty among movers' former and new neighborhood ($n=490$). Model 2 examined the association between child- or parent- reported health status measures on change in percent poverty among movers' former and new neighborhood, adjusting for covariates. The continuous quality of move score was created by subtracting Wave 1 neighborhood poverty from Wave 2 poverty. Thus, a negative beta coefficient can be interpreted as a move to lower poverty neighborhood and positive beta coefficient can be interpreted as a move to a higher poverty neighborhood. Each predictor variable was entered into the model individually; thus, no predictors were entered simultaneously.

In examining the relationship between health status and quality of move measured continuously (refer to Table 3), we found that a greater number of past year health symptoms was positively associated with moving to a lower income neighborhood ($\beta=0.005$; $p=0.007$), adjusting for covariates. However, parental disability ($\beta=-0.058$; $p=0.01$) or fair/poor health ($\beta=-0.053$; $p=0.001$) was *negatively* associated with moving to a lower income neighborhood (i.e., having a parent with a disability was associated with moving to a *higher* income neighborhood).

When we examined move quality using the dichotomous measure, we found that 49% of movers reported transitioning into a census tract with a lower percentage of poverty than the previous tract. After adjusting for covariates (see Model 2 in Supplemental Table S1), only child reports of serious injury (OR=0.48; $p=0.02$) and parent fair/poor health status (OR=0.44; $p=0.02$) were associated with a lower odds of moving to a neighborhood with higher poverty.

Approximately 46% of parents ($n=5906$) reported some or very much desire to move. As shown in Table 4, almost every poor health status indicator was associated with a greater desire to move, even after adjusting for covariates. For example, compared to people with good, very good, or excellent health, those with fair or poor health had 1.23 times the odds of having some or very much desire to move. Desire to move in Wave 1 was associated with actually moving by Wave 2, with those reporting some or very much desire to move having 1.98 times the odds of moving compared to those reporting no desire to move, even after adjusting for covariates (95% CI=1.58–2.48; $p<0.001$). However, desire to move was not associated with quality of move.

4. Discussion

The current study examined the independent contribution of health and socioeconomic factors to neighborhood choice, as defined by moving, move quality, and desire to move. In doing so, we aimed to complement extant studies, which have largely focused on the relationship between neighborhoods and health, in order to increase knowledge of the bidirectional relationship between health and neighborhoods. A secondary goal was also to examine the sorting of individuals by geography, which has been described mostly in theoretical terms and has been examined empirically in a small number of studies (Subramanian et al., 2007), most of which have examined non-neighborhood based moves (e.g., moves between countries). We found that there does appear to be some evidence of neighborhood selection, with the health status indicators we examined oftentimes having stronger associations with move indicators than other demographic indicators, including socioeconomic status. For example, after adjusting for covariates, the odds of moving was 1.44 among youth whose parents had less than a college degree, compared to 1.85 for youth who reported more health related absences from school. Our results also suggest differences between parents and children in the quality of move by health status. For example, we found that a greater number of past year health problems in the child was associated with moving to a *lower* income neighborhood. However, parents with disabilities or fair or poor health were more likely to move to a *higher* income neighborhood. Thus, the influence of poor health on move quality was not uniform across all health conditions. Finally, we also found that almost every health status indicator was associated with desire to move, even after adjusting for covariates. Collectively, these findings suggest that neighborhood moving is influenced by health status and that prior health status may be an issue we must consider when interpreting results from neighborhood effects studies.

It is unclear why health status was not consistently associated with move quality in particular. As noted previously, we found that health problems among the children were more likely to predict moving to a *lower* income neighborhood, whereas health problems among parents were more likely to predict moving to a *higher* income neighborhood. Moreover, when comparing parents to children on the same health condition, we found several instances where the health condition was associated with move quality in adults, but not children, and vice versa. For example, child physical disability was unrelated to move quality, but physical disability in the parent was. Although there are a number of possible explanations for why health status could have a different relationship to moving indicators among parents and children, such as differential access to health services or understanding and reporting of health status, the source of conflicting findings is unclear. Given the dearth of prior empirical work on this topic, future studies are needed.

These results must be interpreted in light of the fact that moving was a relatively rare occurrence in our sample, as only 3.8% of adolescents reported a move in the past year. The low occurrence of moving was likely due to our examination of moves only during a roughly one-year period. Given the association between desire to move and actually moving, it is possible that those with some or very much desire may have been more likely than those with no desire to have moved in the subsequent years. Although a longer longitudinal study would have been more ideal to understand the relationship between health status and moving, our examination of the association between a broad array of health indicators and moving is useful because the outcomes

Table 4
Unadjusted and adjusted logistic regression analysis modeling association between health measures and parent's desire to move (some/very much desire vs. no desire) in the total sample ($N = 12,793$).

Child-reported items	Model 1			Model 2		
	OR	95% CI	*p*-Value	OR	95% CI	*p*-Value
General health status						
Good/very good/excellent	ref	ref		ref	ref	
Fair/poor	1.32	1.16–1.51	< 0.0001	1.23	1.06–1.42	0.006
Health-related absences from school						
None/just a few	ref	ref		ref	ref	
Weekly to every day	1.26	1.06–1.49	0.008	1.26	1.04–1.52	0.02
Worst injury						
Minor	ref	ref		ref	ref	
Serious	1.16	1.05–1.29	0.005	1.17	1.04–1.31	0.008
Physical disability						
No	ref	ref		ref	ref	
Yes	1.15	0.91–1.45	0.25	1.21	0.93–1.56	0.16
Alcohol use						
Never	ref	ref		ref	ref	
A little	1.02	0.94–1.10	0.62	1.09	1.01–1.18	0.03
A lot	1.11	0.97–1.27	0.14	1.17	1.01–1.36	0.04
Marijuana use						
No	ref	ref		ref	ref	
Yes	1.25	1.11–1.40	0.0002	1.26	1.12–1.43	0.0002
Depressive symptoms	1.38	1.24–1.54	< 0.0001	1.35	1.12–1.51	< 0.0001
# of Health symptoms	1.02	1.01–1.03	0.001	1.02	1.01–1.04	0.0002
BMI	1.01	1.00–1.01	0.20	1.00	1.00–1.01	0.72
Parent-reported items						
# of Health problems in child						
0	ref	ref		ref	ref	
1	1.12	1.03–1.19	< 0.0001	1.14	1.06–1.23	0.0008
2 or more	1.29	1.16–1.43	0.006	1.31	0.89–1.46	< 0.0001
# of Health problems in parents						
0	ref	ref		ref	ref	
1–2	1.11	1.01–1.24	0.04	1.17	1.04–1.31	0.007
3 or more	1.47	1.29–1.67	< 0.0001	1.51	1.31–1.73	< 0.0001
Parent disability						
No	ref	ref		ref	ref	
Yes	1.20	0.99–1.45	0.07	1.03	0.81–1.32	0.79
Parent general health status						
Good/very good/excellent	ref	ref		ref	ref	
Fair/poor	1.39	1.22–1.59	< 0.0001	1.26	1.12–1.43	0.0002
Parent alcohol use						
Never	ref	ref		ref	ref	
A little	1.16	1.05–1.29	0.005	1.25	1.13–1.39	< 0.0001
A lot	1.10	0.95–1.29	0.21	1.19	1.03–1.39	0.018
Parent smoking						
No	ref	ref		ref	ref	
Yes	1.48	1.34–1.64	< 0.0001	1.37	1.24–1.51	< 0.0001
Child mental or physical disability						
No	ref	ref		ref	ref	
Yes	1.37	1.24–1.51	< 0.0001	1.35	1.21–1.51	< 0.0001
Covariates						
Sex						
Female	ref	ref		ref	ref	
Male	1.00	0.93–1.07	0.97	1.01	0.94–1.08	0.84
Race						
White	ref	ref		ref	ref	
Black	1.62	1.36–1.94	< 0.0001	1.71	1.45–2.02	< 0.0001
Hispanic	1.14	0.97–1.35	0.10	1.09	0.94–1.26	0.27
Other	1.10	0.94–1.28	0.24	1.03	0.88–1.20	0.73
Parent welfare receipt						
No	ref	ref		ref	ref	
Yes	1.62	1.33–1.96	< 0.0001	1.45	1.21–1.74	< 0.0001
Parent highest education						
College or more	ref	ref		ref	ref	
Less than college	1.29	1.15–1.46	< 0.0001	1.21	1.07–1.37	0.002

Table 4 (*continued*)

Child-reported items	Model 1			Model 2		
	OR	95% CI	*p*-Value	OR	95% CI	*p*-Value
Parent highest occupation						
Professional/manager	ref	ref		ref	ref	
Technical/office worker/sales	1.26	1.13–1.42	< 0.0001	1.17	1.05–1.31	0.006
Service/construction/military	1.25	1.09–1.43	0.001	1.07	0.95–1.2	0.29
Child age (in years)	1.00	0.96–1.04	0.97	1.00	0.97–1.04	0.89

Cell entries are odds ratios (exponentiated beta coefficients), 95% confidence intervals, and p-values. Model 1 examined the unadjusted association between child- or parent-reported health status measures on parent's desire to move (some or very much desire to move vs. no desire to move). Model 2 examined the association between child- or parent- reported health status measures on parent's desire to move, adjusting for covariates. In both models, each predictor variable was entered into the model individually; thus, no predictors were entered simultaneously.

likely occurred close in time to the health measures we examined. However, future studies should examine how incident health problems and specific health conditions predict future moves and examine neighborhood changes over a longer period of time.

Results from our study suggest that examination of health conditions and health behaviors may play a pivotal role in how people "choose" a neighborhood. However, much more research is needed to further understand the ways in which people may choose neighborhoods based on their health status, behaviors, or other structural factors. We envision several different possible areas for future research. First, future studies can examine the meaning residents attribute to a particular residential location and the duration of their intended move. This is an important line of future inquiry, as residence in a particular neighborhood may not be permanent, but rather temporary until particular circumstances improve (Piro et al., 2007). Second, future studies can also more richly take into consideration changes in socioeconomic status, including family income, immigration status, race/ethnicity, and single parenthood over time. One of the main covariates that our study took into consideration was educational attainment, which has been consistently related to health (van Lenthe et al., 2007). However, other dimensions of socioeconomic status, such as income, may fluctuate over time and have stronger relationships with moving based on some health conditions. Third, future studies should also examine how different measurement approaches and techniques can capture health in relation to moving. In our study, certain health measures were retrieved through self-report; however, as was reported in previous studies (Mackenbach et al., 1996), using other assessments or medical diagnoses in tandem may be necessary to capture valid representation of certain health conditions. Finally, longitudinal studies are also needed to disentangle the complex relationship between neighborhoods, health, and residential mobility. By having repeated measures of neighborhood characteristics and health over time and being able to observe the predictors and consequences of change in neighborhood residence over time, longitudinal studies will provide crucial new knowledge to both quantify and separate out potential selection effects. To that end, greater use of propensity score approaches, which epidemiologists have not yet widely adopted in neighborhood effects studies but allow investigators to address possible unmeasured confounding (Joffe and Rosenbaum, 1999), may also help to better disentangle selection effects, even in the context of an experimental study.

If health does influence moving, as our study suggests, how can we interpret prior and future studies describing a relationship between neighborhoods and health? Findings from the current study suggest that caution is warranted when applying a causal interpretation to both observational and experimental neighborhood effects studies. Results from this study suggest that because health status is associated with move indicators, empirical studies may over or underestimate the relationship between neighborhoods and health depending on whether they take into account prior health status. In some cases, estimates of neighborhood effects on health may be over- or underestimated based on selection or confounding as individuals non-randomly "choose" to live in different neighborhoods. To further advance the science of neighborhood effects and health, additional studies are needed to quantify the degree of potential selection resulting from health-based moving and how prior health status may play a role in move and desire to move.

Acknowledgment

SVS is supported by a Robert Wood Johnson Investigator Award in Health Policy Research.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.healthplace.2014.08.007.

References

Ahern, J., Galea, S., 2011. Collective efficacy and major depression in urban neighborhoods. Am. J. Epidemiol. 173 (12), 1453–1462.
Bentham, G., 1988. Migration and morbidity: implications for geographical studies of disease. Soc. Sci. Med. 26, 49–54.
Brenner, A., Zimmerman, M., Bauermeister, J., Caldwell, C., 2013. Neighborhood context and perceptions of stress over time: an ecological model of neighborhood stressors and intrapersonal and interpersonal resources. Am. J. Community Psychol. 51 (3–4), 544–556.
Browning, C., Cagney, K., 2003. Moving beyond poverty: neighborhood structure, social processes, and health. J. Health Soc. Behav. 44 (4), 552–571.
Busacker, A., Kasehagen, L., 2012. Association of residential mobility with child health: an analysis of the 2007 National Survey of Children's Health. Matern. Child Health J. 16, S78–S87.
Connolly, S., O'Reilly, D., Rosato, M., 2007. Increasing inequalities in health: is it an artefact caused by the selective movement of people? Soc. Sci. Med. 64 (10), 2008–2015.
Cox, M., Boyle, P.J., Davey, P., Morris, A., 2007. Does health-selective migration following diagnosis strengthen the relationship between Type 2 diabetes and deprivation? Soc. Sci. Med. 65 (1), 32–42.
Diez Roux, A., Mair, C., 2010. Neighborhoods and health. Ann. N. Y. Acad. Sci. 1186, 125-145.
Dunn, E.C., Masyn, K.E., Yudron, M., Jones, S.M., Subramanian, S.V., 2014. Translating multilevel theory into multilevel research: challenge and opportunities for understanding the social determinants of psychiatric disorders. Soc. Psychiatry Psychiatr. Epidemiol. 49, 859–872.
Epstein, H., 2003. Enough to Make You Sick? The New York Times.
Erlanger, S., 2011. Attack Reignites Immigration Debate in Divided Oslo. The New York Times.
Faris, R.E.L., Dunham, H.W., 1939. Mental Disorders in Urban Areas: An Ecological Study of Schizophrenia and Other Psychoses. The University of Chicago Press, Chicago, IL.

Feins, J., Shroder, M., 2005. Moving to opportunity: the demonstration's design and its effects on mobility. Urban Stud. 42 (8), 1275–1299.
Findlay, S.E., 1988. The directionality and age selective of the health-migration relation: evidence from sequences of disability and mobility in the United States. Int. Migr. Rev. 22, 4–29.
Galea, S., Ahern, J., Rudenstine, S., Wallace, Z., Vlahov, D., 2005. Urban built environment and depression: a multilevel analysis. J. Epidemiol. Community Health. 59 (10), 822–827.
Galea, S., Ahern, J., Nandi, A., Tracy, M., Beard, J., Vlahov, D., 2007. Urban neighborhood poverty and the incidence of depression in a population-based cohort study. Ann. Epidemiol. 17, 171–179.
Halliday, T.J., Kimmitt, M.C., 2008. Selective migration and health in the USA, 1984–1993. Popul. Stud. 62 (3), 321–334.
Harris, K., 2013. The Add Health Study: Design and Accomplishments. Chapel Hill, NC.
Ioannides, Y., Zabel, J., 2008. Interactions, neighborhood selection and housing demand. J. Urban Econ. 63 (1), 229–252.
Jencks, C., Mayer, S.E., 1990. The social consequences of growing up in a poor neighborhood. In: Lynn Jr., L.E., McGeary, M.G.H. (Eds.), Inner City Poverty in the United States. National Academy Press, Washington, DC.
Jelleyman, T., Spencer, N., 2008. Residential mobility in childhood and health outcomes: a systematic review. J. Epidemiol. Community Health 62 (7), 584–592.
Joffe, M.M., Rosenbaum, P.R., 1999. Invited commentary: propensity scores. Am. J. Epidemiol. 150, 327–333.
Neighborhoods and Health. In: Kawachi, I., Berkman, L.F. (Eds.), Oxford University Press, New York, NY.
Larson, A., Bell, M., Young, A.F., 2004. Clarifying the relationships between health and residential mobility. Soc. Sci. Med. 59, 2149–2160.
van Lenthe, F.J., Martikainen, P., Mackenbach, J.P., 2007. Neighbourhood inequalities in health and health-related behaviour: results of selective migration? Health Place 13 (1), 123–137.
Leventhal, T., Brooks-Gunn, J., 2000. The neighborhoods they live in: the effects of neighborhood residence on child and adolescent outcomes. Psychol. Bull. 126 (2), 309–337.
Ludwig, J., Kling, J.R., Katz, L.F., Sanbonmatsu, L., Liebman, J.B., Duncan, G.J., et al., 2008. What can we learn about neighborhood effects from the Moving to Opportunity experiment? Am. J. Sociol. 114 (1), 144–188.
Mackenbach, J.P., Looman, C.W., van der Meer, J.B., 1996. Differences in the misreporting of chronic conditions, by level of education: the effect on inequalities in prevalence rates. Am. J. Public Health 86 (5), 706–711.
Mair, C.F., Diez Roux, A.V., Galea, S., 2008. Are neighborhood characterisics associated with depressive symptoms? A critical review. J. Epidemiol. Community Health 62 (11), 940–946.
March, D., Hatch, S.L., Morgan, C., Kirkbride, J.B., Bresnahan, M., Fearon, P., et al., 2008. Psychosis and place. Epidemiol. Rev. 30 (1), 84–100.
Mayer, S.E., Jencks, C., 1989. Growing up in poor neighborhoods: how much does it matter? Science 243 (4897), 1441–1445.
Norman, P., Boyle, P., Rees, P., 2005. Selective migration, health, and deprivation: a longitudinal analysis. Soc. Sci. Med. 60, 2755–2771.
Piro, F., Næss, Ø., Claussen, B., 2007. Area deprivation and its association with health in a cross-sectional study: are the results biased by recent migration? Int. J. Equity Health 6 (1), 10.
Rosenbaum, J.E., 1995. Changing the geography of opportunity by expanding residential choice: lessons from the Gautreaux program. Hous. Policy Debate 6 (1), 231–269.
Sampson, R.J., 2008. Moving to inequality: neighborhood effects and experiments meet social structure. Am. J. Sociol. 114 (1), 189–231.
Sampson, R.J., Sharkey, P., 2008. Neighborhood selection and the social reproduction of concentrated racial inequality. Demography 45 (1), 1–29.
Silver, E., Mulvey, E.P., Swanson, J.W., 2002. Neighborhood structural characteristics and mental disorder: Faris and Dunham revisited. Soc. Sci. Med. 55 (8), 1457–1470.
Subramanian, S.V., Glymour, M.M., Kawachi, I., 2007. Identifying causal ecologic effects on health: a methodological assessment. In: Galea, S. (Ed.), Macrosocial Determinants of Population Health. Springer Media, New York, pp. 301–331.
Teater, B., 2009. Factors predicting residential mobility among the recipients of the Section 8 Housing Choice Voucher Program. J. Sociol. Soc. Welf. 36 (3), 159–178.
Theall, K.P., Brett, Z.H., Shirtcliff, E.A., Dunn, E.C., Drury, S.S., 2013. Neighborhood disorder and telomeres: connecting children's exposure to community level stress and cellular response. Soc. Sci. Med. 85, 50–58.
Verheij, R.A., van de Mheen, H.D., de Bakker, D.H., Groenewegen, P.P., Mackenbach, J.P., 1998. Urban–rural variations in health in the Netherlands: does selective migration play a part? J. Epidemiol. Community Health 52, 487–493.
Wilson, W., 2011. When Work Disappears: The World of the New Urban Poor. Random House LLC, New York, NY.
Woldoff, R., Ovadia, S., 2008. Not getting their money's worth: African-American disadvantages in converting income, wealth, and education into residential quality. Urban Aff. Rev. 45 (1), 66–91.

Web Table 1. Nested Models Describing Association Between Predictors and Binary Smoking Outcome (0=not smoked in past 30 days; 1=smoked in past 30 days) in the National Longitudinal Study of Adolescent Health (N=16,070)

	Model 1			**Model 2**		
Fixed Effect Estimates	School Only	Neighborhood Only	Cross-Classified	School Only	Neighborhood Only	Cross-Classified
Intercept (SE)	-1.22 (0.06)	-1.20 (0.03)	-1.23 (0.06)	0.03 (0.01)	0.02 (0.005)	0.03 (0.01)
Individual-level						
Age				1.19 (1.16, 1.22)	1.21 (1.18, 1.24)	1.19 (1.17, 1.23)
Female				1.01 (0.94, 1.09)	1.00 (0.92, 1.07)	1.01 (0.94, 1.09)
Public Assistance				1.40 (1.22, 1.58)	1.43 (1.25, 1.63)	1.41 (1.23, 1.60)
High School degree				1.00 (0.90, 1.12)	1.01 (0.90, 1.15)	1.01 (0.89, 1.13)
Race						
White				Ref	Ref	Ref
Black				0.30 (0.26, 0.34)	0.28 (0.24, 0.31)	0.30 (0.26, 0.34)
Hispanic				0.68 (0.59, 0.78)	0.54 (0.48, 0.60)	0.68 (0.60, 0.78)
School-level						
Public Assistance						
High School Degree						
Percent White						
Neighborhood-level						
Public Assistance						
High School Degree						
Percent White						
Random Effect Estimates						
U3 neighborhood (SE)		0.31 (0.24, 0.38)*	0.05 (0.03, 0.09)*		0.10 (0.06, 0.15)*	0.01 (0.00, 0.02)
U2 school (SE)	0.36 (0.26, 0.48)*		0.36 (0.25, 0.49)*	0.14 (0.09, 0.20)*		0.13 (0.08, 0.20)*
Fit Statistics						
DIC			17522			17129

Model 1 presents the results for a null model (i.e., no covariates) for each model type: school-only multilevel model, neighborhood-only multilevel model, and the cross-classified multilevel model. Model 2 presents the same models as Model 1, except Model 2 includes individual-level predictors and covariates. For the fixed effect estimates, cell entries are odds ratio (OR) estimates and credible intervals. The intercept is presented as parameter estimate and standard error (SE). Random effects are presented as estimate and credible intervals. DIC refers to Deviance Information Criterion, a measure of model fit and complexity, and is only reported for the CCMM. Significant random effects are indicated by * (p<0.05).

Web Table 2. Nested Models Describing Association Between Covariates and Binary Smoking Outcome (0=not smoked in past 30 days; 1=smoked in past 30 days) in the National Longitudinal Study of Adolescent Health (N=16,070)

	Model 3		**Model 4**		**Model 5**
Fixed Effect Estimates	School Only	Cross-Classified	Neighborhood Only	Cross-Classified	Cross-Classified
Intercept (SE)	0.02 (0.003)	0.02 (0.003)	0.02 (0.004)	0.03 (0.01)	0.02 (0.004)
Individual-level					
Age	1.19 (1.16, 1.21)	1.19 (1.17, 1.21)	1.22 (1.20, 1.24)	1.18 (1.16, 1.20)	1.20 (1.17, 1.23)
Female	1.01 (0.94, 1.09)	1.01 (0.93, 1.09)	1.00 (0.93, 1.07)	1.01 (0.94, 1.09)	1.01 (0.94, 1.09)
Public Assistance	1.37 (1.20, 1.58)	1.37 (1.20, 1.56)	1.43 (1.25, 1.63)	1.39 (1.21, 1.60)	1.36 (1.18, 1.56)
High School degree	0.98 (0.87, 1.09)	0.99 (0.86, 1.12)	0.99 (0.87, 1.14)	1.00 (0.88, 1.16)	1.00 (0.88, 1.13)
Race					
White	Ref	Ref	Ref	Ref	Ref
Black	0.31 (0.27, 0.36)	0.32 (0.28, 0.36)	0.3 (0.25, 0.34)	0.3 (0.26, 0.36)	0.31 (0.26, 0.35)
Hispanic	0.72 (0.62, 0.81)	0.73 (0.64, 0.83)	0.55 (0.48, 0.63)	0.68 (0.59, 0.78)	0.71 (0.61, 0.81)
School-level					
Public Assistance	1.02 (1.01, 1.03)	1.02 (1.01, 1.03)			1.02 (1.01, 1.04)
High School Degree	0.99 (0.98, 1.00)	0.99 (0.98, 1.00)			0.99 (0.98, 1.00)
Percent White	1.01 (1.00, 1.01)	1.01 (1.00, 1.01)			1.01 (1.00, 1.01)
Neighborhood-level					
Public Assistance			1.00 (0.99, 1.01)	1.00 (0.99, 1.01)	0.99 (0.98, 1.00)
High School Degree			1.00 (0.99, 1.01)	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)
Percent White			1.00 (0.99, 1.01)	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)
Random Effect Estimates					
U3 neighborhood (SE)		0.02 (0.00, 0.04)	0.10 (0.06, 0.14)*	0.01 (0.00, 0.01)	0.02 (0.01, 0.04)*
U2 school (SE)	0.10 (0.06, 0.16)*	0.10 (0.06, 0.16)*		0.14 (0.09, 0.20)*	0.10 (0.06, 0.15)*
Fit Statistics					
DIC		17131		17129	17133

Model 3 presents the results of the school-only multilevel model and CCMM containing individual-level variables as well as the *school-level* measure of SES and race/ethnicity. Model 4 presents the results of the neighborhood-only multilevel model and CCMM containing individual-level variables combined with the *neighborhood-level* measure of SES and race/ethnicity. Model 5 presents the results of a CCMM containing all individual-, school-, and neighborhood-level variables. For the fixed effect estimates, cell entries are odds ratio (OR) estimates and credible intervals. The intercept is presented as parameter estimate and standard error (SE). Random effects are presented as estimate and credible intervals. DIC refers to Deviance Information Criterion, a measure of model fit and complexity, and is only reported for the CCMM. Significant random effects are indicated by * ($p<0.05$).

Web Table 3. Nested Cross-Classified Multilevel Models (CCMM) Describing Association Between Predictors and Number of Days Smoked in the Past 30 Days and Binary Smoking Outcome in the National Longitudinal Study of Adolescent Health (N=16,070) adjusting for neighborhoods with one respondent (n=970)

	Number of Days Smoked in Past 30 Days	Binary Smoking
Fixed Effect Estimates		
Intercept (SE)	-8.40 (1.05)	0.02 (0.01)
Individual-level		
Age	0.81 (0.71, 0.90)	1.18 (1.15, 1.21)
Female	0.06 (-0.21, 0.35)	1.01 (0.94, 1.08)
Public Assistance	0.68 (0.18, 1.17)	1.36 (1.19, 1.56)
High School degree (parent)	-0.22 (-0.69, 0.25)	0.98 (0.86, 1.12)
Race		
White	Ref	
Black	-4.13 (-4.63, -3.59)	0.30 (0.26, 0.35)
Hispanic	-1.84 (-2.36, -1.32)	0.70 (0.61, 0.81)
Neighborhood has one respondent	0.44 (-0.17, 1.05)	1.06 (0.89, 1.24)
School-level		
Public Assistance	0.07 (0.02, 0.12)	1.02 (1.01, 1.03)
High School Degree	-0.04 (-0.09, 0.003)	0.99 (0.98, 1.00)
Percent White	0.02 (0.001, 0.03)	1.01 (1.00, 1.01)
Neighborhood-level		
Public Assistance	-0.005 (-0.04, 0.03)	0.99 (0.98, 1.01)
High School Degree	0.01 (-0.01, 0.03)	1.005 (1.00, 1.01)
Percent White	0.0002 (-0.01, 0.01)	0.998 (0.995, 1.00)
Random Effect Estimates		
U3 neighborhood (SE)	0.21 (0.05, 0.50)*	0.01 (0.003, 0.02)
U2 school (SE)	1.79 (1.17, 2.60)*	0.10 (0.06, 0.16)*
U1 individual (SE)	80.6 (78.8, 82.4)*	-

For the fixed effect estimates, cell entries are parameter (beta) estimates and credible intervals for continuous days smoked outcome and parameter (OR) estimates and credible intervals for binary smoking outcome. The intercept is presented as parameter estimate and standard error (SE). Random effects are presented as estimate and credible intervals. Significant random effects are indicated by * ($p<0.05$).

Technical Appendix Part 1:
Running Cross-Classified Multilevel Models in MLWiN

**If you used this appendix for your analysis, please cite us:

Dunn, E.C., Richmond, T.K., Milliren, C.E., & Subramanian, S.V. Using Cross-Classified Multilevel Models to Disentangle School and Neighborhood Effects: An Example Focusing on Smoking Behaviors among Adolescents in the United States. *Health and Place*

Introduction

This first technical appendix is intended to show MLwiN users how to fit cross-classified multilevel models in MLwiN. A second technical appendix shows users how MLwiN can be executed through STATA.

Here, we provide detailed instruction on how to fit a cross-classified model with a continuous outcome followed by a brief overview of how to fit a cross-classified model with a binary outcome. Although the same general set of steps are taken for either model, we think it is easier to understand the procedures to fit a linear model and thus recommend readers start analyzing cross-classified models with continuous outcomes.

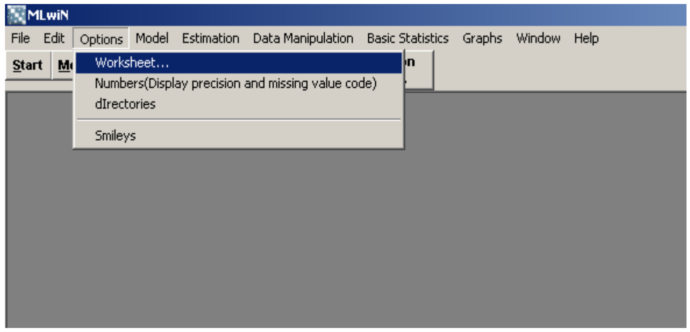
Our illustration uses data previously constructed, cleaned, and analyzed in the paper by Dunn and colleagues noted above. We used MLwiN version 2.26 for all analyses (Rasbash et al. 2012, Center for Multilevel Modelling).

In our analysis, all data manipulation (e.g., creating derived variables, recoding variables, etc.) was performed prior to importing the data into MLwiN. For ease of implementation, we recommend all data manipulation (e.g., creating derived variables, recoding variables, etc) occur in other programs (e.g., SAS, STATA) outside of MLwiN. Readers interested in specific data manipulation capabilities should refer to the MLwiN manual (Rasbash et al. 2012).

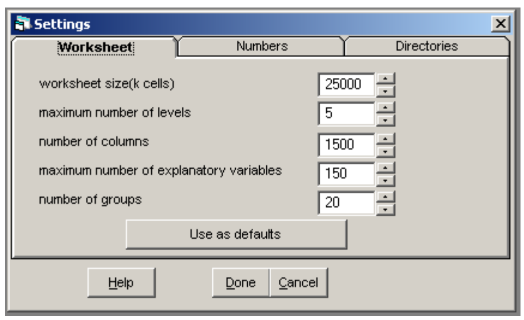
We would also like to note that MLwiN requires a constant variable consisting of a vector of 1's for all observations. The constant variable is necessary for modeling the intercept and thus is required to fit a model with random intercepts. We recommend you create this constant variable before importing the data into the program.

Getting Started in MLwiN

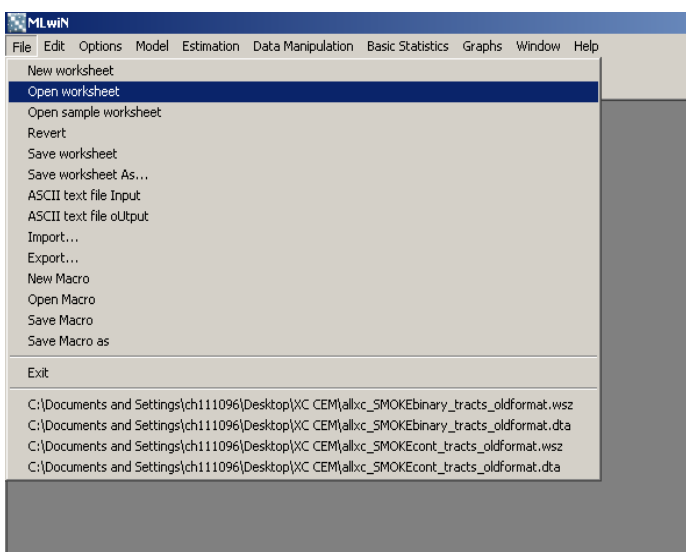
Before opening a dataset, the worksheet size should be adjusted to accommodate the size of data (i.e., number of observations) by clicking on *Worksheet* under the **Options** menu.If the number of cells listed in the worksheet is less than the actual number of observations in the data, some observations may be lost in the imported dataset.



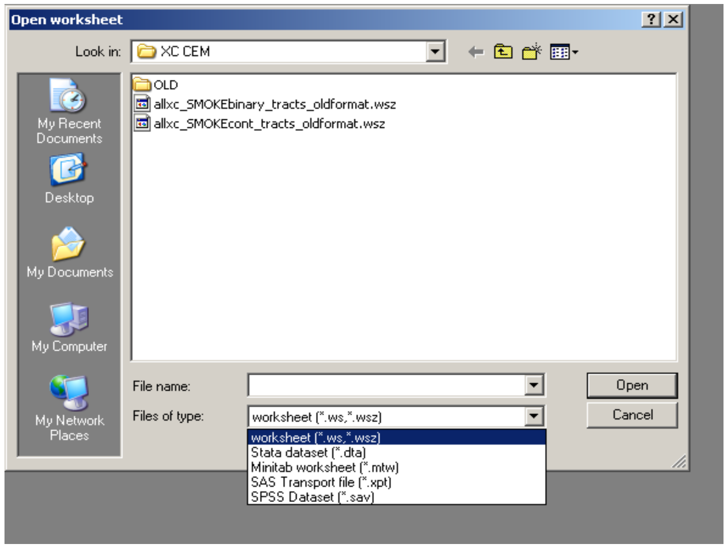
On the Worksheet tab, adjust the number of cells in the worksheet (i.e., number of observations), maximum number of levels (e.g. 2 for individual and school), number of columns (i.e., variables), number of explanatory variables allowed in a single model, and number of groups (i.e., number of categories within the categorical variables) to fit the dataset. The number of cells must be at least equal to the number of observations in the dataset. Adding more cells, levels, or other features to the worksheet, that exceed the actual number in the dataset, will not affect model specification or results in any way. Thus, it is reasonable to always choose numbers larger than what you think you will need. After all adjustments are made, click the "**Done**" button for the changes to take effect.



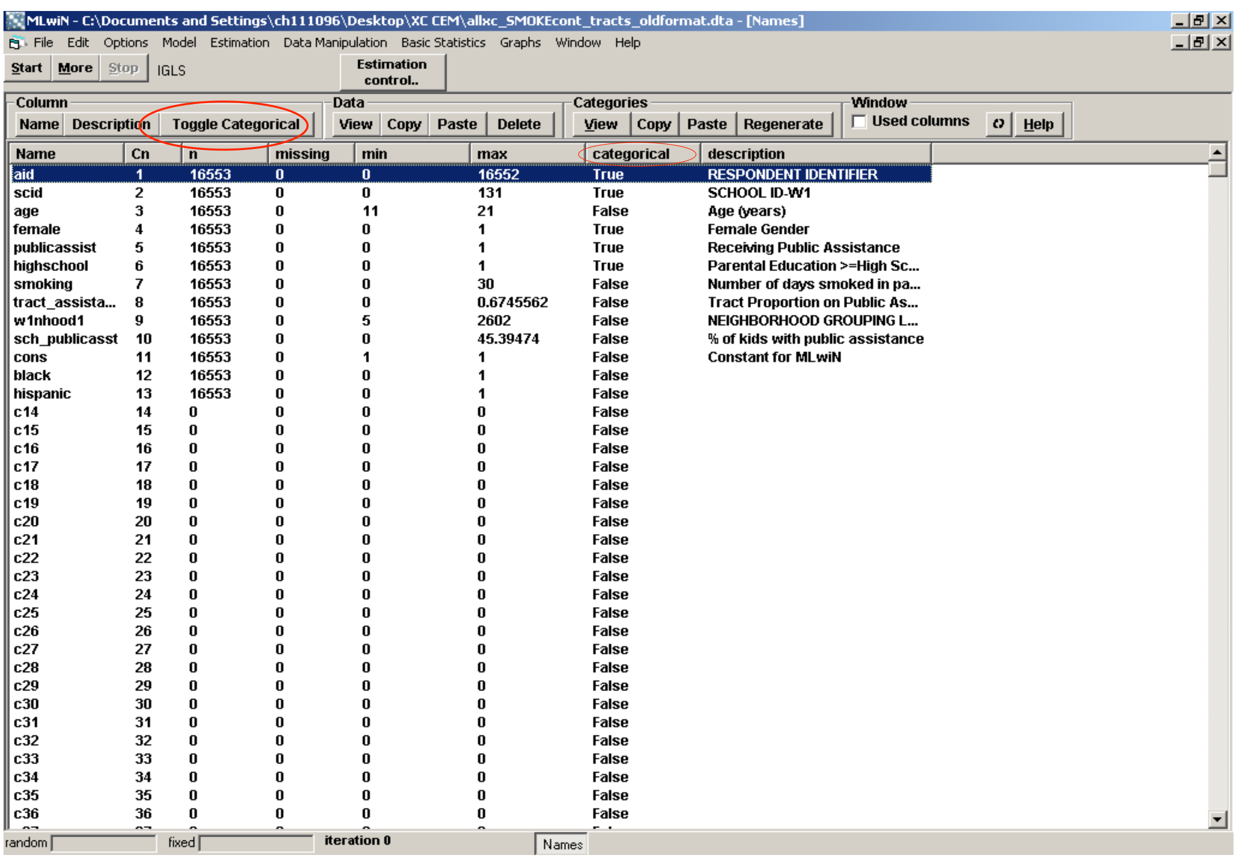
In order to open your dataset, go to the **File** menu and select *Open worksheet*.



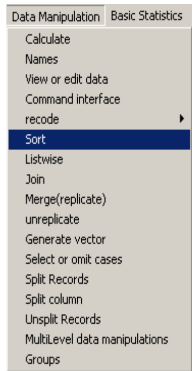
You can open previously saved MLwiN worksheets (.wsz or .ws) as well as datasets from other programs (Stata, SAS, SPSS) by selecting the file type from the drop-down menu



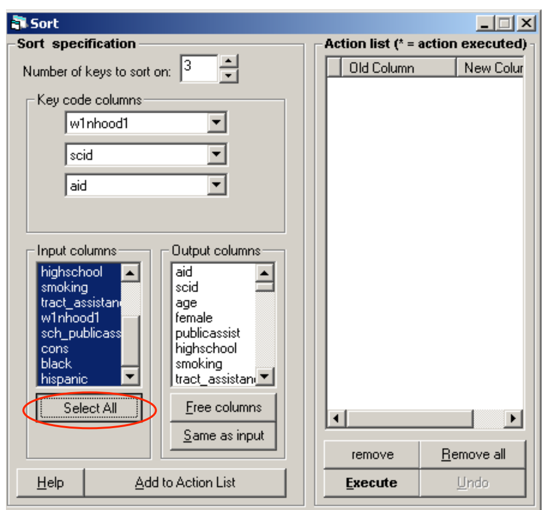
Once a dataset has been opened, the variable names and attributes are displayed in the Names window. A 'categorical' column indicates whether MLwiN recognizes the variable as categorical (indicated by True in the Categorical column) or continuous (indicated by False in the Categorical column) based on the values. This attribute may be changed using the "**Toggle Categorical**" button. Miscategorization of variables will result in continuous variables being treated as categorical and vice versa.



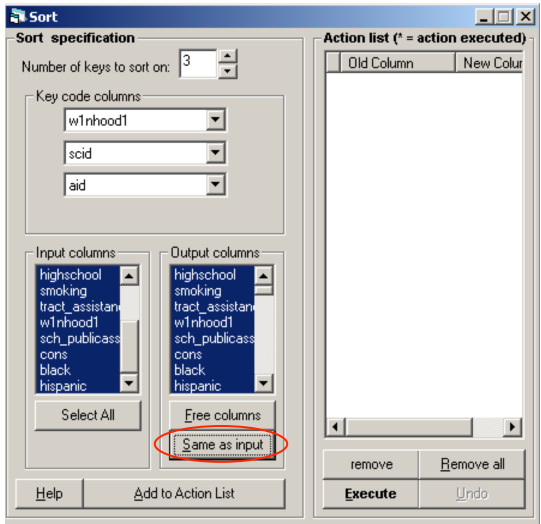
To implement multilevel models, the data must be sorted in ascending order by the identifier variable within each level. Thus, the level 1 identifier (e.g., individual subject identification number) must be sorted within level 2 (e.g., school identification number), level 2 within level 3, etc.. The *Sort* function can be accessed from the **Data Manipulation** menu.



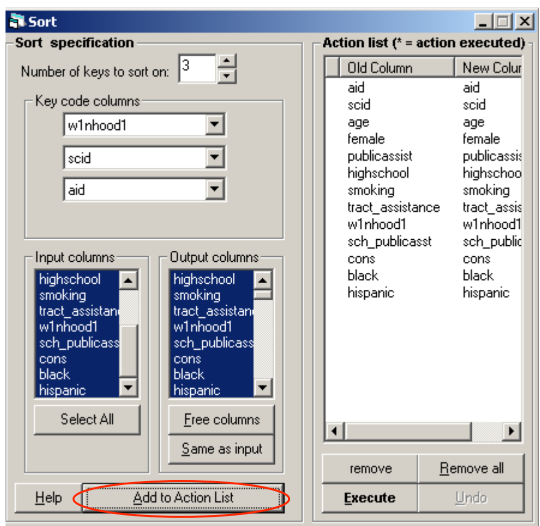
To sort the data, first select the number of levels to sort on (MlwIN refers to this as the number of 'keys' to sort on) and choose the level identifiers from the drop-down menu (under 'key code columns'). The highest level in the hierarchy (e.g. school) should be the first variable and lowest level should be last (e.g. individual). The level identifier variables to be sorted on should be selected next under the Input Columns. Variables can be chosen individually, or click the "**Select All**" button to sort all variables.



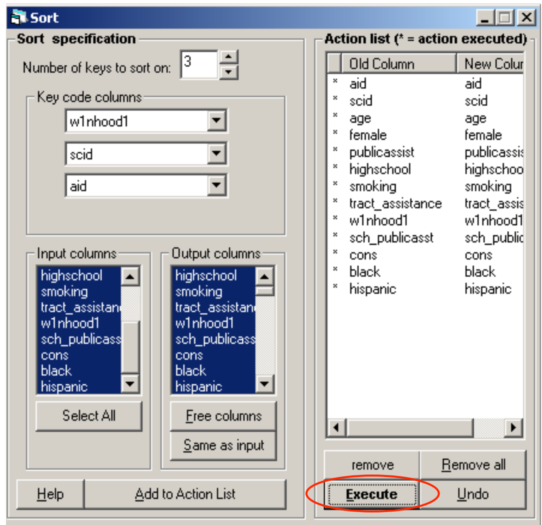
To save these sort changes, the data must be "output." To output all of the sorted variables without creating copies of the variables (recommended), click the "**Same as input**" button under Output columns. The sorted variables can also be output to empty columns by clicking "**Free columns**" instead.



Clicking the "**Add to Action List**" button will create a queue for the variables to be sorted.



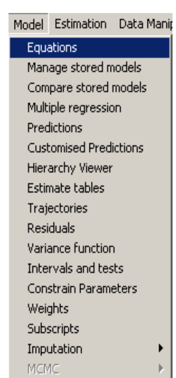
To perform the sort, click the "**Execute**" button. Once the sort is complete, an asterisk will appear next to all the variable names that have been sorted.



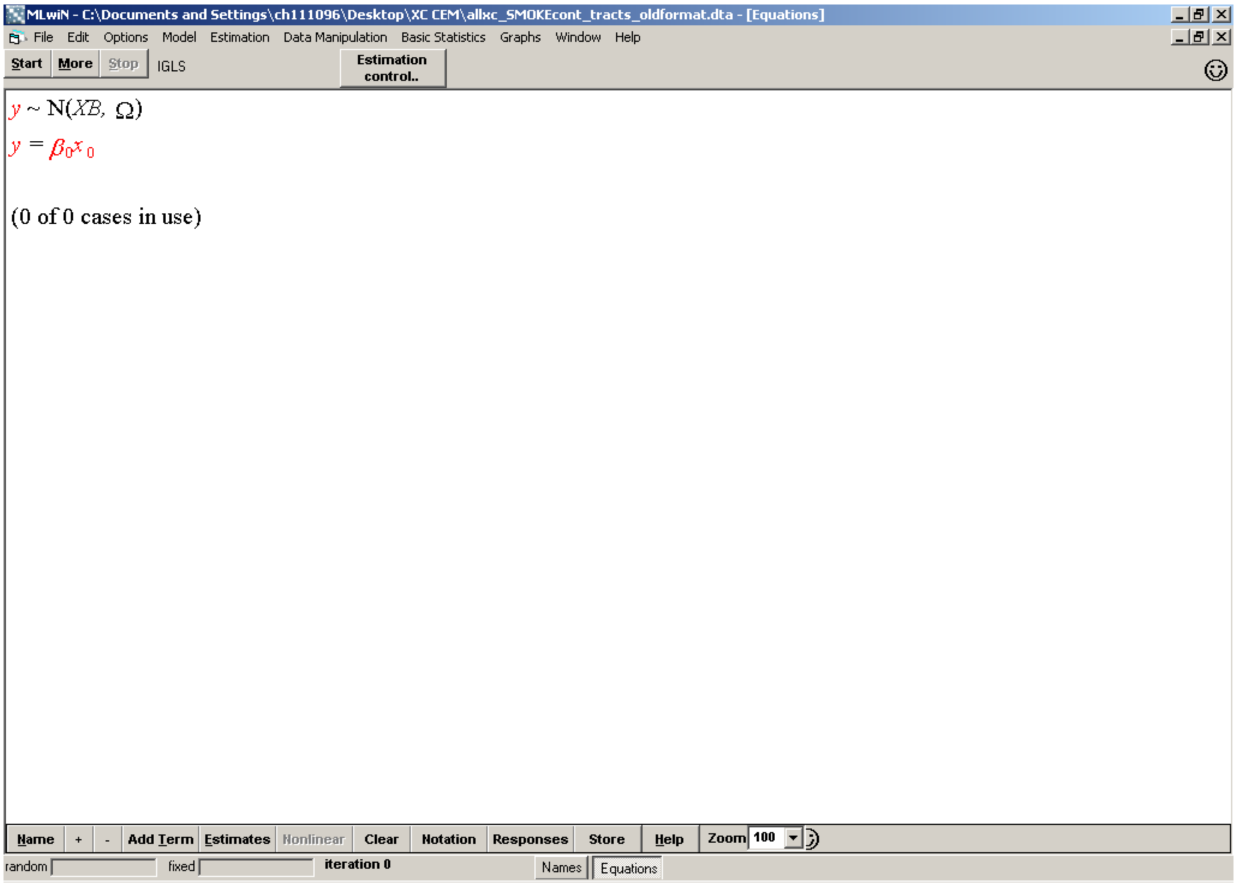
Specifying the Model: Example with a Continuous Outcome

Note: When fitting any cross-classified multilevel model in MLwiN, the model must first be specified assuming a nested hierarchical structure (i.e., individuals clustered in schools and schools clustered in neighborhoods). This initial model provides starting values. The model must then be refit with the cross-classified structure, using the parameters obtained from the first ("nested structure") model as starting values. MLwiN uses the Iterative Generalized Least Squares method to implement hierarchically nested models while cross-classified models must be implemented using Markov Chain Monte Carlo (MCMC) methods. MCMC is a simulation-based method estimating the parameters by re-sampling the data to produce more accurate estimates of the unknown parameters. IGLS and traditional least-squares regression methods give point estimates for unknown parameters calculated from the sample data; however there is no re-sampling performed. MCMC allows for more complex models to be fitted, including cross-classified models. For more information on MCMC, see Chapter 1 of the MCMC manual (Browne, 2012).

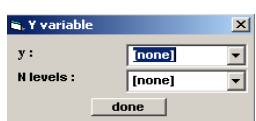
To specify the first model, choose *Equations* from the **Model** menu.



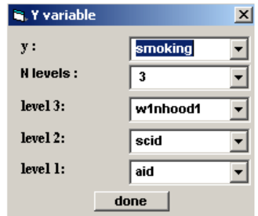
This will bring up the Equations window, where the first model can be specified. Here, you will specify the dependent and independent variables, distribution function of the dependent variable, and the number of levels to model. Red text in the Equations window indicates elements of the model that have not yet been specified. The nested structure model is run using Iterative Generalized Least Squares regression which is indicated by the IGLS under the toolbar. This regression estimation method is the default when MLwiN is opened.



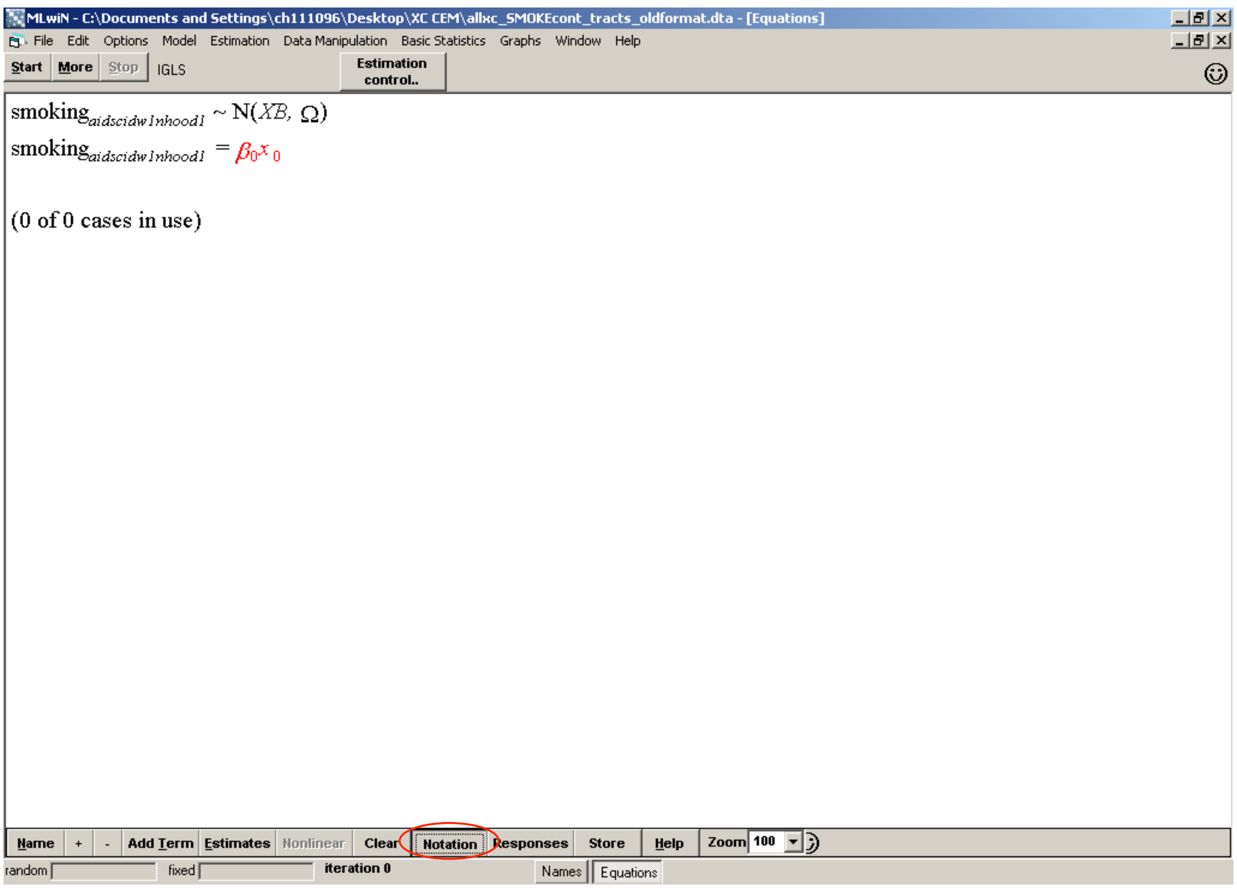
To specify the model, click on the red "y" on the left side of the equation to specify the dependent variable as well as the number of levels. Choose the outcome variable from the drop-down list. Then choose the appropriate number of levels to include in the model.



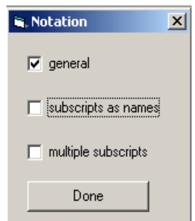
Once the number of levels has been specified, drop-down menus will appear for each level. The level identifiers are specified here. Level 3 is the highest level unit (in our example it is the neighborhood, coded as: w1nhood1), while level 1 is the lowest level (it is the individual, coded as: aid in our example).



Here the variable *smoking* has been specified as the dependent variable, being modeled at 3 levels neighborhood: winhood1, school: scid, and individual: aid. Click done when the dependent variables and levels have been specified. Clicking the "**Notation**" button, we can change the way subscripts are displayed for the levels.

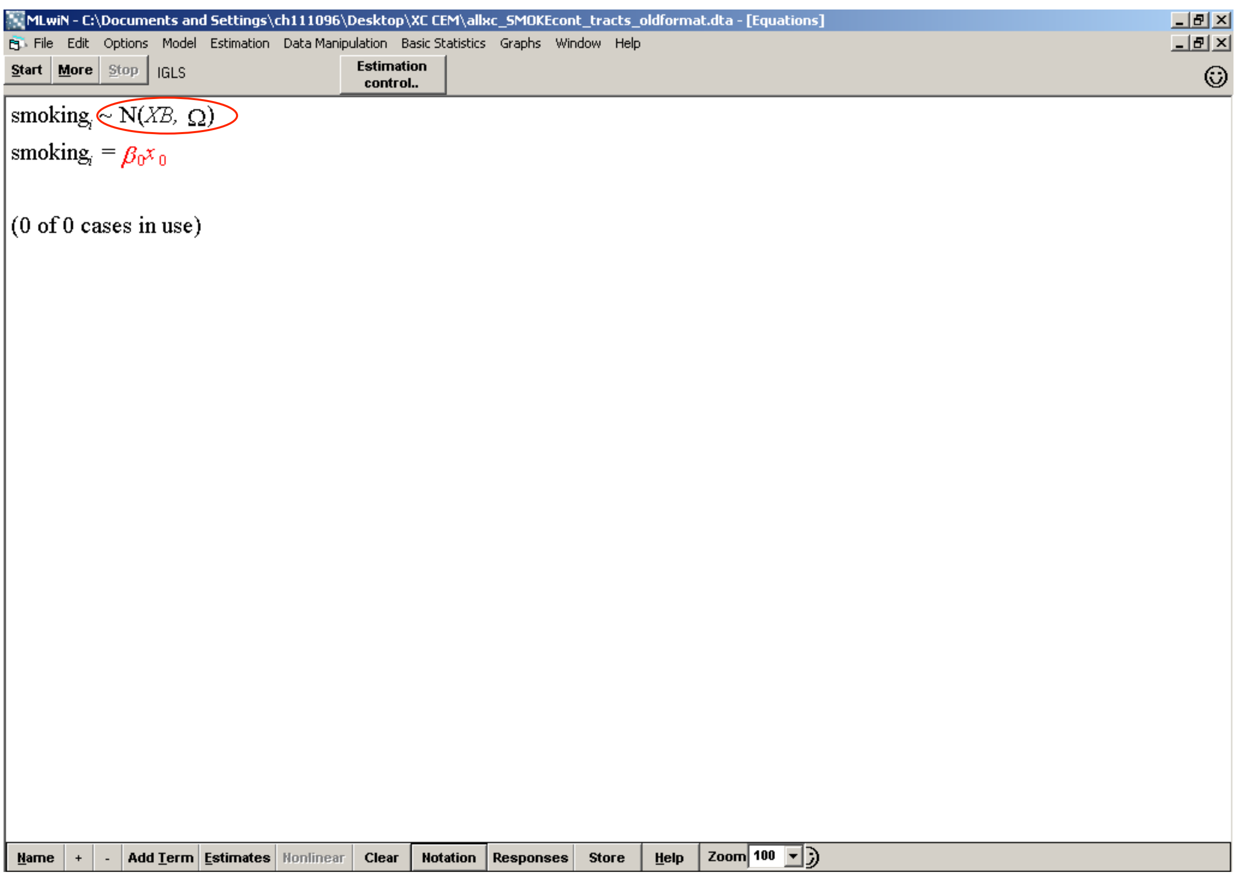


It is particularly useful when implementing the cross-classified model to display a single letter subscripts for all levels. This is because the levels are no longer nested which is not properly accounted for by the traditional hierarchical notation (ith student in jth school in kth neighborhood). There are more complex relationships between school and neighborhood in a cross-classified model (i.e. there can be both crossed and nested relationships between the two levels) which cannot easily be indicated in the traditional notation. For this reason, we use a classification notation with a single subscript to indicate that the levels are not entirely nested. To display letter subscripts for all levels, uncheck the box for "subscripts as names." To display a single "i" for all three levels, unclick 'multiple subscripts,' to further simplify the display of the model.



Now the single subscript is displayed instead of the level identifiers.

The default distribution function for the dependent variable is a normal distribution, which is indicated by the *N* on the right side of the top equation. This indicates a normal distribution (N=normal) for the fixed estimate, Xβ and a random part indicated by Ω. The distribution of the dependent variable can be changed by double-clicking the *N*.



Technical Appendix Part 2:
Running Cross-Classified Multilevel Models in MLwiN through STATA using the runmlwin package

**If you used this appendix for your analysis, please cite us:

Dunn, E.C., Richmond, T.K., Milliren, C.E., & Subramanian, S.V. Using Cross-Classified Multilevel Models to Disentangle School and Neighborhood Effects: An Example Focusing on Smoking Behaviors among Adolescents in the United States. *Health and Place*.

Introduction

This second technical appendix is intended to show users how to fit cross-classified multilevel models in MLwiN via STATA. With both STATA and MLwiN installed, MLwiN can be executed through STATA, which is convenient for data management as well as looking at and interpreting output. The following tutorial uses MLwiN version 2.29 and STATA version 13.1 (College Station, TX). Additional details are also available here:

> Leckie, G, C Charlton. Runmlwin: A program to run the MLwiN multilevel modeling software f rom within Stata. 2012. Journal of Statistical Software 52(11): 1-40.

The runmlwin package must first be installed from the Statistical Software Components (SSC) archive, which is a repository of user contributed STATA commands. The following accesses the SSC and installs the runmlwin command.

```
ssc install runmlwin
```

The filepath where MLwiN is located must be specified so that STATA can find the program to execute any runmlwin commands. The following command specifies the filepath where MLwiN is located on the user's computer (users should substitute their own filepath).

```
global MLwiN_path "C:\Program Files\MLwiN v2.29\i386\mlwin.exe"
```

Linear Models
The following command fits a two-level hierarchical (i.e., multilevel) linear null model with random intercepts for school and individual predicting the number of days smoked in the past 30. The data must first be sorted by the level identifiers just as it would be sorted within MLwiN. Additionally, a constant variable (here called *cons*) with a value of 1 for every observation must first be constructed in order to fit the intercept. The level 2 school identifier is *scid* and level 1 individual identifier is *aid*.

```
sort scid aid
runmlwin smoking cons, level2(scid:cons) level1(aid:cons) nopause
```

The following output generated by the above command appears in the STATA results window after MLwiN opens, runs, and closes.

```
MLwiN 2.29 multilevel model                    Number of obs      =     16070
Normal response model
Estimation algorithm: IGLS

-----------------------------------------------------------
                 |   No. of       Observations per Group
  Level Variable |   Groups    Minimum    Average    Maximum
-----------------+-----------------------------------------
            scid |      128         18      125.5       1012
-----------------------------------------------------------

Run time (seconds)   =       5.94
Number of iterations =          3
Log likelihood       = -58446.645
Deviance             =  116893.29
------------------------------------------------------------------------------
     smoking |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cons |   3.825463    .220559    17.34   0.000     3.393175    4.257751
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
Level 2: scid                |
                   var(cons) |   5.330074   .7726001      3.815806    6.844343
-----------------------------+------------------------------------------------
Level 1: aid                 |
                   var(cons) |   83.08062   .9305611      81.25675    84.90449
------------------------------------------------------------------------------
```

The output gives information about the number of level 2 groups, model run time, iterations, and fit statistics as well as the parameter estimate, test statistic and 95% confidence interval for the fixed effects (in this case just the intercept). Additionally, the random effect parameter estimates for school and individual variance in the intercept appear in the last output table.

To fit a cross-classified 3-level null model predicting number of smoking days with random intercepts for neighborhood, school, and individual the model must first be fit using the iterative generalized least squares (IGLS) algorithm with a hierarchical structure and re-run in the Bayesian Markov Chain Monte Carlo (MCMC) framework in order to properly account for the cross-classified data structure. The parameter estimates from the naïve IGLS model are used as starting values or priors for the MCMC cross-classified model. Again, the data should first be sorted by the level identifiers. The level 3 neighborhood identifier is *w1nhood1*, level 2 school identifier is *scid*, and level 1 individual identifier is *aid*.

The first model is quietly run (**quietly runmlwin**) and fits the data using a hierarchical structure. The second model is fit using the cross-classified structure in a Bayesian framework using MCMC. The option **mcmc(cc)** indicates that the data are cross-classified (cc) and the model should be fit using MCMC while the option **initsprevious** indicates that the parameter estimates from the first model (assuming hierarchical structure) should be used as starting values for the second model. The output from the first model will not appear in the output window if the **quietly** option is used.

```
sort w1nhood1 scid aid
quietly runmlwin smoking cons, level3(w1nhood1:cons) level2(scid:cons)
level1(aid:cons) nopause

runmlwin smoking cons, level3(w1nhood1:cons) level2(scid:cons)
level1(aid:cons) mcmc(cc) initsprevious nopause
```

After submitting the above command, the cross-classified results on the following page appear in the STATA results window.

The MCMC cross-classified model output gives information on the number of observations nested within each level, how long the model was run and for how many iterations (burnin, chain, thinning, etc.) and model fit statistics (deviance, DIC) in addition to the parameter estimates for the fixed and random effects at each level along with 95% credible intervals (Bayesian confidence interval).

```
MLwiN 2.29 multilevel model                    Number of obs      =      16070
Normal response model
Estimation algorithm: MCMC

-----------------------------------------------------------
                |   No. of       Observations per Group
 Level Variable |   Groups    Minimum    Average    Maximum
----------------+------------------------------------------
       w1nhood1 |     2111          1        7.6        260
           scid |      128         18      125.5       1012
-----------------------------------------------------------

Burnin                     =        500
Chain                      =       5000
Thinning                   =          1
Run time (seconds)         =       20.6
Deviance (dbar)            =  116566.02
Deviance (thetabar)        =  116396.16
Effective no. of pars (pd) =     169.87
Bayesian DIC               =  116735.89
------------------------------------------------------------------------------
     smoking |      Mean    Std. Dev.     ESS     P       [95% Cred. Interval]
-------------+----------------------------------------------------------------
        cons |  3.881897   .2225641      293   0.000     3.463005    4.319552
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters |     Mean   Std. Dev.    ESS      [95% Cred. Int]
----------------------------+-------------------------------------------------
Level 3: w1nhood1           |
                  var(cons) |  .4572102  .1893233      14   .1262233  .8794164
----------------------------+-------------------------------------------------
Level 2: scid               |
                  var(cons) |  5.362256  .8018588    2527   3.952652  7.065619
----------------------------+-------------------------------------------------
Level 1: aid                |
                  var(cons) |  82.72698  .9336718    3899   80.93797  84.58237
------------------------------------------------------------------------------
```

Additional fixed effects can be added to the model by including them after the outcome variable specification. The following example adds additional predictors at the individual, school, and neighborhood levels first fitting a naïve hierarchical model in IGLS and then using the parameter estimates as starting values for the cross-classified MCMC model.

```
sort w1nhood1 scid aid
quietly runmlwin smoking age female publicassist highschool black
hispanic sch_publicasst sch_hsed sch_propwhite tract_assistance
tractprop_hs tractprop_white cons, level3(w1nhood1:cons)
level2(scid:cons) level1(aid:cons) nopause

runmlwin smoking age female publicassist highschool black hispanic
sch_publicasst sch_hsed sch_propwhite tract_assistance tractprop_hs
```

```
tractprop_white cons, level3(w1nhood1:cons) level2(scid:cons)
level1(aid:cons) mcmc(cc) initsprevious nopause
```

Output for the cross-classified model with additional fixed effects can be found on the following page.

```
MLwiN 2.29 multilevel model                    Number of obs      =     16070
Normal response model
Estimation algorithm: MCMC
```

Level Variable	No. of Groups	Observations per Group Minimum	Average	Maximum
w1nhood1	2111	1	7.6	260
scid	128	18	125.5	1012

```
Burnin                         =        500
Chain                          =       5000
Thinning                       =          1
Run time (seconds)             =       34.5
Deviance (dbar)                =  116131.02
Deviance (thetabar)            =  115998.01
Effective no. of pars (pd)     =     133.01
Bayesian DIC                   =  116264.02
```

smoking	Mean	Std. Dev.	ESS	P	[95% Cred.	Interval]
age	.8099638	.0505673	2956	0.000	.7131925	.9083905
female	.0631583	.1415275	4474	0.333	-.2096058	.3383548
publicassist	.6808348	.2573457	5083	0.003	.1830963	1.18176
highschool	-.2111972	.2340137	5037	0.180	-.6790206	.2505704
black	-4.112608	.2591757	4401	0.000	-4.630536	-3.612959
hispanic	-1.835435	.2730815	4027	0.000	-2.364895	-1.316999
sch_public~t	.0680946	.0251765	861	0.004	.0184816	.1170105
sch_hsed	-.0431506	.0222876	888	0.026	-.0859315	.0008327
sch_propwh~e	.0184719	.008447	885	0.014	.001701	.0349013
tract_assi~e	-.0049191	.0192253	2964	0.406	-.0425568	.0325781
tractprop_hs	.0107189	.0101577	2632	0.144	-.0089962	.0307047
tractprop_~e	.0002768	.0055375	2564	0.485	-.0105045	.010997
cons	-8.445673	1.036984	1900	0.000	-10.43999	-6.442746

Random-effects Parameters	Mean	Std. Dev.	ESS	[95% Cred.	Int]
Level 3: w1nhood1					
var(cons)	.2507093	.1310143	14	.0556763	.5244083
Level 2: scid					
var(cons)	1.719584	.3491465	1228	1.123541	2.485247
Level 1: aid					
var(cons)	80.53305	.9046208	4257	78.81626	82.29888

Logistic Models
Fitting cross-classified logistic models with a binary outcome is very similar to the linear models outlined above. The distribution and link function must be specified as binomial and logit, respectively instead of the default normal distribution used in the linear models (NOTE: MLwiN has other distribution and link functions not outlined in this tutorial). An additional constant variable for the denominator must be constructed. In the following example the denominator variable is called *denom* and has a value of 1 for all observations.

The following command fits a hierarchical null model with school as level 2 and individual as level 1 predicting smoking status (*smoking2* coded 0=non-smoker, 1=smoker). A random intercept is fitted for school. The **discrete** option specifies that the outcome variable is discrete and not continuous and that the distribution function is binomial (**dist(binomial)**) with a logit link function (**link(logit)**). The **denom** option specifies the variable that should be used to calculate the denominators for the higher level units which is called *denom*.

```
sort scid aid
runmlwin smoking2 cons, level2(scid:cons) level1(aid:)
discrete(dist(binomial) link(logit) denom(denom)) nopause
```

Output from the above command is shown below. The output gives information on the number of lower level units nested in the higher level units, run time and number of iterations as well as parameter estimates and 95% confidence intervals for the fixed and random effects. With logistic models, there is no variance parameter output for the lowest level (in this case individual) because individual variance in a binary outcome is a function of the proportion of individuals who have the outcome.

```
MLwiN 2.29 multilevel model                    Number of obs      =     16070
Binomial logit response model
Estimation algorithm: IGLS, MQL1
```

Level Variable	No. of Groups	Observations per Group Minimum	Average	Maximum
scid	128	18	125.5	1012

```
Run time (seconds)    =        4.95
Number of iterations =            5
```

smoking2	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
cons	-1.130177	.0517373	-21.84	0.000	-1.23158	-1.028774

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
Level 2: scid				
var(cons)	.2845968	.0425854	.2011309	.3680626

Next a 3-level cross-classified null model with random intercepts for school and neighborhood is fit. Just as with the linear models, the logistic models must first be run in IGLS assuming a hierarchical structure and then refit in MCMC to account for the cross-classified structure using the parameter estimates from the first model as the starting values for the Bayesian MCMC. Similar to the linear models, the naïve hierarchical model is quietly run to suppress the output followed by the cross-classified model.

```
sort w1nhood1 scid aid
```

```
quietly runmlwin smoking2 cons, level3(w1nhood1:cons)
level2(scid:cons) level1(aid:) discrete(dist(binomial) link(logit)
denom(denom)) nopause

runmlwin smoking2 cons, level3(w1nhood1:cons) level2(scid:cons)
level1(aid:) discrete(dist(binomial) link(logit) denom(denom))
mcmc(cc) initsprevious nopause
```

The output for the cross-classified null logistic model can be found below.

```
MLwiN 2.29 multilevel model                     Number of obs      =     16070
Binomial logit response model
Estimation algorithm: MCMC
```

Level Variable	No. of Groups	Observations per Group Minimum	Average	Maximum
w1nhood1	2111	1	7.6	260
scid	128	18	125.5	1012

```
Burnin                        =        500
Chain                         =       5000
Thinning                      =          1
Run time (seconds)            =       63.6
Deviance (dbar)               =   17331.04
Deviance (thetabar)           =   17140.24
Effective no. of pars (pd)    =     190.80
Bayesian DIC                  =   17521.84
```

smoking2	Mean	Std. Dev.	ESS	P	[95% Cred.	Interval]
cons	-1.226344	.0564287	92	0.000	-1.337117	-1.120508

Random-effects Parameters	Mean	Std. Dev.	ESS	[95% Cred.	Int]
Level 3: w1nhood1					
var(cons)	.0482382	.0168734	11	.0256405	.0872135
Level 2: scid					
var(cons)	.3561811	.0605399	1012	.2525054	.4894393

More predictors can be added to the model by including additional variables after the outcome variable. Here fixed effects are added at the individual, school, and neighborhood levels. The last line in the cross-classified model requests that a table of odds ratios be output in addition to the table of logit parameter estimates.

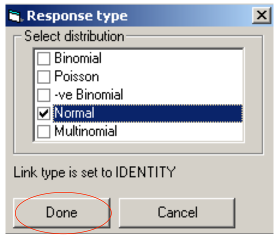
```
sort w1nhood1 scid aid
quietly runmlwin smoking2 age female publicassist highschool black
hispanic sch_publicasst sch_hsed sch_propwhite tract_assistance
tractprop_hs tractprop_white cons, level3(w1nhood1:cons)
level2(scid:cons) level1(aid:) discrete(dist(binomial) link(logit)
denom(denom)) nopause
```

```
runmlwin smoking2 age female publicassist highschool black hispanic
sch_publicasst sch_hsed sch_propwhite tract_assistance tractprop_hs
tractprop_white cons, level3(w1nhood1:cons) level2(scid:cons)
level1(aid:) discrete(dist(binomial) link(logit) denom(denom))
mcmc(cc) initsprevious nopause
runmlwin, noheader noretable or
```

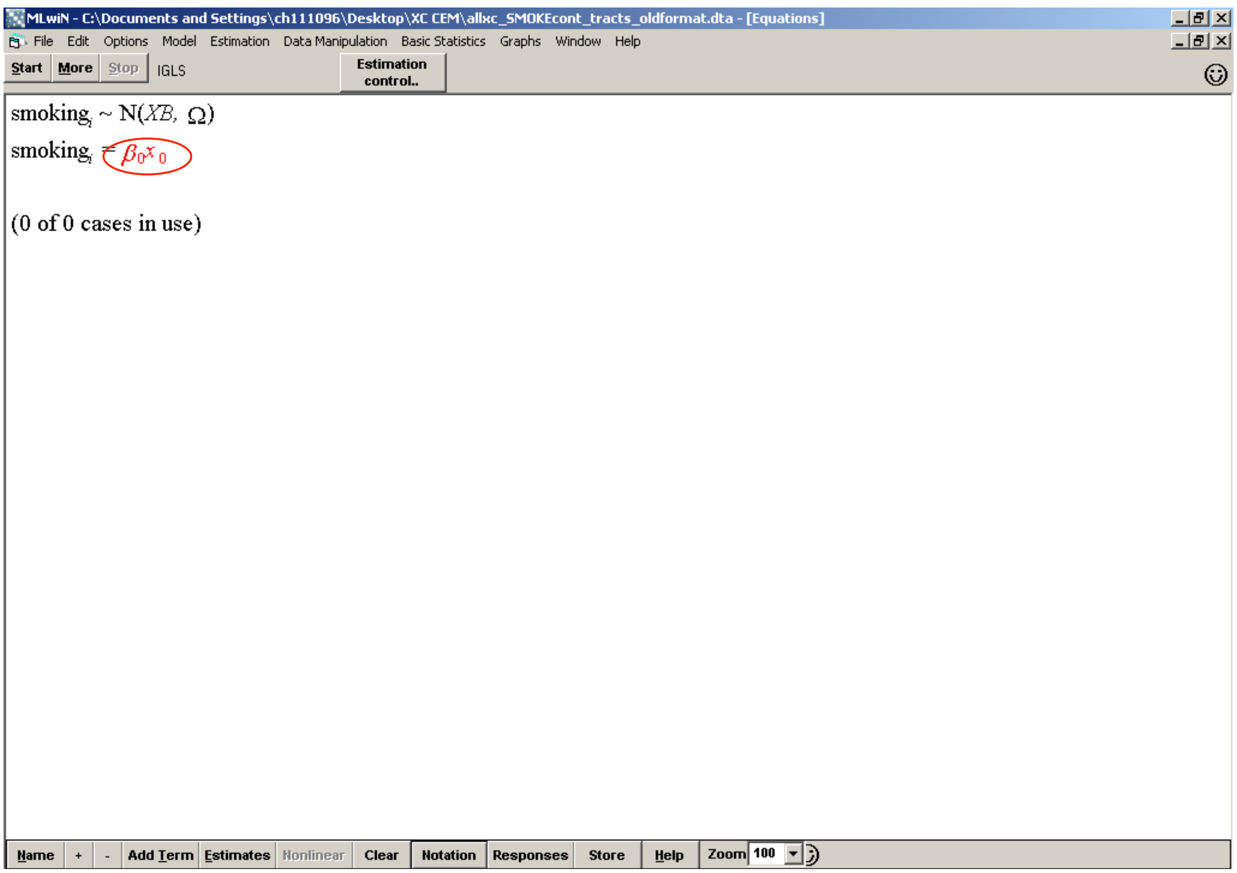
The output for the model with additional fixed effects can be found on the following page. Using the odds ratio option will give two tables of parameter estimates, one of the logits and one with odds ratios.

```
MLwiN 2.29 multilevel model                     Number of obs      =     16070
Binomial logit response model
Estimation algorithm: MCMC

-----------------------------------------------------------
                |   No. of       Observations per Group
 Level Variable |   Groups    Minimum    Average    Maximum
----------------+------------------------------------------
       w1nhood1 |     2111          1        7.6        260
           scid |      128         18      125.5       1012
-----------------------------------------------------------

Burnin                     =        500
Chain                      =       5000
Thinning                   =          1
Run time (seconds)         =        153
Deviance (dbar)            =   17004.39
Deviance (thetabar)        =   16875.57
Effective no. of pars (pd) =     128.82
Bayesian DIC               =   17133.21
------------------------------------------------------------------------------
    smoking2 |      Mean    Std. Dev.     ESS      P       [95% Cred. Interval]
-------------+----------------------------------------------------------------
         age |   .1833469   .0123686       8    0.000     .1554669    .2065977
      female |   .0099364   .0381066     577    0.398    -.0617398    .0852968
publicassist |   .3061968   .0719862     628    0.000     .1652309    .4457451
  highschool |  -.0023873   .0615281      85    0.479    -.1260161    .1207112
       black |  -1.184582   .0784246     175    0.000    -1.341672   -1.040694
    hispanic |  -.3479473   .0710086     313    0.000    -.4930292   -.2129315
sch_public~t |   .0235929    .006303      48    0.000     .0121545    .0363701
    sch_hsed |  -.0129069   .0056384      45    0.012    -.0234648   -.0011173
sch_propwh~e |   .0065937   .0021187      33    0.000     .0026996    .0109378
tract_assi~e |   -.006447   .0058402      48    0.144    -.0167468    .0049644
tractprop_hs |   .0049704   .0028578      46    0.039    -.0003789    .0107634
tractprop_~e |  -.0012386   .0016749      19    0.250    -.0052289    .0017178
        cons |  -4.126022   .2606254       6    0.000    -4.566837    -3.59991
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters |    Mean    Std. Dev.   ESS     [95% Cred. Int]
----------------------------+-------------------------------------------------
Level 3: w1nhood1           |
                  var(cons) |   .020705  .0072227      9    .0103887  .0371728
----------------------------+-------------------------------------------------
Level 2: scid               |
                  var(cons) |  .0988218  .0236526    296    .0589949  .1508862
------------------------------------------------------------------------------

. runmlwin, noheader noretable or

------------------------------------------------------------------------------
    smoking2 | Odds Ratio   Std. Dev.     ESS      P       [95% Cred. Interval]
-------------+----------------------------------------------------------------
         age |   1.201323    .014823       8    0.000     1.168203    1.229488
      female |    1.01072   .0385453     578    0.398     .9401275     1.08904
publicassist |   1.361772   .0980714     631    0.000     1.179665    1.561653
  highschool |   .9995055    .061538      85    0.479     .8816007    1.128299
       black |   .3068144   .0240339     177    0.000     .2614083    .3532096
    hispanic |   .7079151   .0501854     320    0.000     .6107734    .8082115
sch_public~t |   1.023894    .006457      48    0.000     1.012229     1.03704
    sch_hsed |   .9871917   .0055684      45    0.012     .9768084    .9988834
sch_propwh~e |   1.006618   .0021332      33    0.000     1.002703    1.010998
tract_assi~e |   .9935907   .0058048      48    0.144     .9833926    1.004977
tractprop_hs |   1.004987   .0028725      46    0.039     .9996211    1.010822
tractprop_~e |   .9987636   .0016721      19    0.250     .9947848    1.001719
        cons |   .0167114   .0044696       7    0.000     .0103908    .0273262
------------------------------------------------------------------------------
```

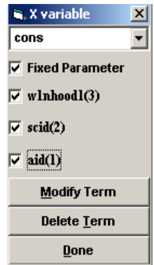
In our example, the outcome variable *smoking* is continuous and for illustrative purposes will be treated as though it is normally distributed. Therefore the model will be specified using the normal distribution. Binomial, Poisson, negative binomial, and multinomial distributions are other options depending on the distribution of the outcome. Once the distribution has been selected, click the "**Done**" button to continue specifiying the model.



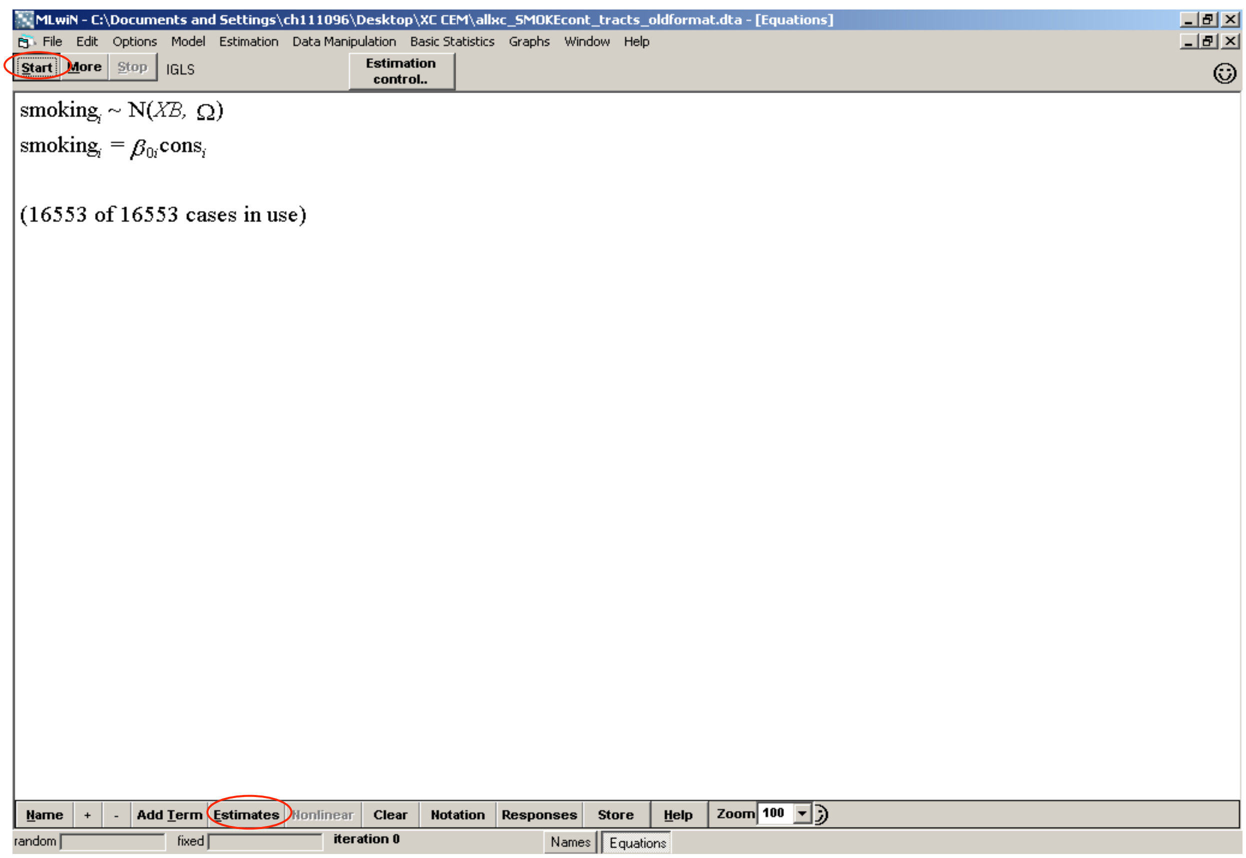
To add a random intercept to the model, double click on the red $\beta_0 x_0$.



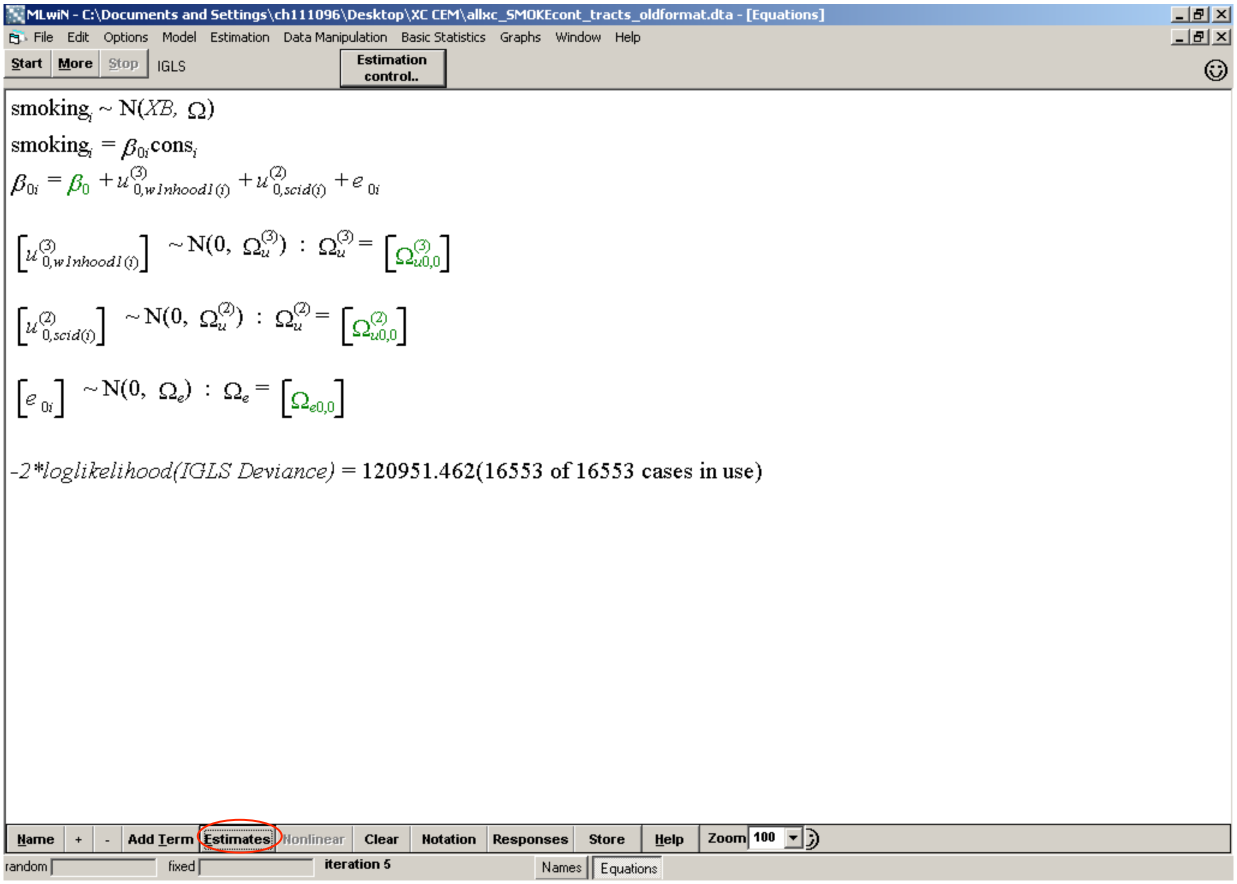
Choose the constant variable (i.e. the vector of 1's) from the drop-down menu (we called this variable *cons*). As noted previously, the constant variable is necessary for fitting the intercept(s). For a random-intercepts model, which allows for random intercepts at all levels, check the boxes for each level as well as the Fixed Parameter. Click done when you are finished.



Now that the null random-intercepts model has been fully specified, the equations are black indicating that specification is complete. The number of cases in use is also populated. Compare the number of cases here to the sample size in your dataset to ensure there are no missing values and all cases are used. The model can be now be run using IGLS by clicking the "**Start**" button. To view the estimates, click the "**Estimates**" button twice. Doing so will allow you to view the full algebraic specification of the model.

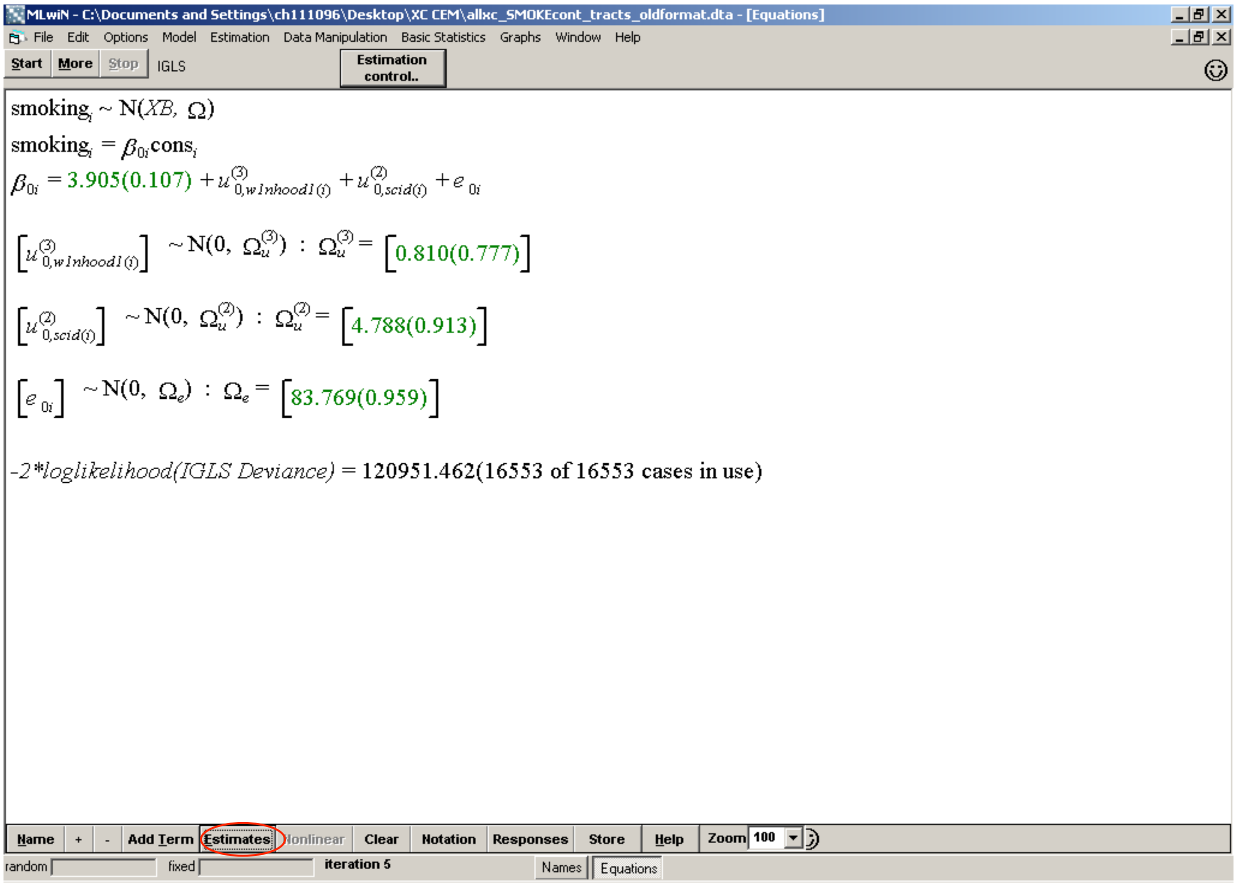


After clicking the "**Estimates**" button once, the specification of the fixed and random parts of the model can be viewed.

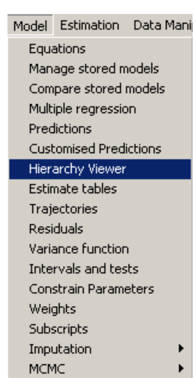


Note the subscripts are uniformly *i* above, as we had indicated earlier.

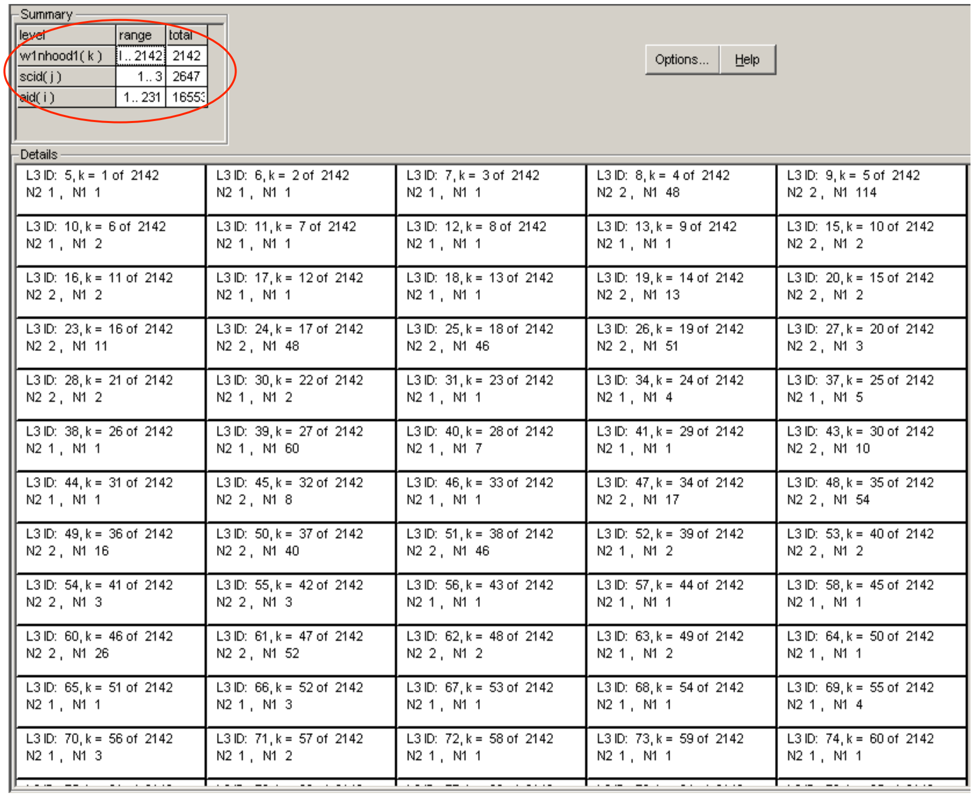
Clicking the "**Estimates**" button again, we can see the converged estimates, specifically the parameter estimate, (standard error), the variance components for each level in brackets, along with the standard error in parentheses. **Note**: The model is currently assuming a hierarchical structure *not* cross-classified. The estimates should not be interpreted. It is necessary to obtain these estimates as starting values for the cross-classified structure.



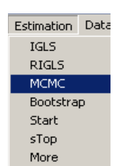
To check the hierarchical structure, or the nesting structure of the levels, go to the *Hierarchy Viewer* on the **Model** menu.



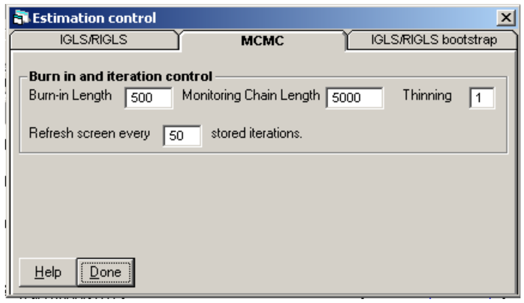
Here we see the nested structure of 16,553 students (aid) in 2,647 schools (scid) in 2,142 neighborhoods (w1nhood). As defined in the non-cross classified hierarhical model, the nesting structure is treating different combinations of school and neighborhood as though they are unique schools (e.g, if a single school has individuals from 2 different neighborhoods, this hierarchical nested model would count schools as 2 separate schools). This is not what you want for a cross-classified model.



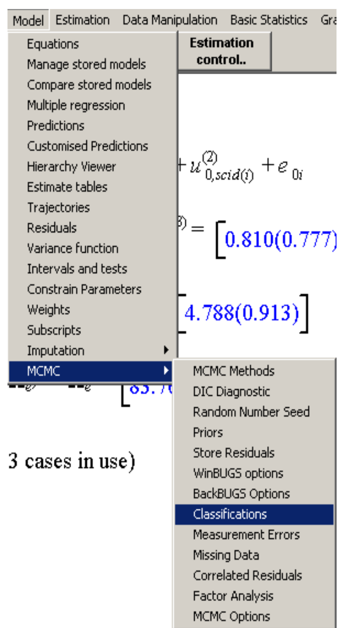
To treat the levels as cross-classified instead of nested, the model needs to be re-run using MCMC reestimation which can be selected from the **Estimation** menu.



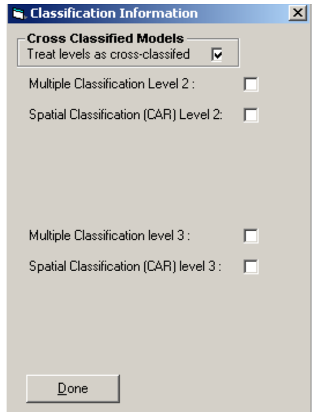
The resampling strategy (e.g. burn in length and number of iterations of the respamples MCMC is taking) can be changed in the Estimation control menu in the MCMC tab. After making any changes, click the "**Done**" button.



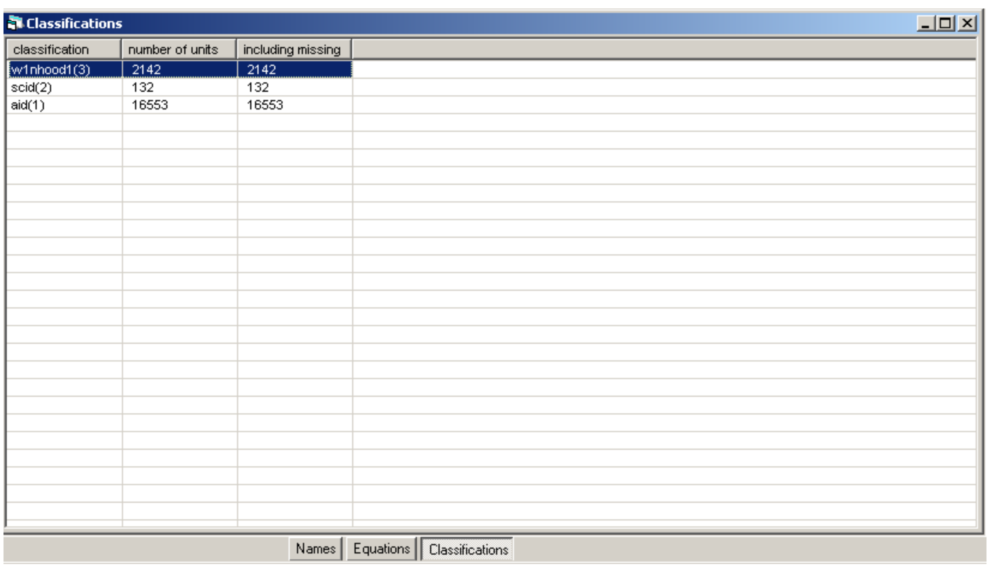
In order to treat neighborhood and school levels as cross-classified instead of hierarchically nested, we need to change the structure in Classifications Information which can be accessed by choosing *Classifications* from the **Model** menu.



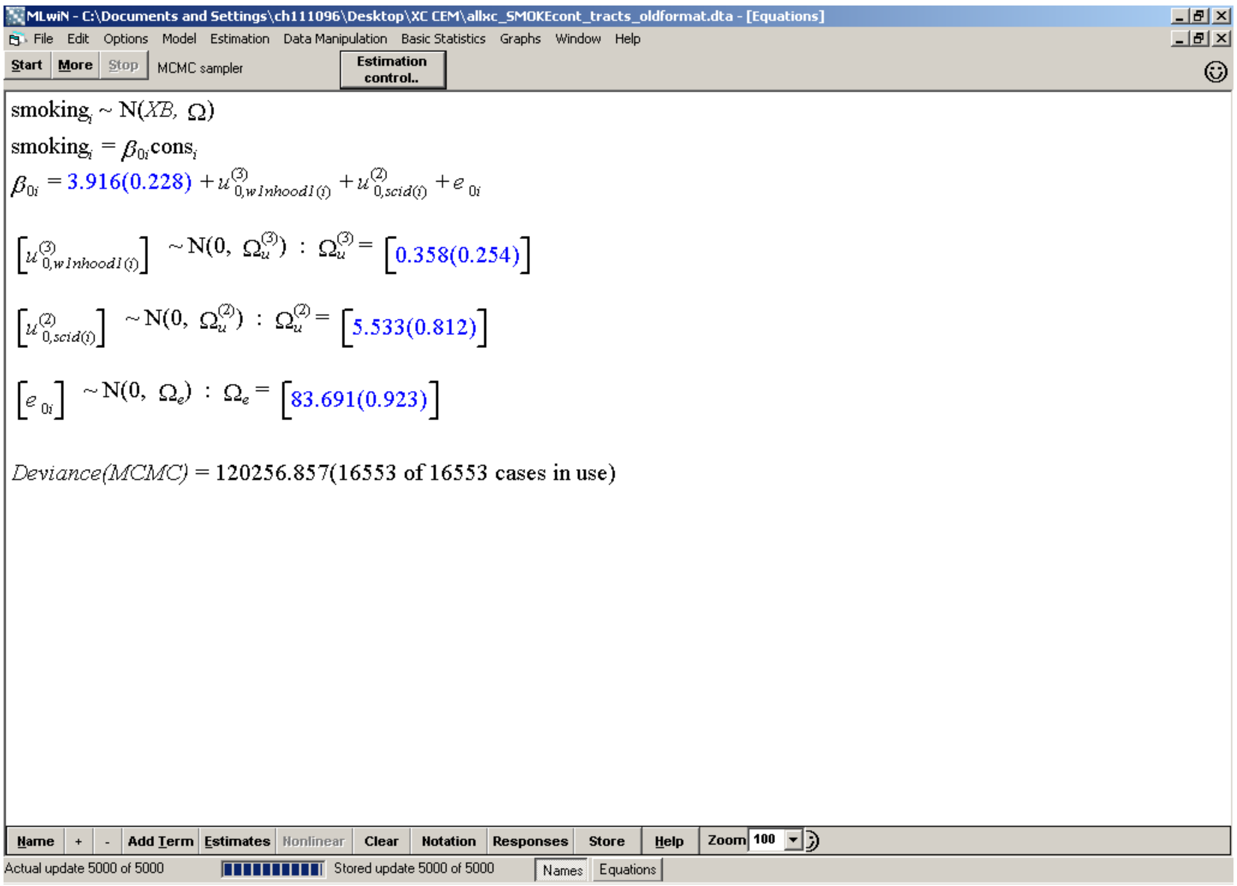
Check the box "Treat levels as cross-classified" in the Classification Information window to run the model with a cross-classified structure.



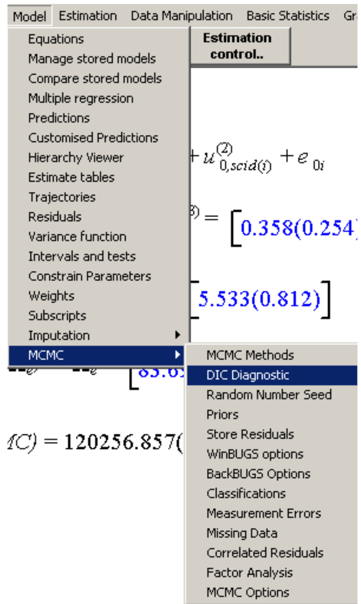
Looking at the Hierarchy Viewer again, we see there are now 16,553 students nested in 132 schools from 2,142 neighborhoods. The number of schools is reduced by treating schools and neighborhoods as cross-classified levels.



Now the model can be run with a cross-classified structure using MCMC re-estimation by clicking the "**Start**" button again. The MCMC re-estimation may take a few minutes depending on the sample size and extent of cross-classification in your data. The estimates are changed from the ones we computed using IGLS. In particular, you could notice a change in the variance components. With the model correctly specified and the cross-classification of the levels accounted for, these estimates are fully interpretable. However, the Deviance statistic reported is not interpretable for these models and should be ignored. This is because the Deviance statistic does not utilize the information from the MCMC re-sampling. The Deviance Information Criterion or DIC is a better diagnostic of model fit in cross-classified models which is discussed on the next page.



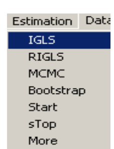
The Deviance Information Criterion (DIC) can be used to assess model fit in cross-classified models; the Deviance, reported in the Equations window, should be ignored when examining the cross-classified model. To obtain this statistic, choose "DIC Diagnostic" from the *MCMC* drop-down menu in the **Model** menu.



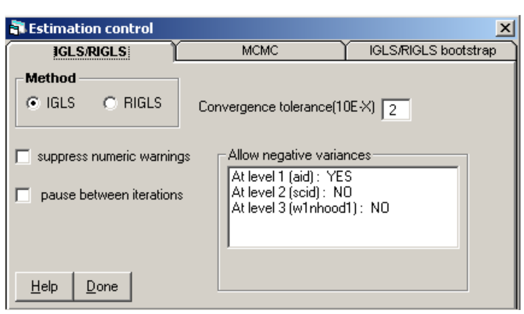
Here the DIC is 120419.9. We use this DIC to compare the fit across cross-classified models. A lower value of the DIC indicates better model fit. We will come back to this value later to compare the null random-intercepts cross-classified model to a random-intercepts model with more predictors.

```
->BDIC
Bayesian Deviance Information Criterion (DIC)
  Dbar     D(thetabar)     pD       DIC
120256.86   120093.83   163.03  120419.88
```

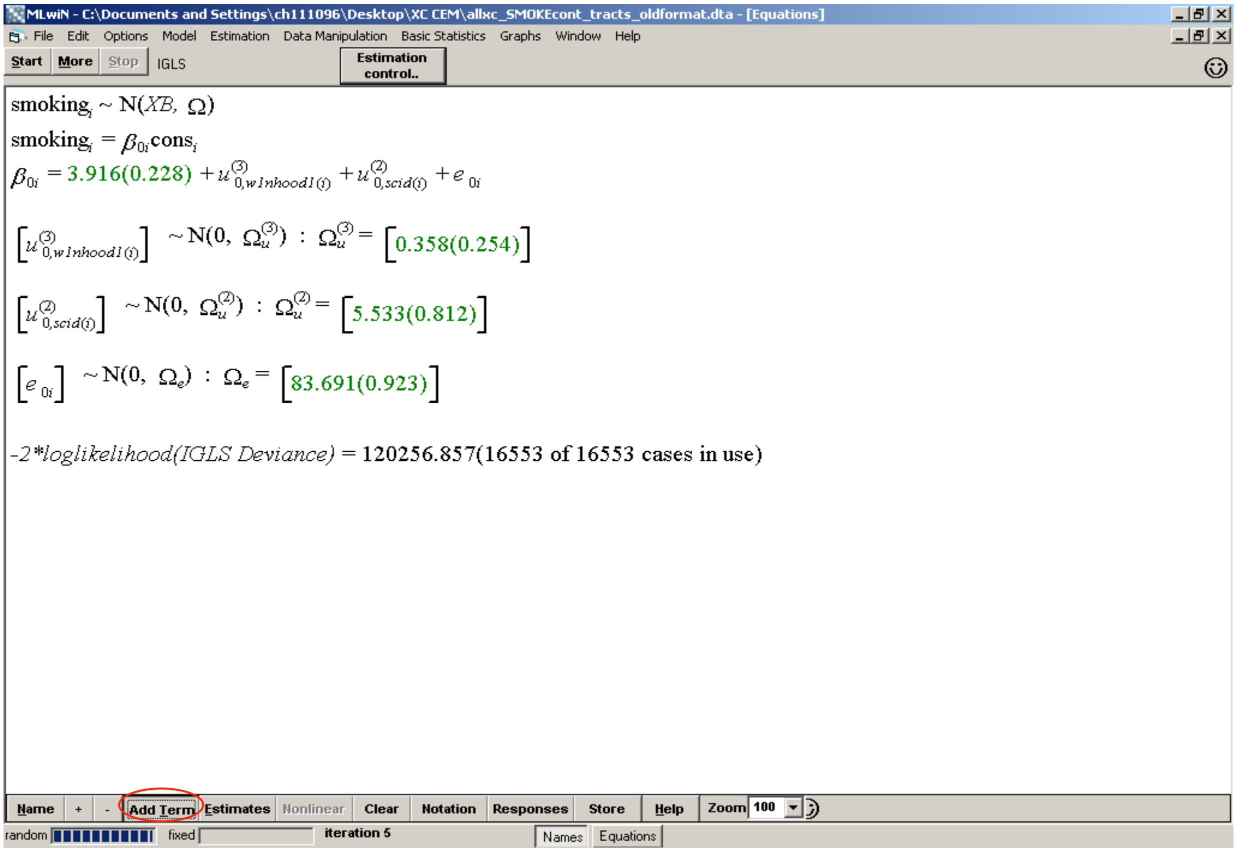
To add more predictors to the model, we need to switch the estimation back to IGLS from MCMC by selecting *IGLS* from the **Estimation** menu.



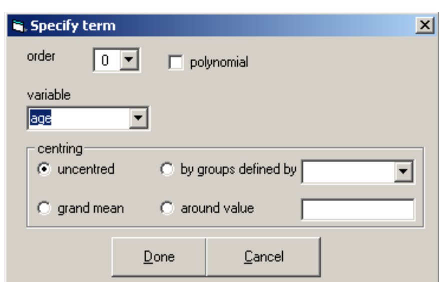
Click the "**Done**" button on the Estimation control window to use the default options in IGLS.



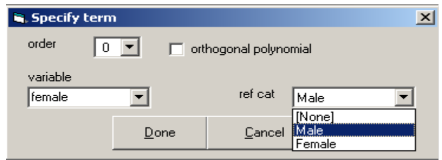
Now more predictors can be added to the model by clicking the "**Add Term**" button on the bottom toolbar.



Choose the variable name from the 'variable' drop-down menu in the Specify term window to indicate the term you want to add to the model. Depending on whether the variable is continuous or categorical, options to center the variable or choose the reference category will appear. With a continuous variable, choose whether values should be uncentered or centered, and if so, around what value. Click the "Done" button when finished choosing options for the variable. Here uncentered *age* is added to the model.

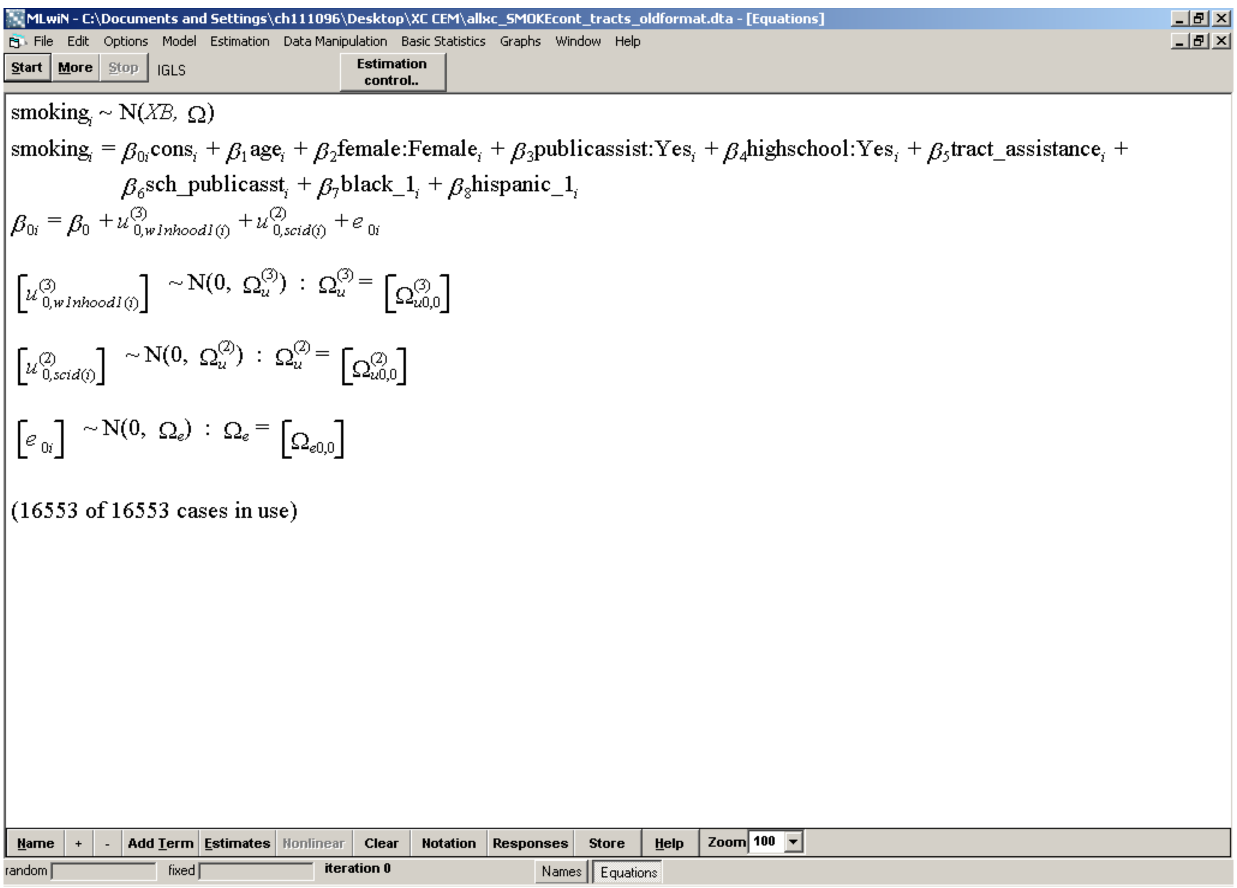


The reference category can be specified for categorical variables. Here the variable for sex called *female* is added to the model with males as the reference category.

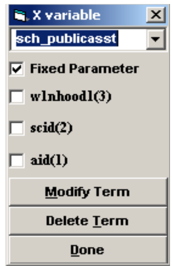


Once all variables have been added to the model, each variable can be modeled as a fixed effect or random effect, which would allow the slopes to vary at different levels. The default is to treat all predictors as fixed parameters. By double-clicking on the variable name, a random slope can be fitted for the variable(s). If two variables with the same formatting/labeling (e.g. "Yes"/"No") are added to a model, an error message stating "Category name clash, using extended names" will appear. This can be fixed by using a more descriptive name for one of the categories.

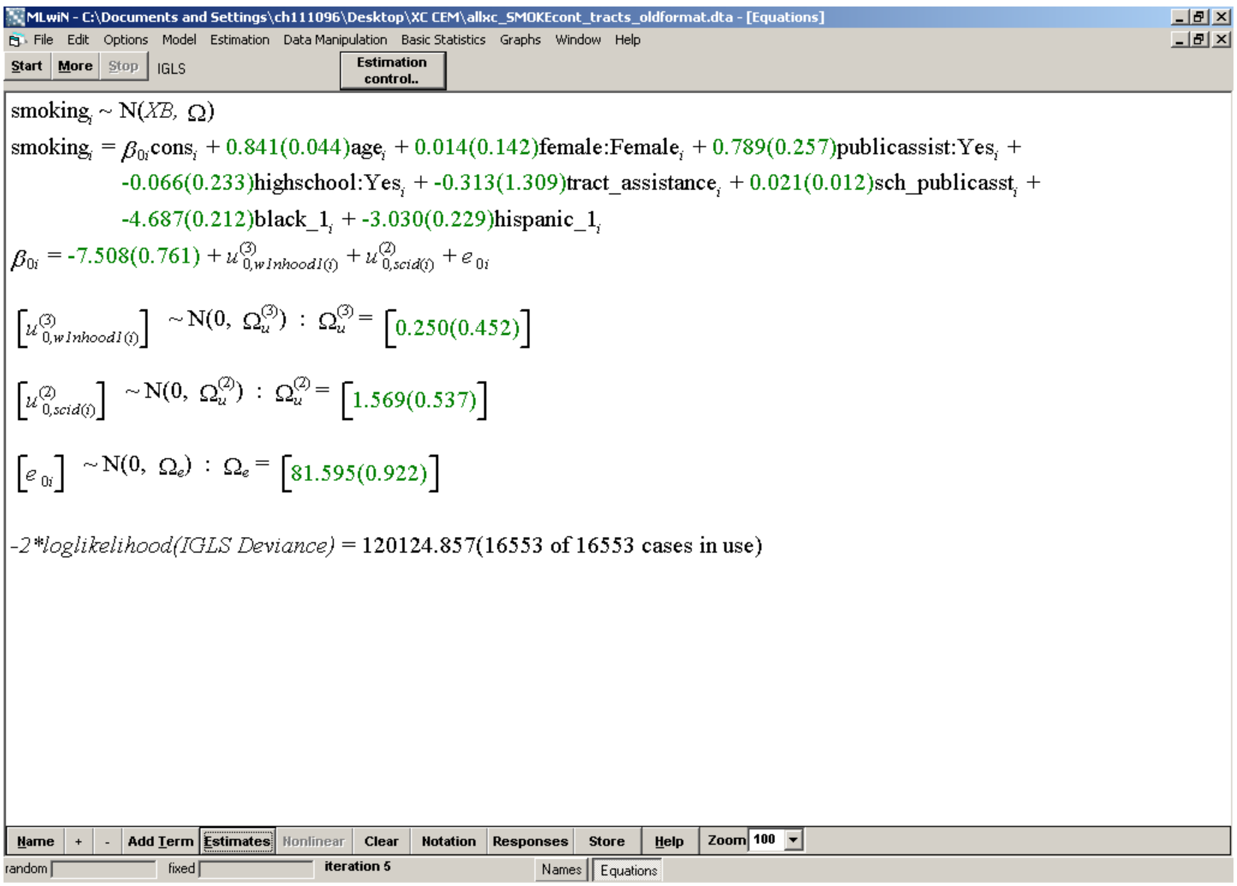
By clicking "**Ok,**" the variable will be added to the model. The full name of the variable will appear. For categorical variables, the full variable name of the variable will appear, followed by a colon and the non-referent category. In this example the variables *publicassist* and *highschool* were both coded "Yes" and "No" with "No" as the reference category.



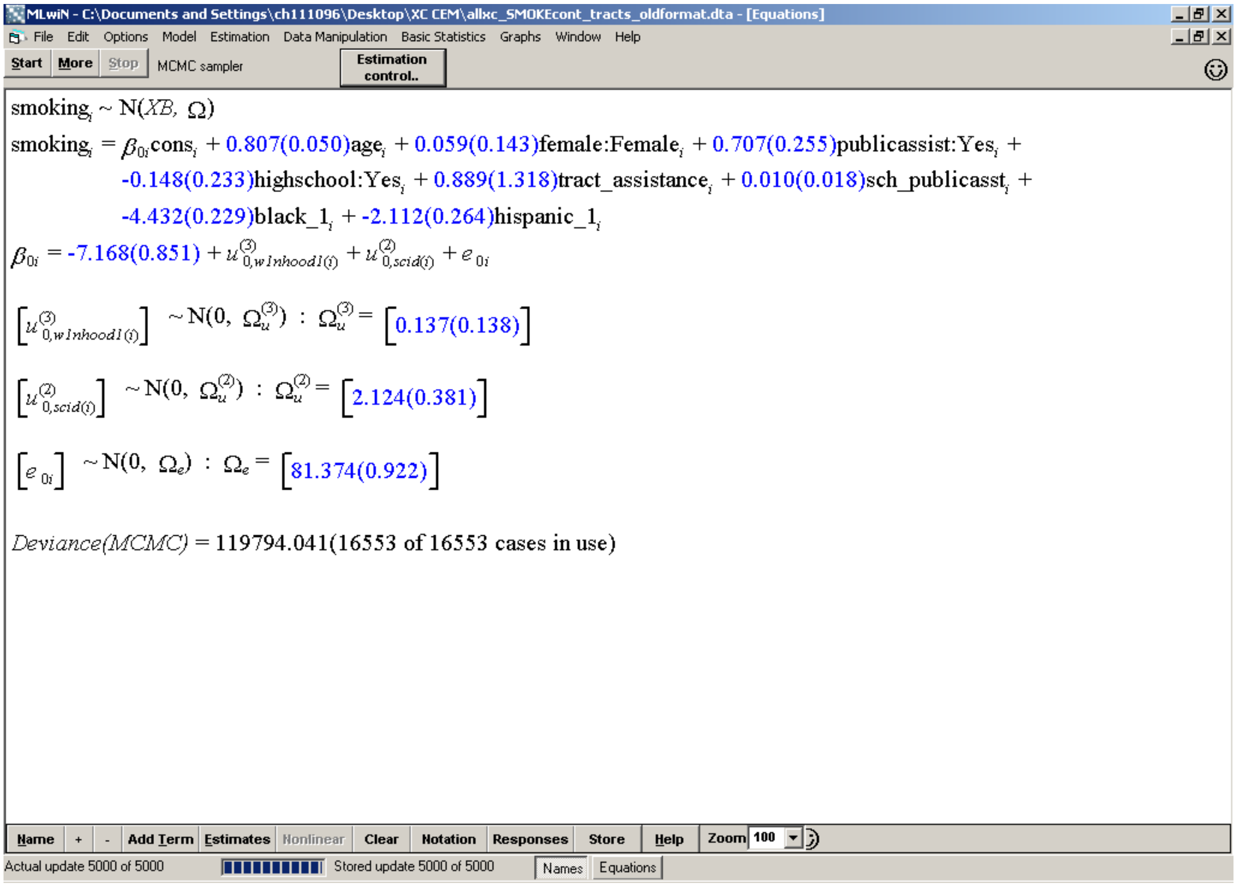
To fit a random slope for a specific level, check the box next to that level identifier. In these models, we are only modeling a random intercept, not random slopes.



Once you are finished adding variables to the model, fit the model in IGLS to get the estimates assuming a nested structure.



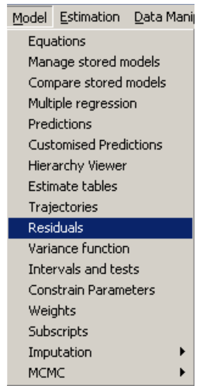
Run the model again using MCMC re-estimation to fit the model with a cross-classified structure with the starting values obtained from IGLS which assume a nested structure. As noted previously, ignore the Deviance statistic and instead use the DIC, which can be accessed from the **Model** menu, under *MCMC*, and "DIC").



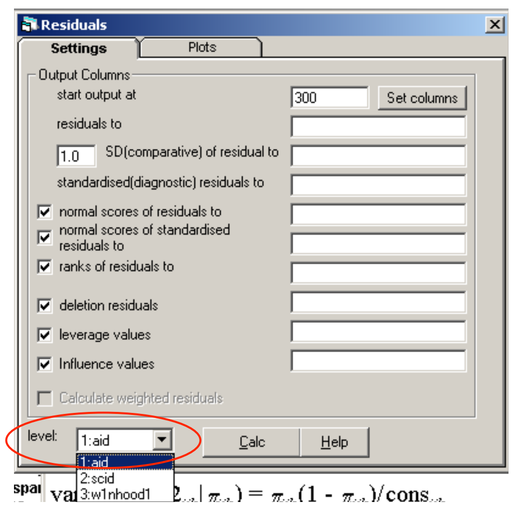
The DIC for the full model with all predictors is less than the null model. This indicates that the full model is a better fitting model.

```
->BDIC
Bayesian Deviance Information Criterion (DIC)
  Dbar      D(thetabar)     pD      DIC
119794.07   119670.46    123.60  119917.67
```

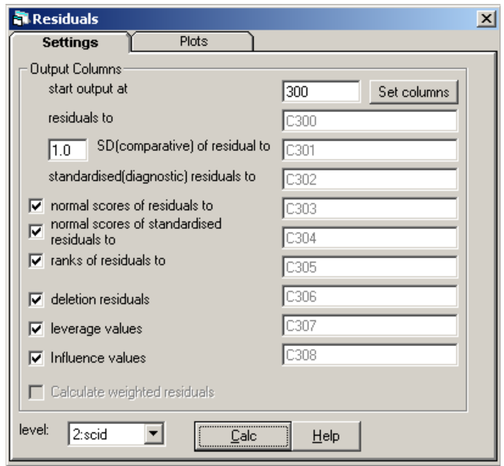
To check the assumptions of the linear model, we can look at the residuals which can be accessed from the **Model** menu under *Residuals*.



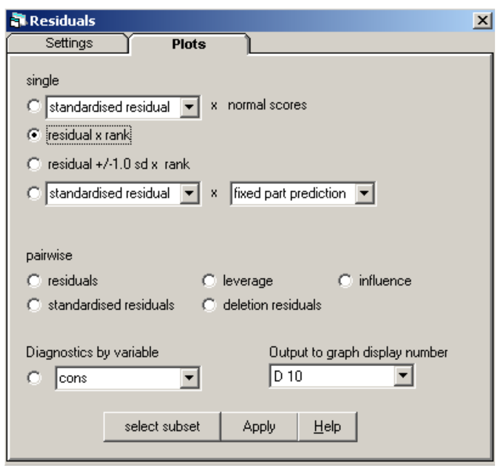
This will open the Residuals window. Residuals can be calculated and plotted for each level to look for issues with overall model fit as well as outliers and influential observations. The level on which to calculate residuals must be chosen first from the drop-down menu on the Settings tab.



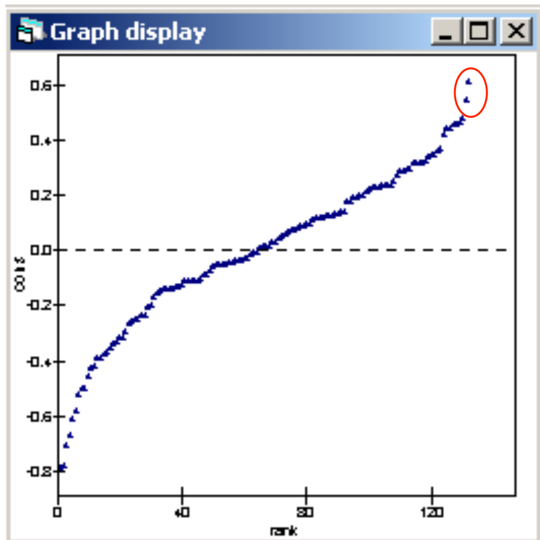
Here the residuals at the school-level are calculated by choosing the school identifier and clicking the "**Calc**" button.



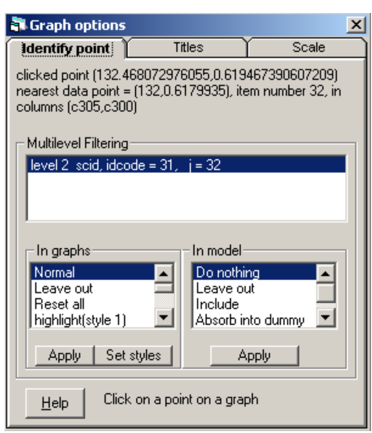
Once the residuals have been calculated at the desired level, different plots may be selected from the Plots tab. Here we will plot the school residuals by their rank on the x-axis. This will produce a plot with the smallest residuals on the left up to the largest residuals on the right. Once the desired plot is selected, click "**Apply**."



The following residual plot is produced for the school level Each triangle represents a different school. The residual plot indicates good model fit if there is an overall linear trend with residuals close to zero (indicating perfect correlation between model-predicted outcome and observed outcome for a given school). A slight s-shaped curve toward the minimum and maxium is common as there is usually sparse data. Observations with large residuals may be concerning. Here we may be concerned about the two outlying schools with large positive residuals indicating that the observed average smoking days for these schools were greater than what was predicted by the model. By double-clicking on one of these extreme observations, we can determine which school it represents and highlight it in the graph or remove it from the model altogether.



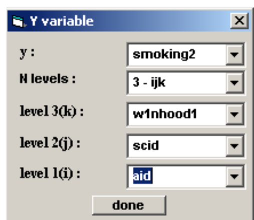
Here we see that the largest positive residual comes from school 31 (SCID=31) which is the 32nd school in the model (j=32 because SCID starts at 0). By making changes to the "In graphs" inset, we can leave this school in the residual plot, leave it out entirely, or highlight it using a different color if desired. Additionally, we could choose to leave this outlying school out of the model or model this school as a dummy variable to try to capture the excess variability contributed by this school. For more information on model diagnostics and residuals in MLwiN, see Chapter 15 of the manual for general information on diagnostics (Rasbash, et al. 2012), and Chapter 15 of the MCMC manual for information specific to cross-classified models (Browne, 2012).



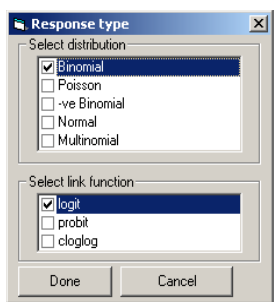
Specifying the Model: Example with a Binary Outcome

Fitting a model with a binary outcome is very similar to fitting a model with a continuous outcome in MLwiN. The major difference is that the distribution function will change. There are a few minor differences between fitting the binary model that are highlighted here. Specifying the dependent variable, running the model, and adding predictors are practically identical to fitting a model with a continuous outcome. As before, the data must be sorted by level using the *Sort* function (Click Remove All before starting the Sort) and any categorical variables should be classified as such using the "**Toggle Categorical**" button in the Names window. Additionally, a constant variable (vector of 1's for all observations) is also necessary when fitting a binary model and random intercepts. **Note**: This appendix provides instructions for modeling binary data. For modeling proportions at a higher level, see the MLwiN User Manual.

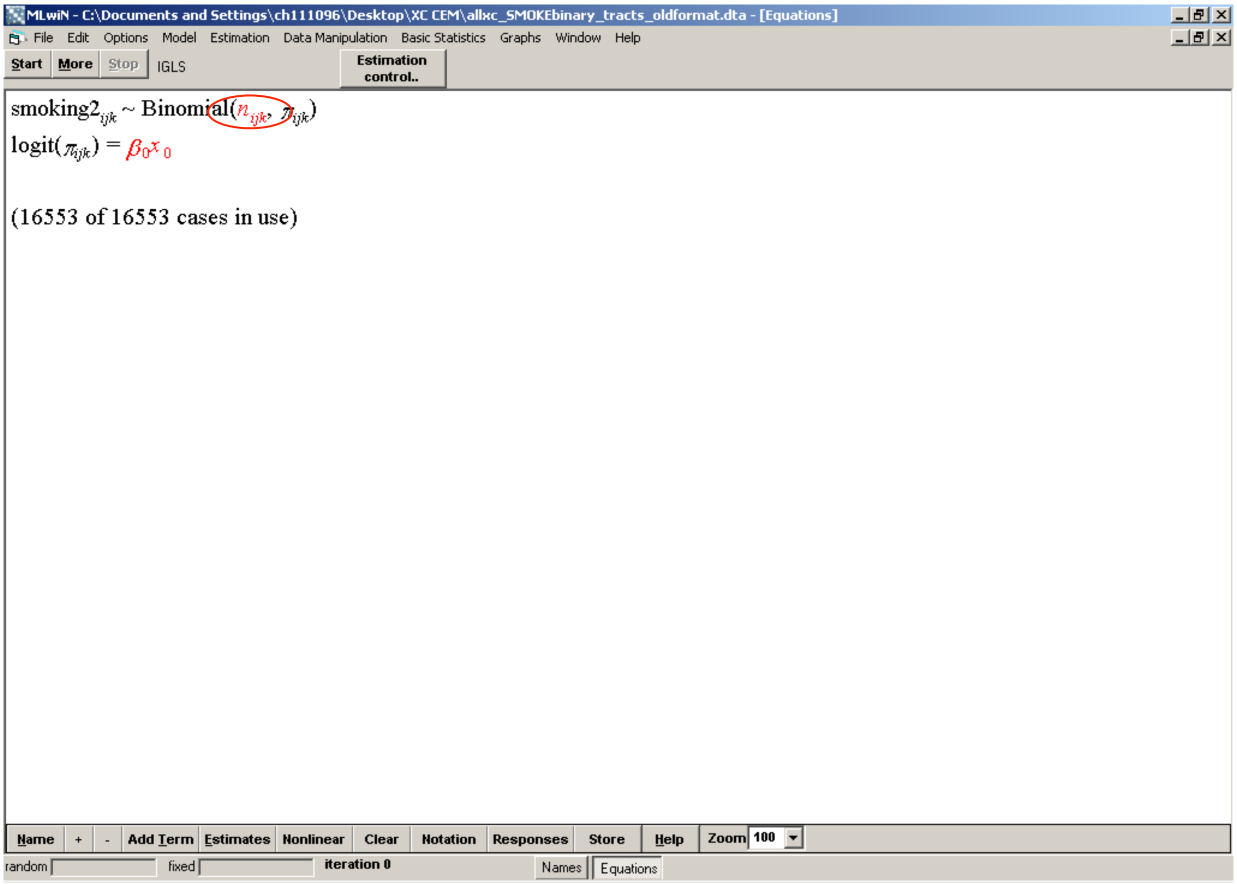
The model with a binary outcome can again be specified from the Equations window (if there is an existing equation you can click on terms; a menu will then appear, allowing you to 'delete term'. The variable *smoking2* (0=non-smoker, 1=smoker) is specified as the dependent variable for the binary model. The number of levels is chosen and the level identifiers are selected from the drop-down menus.



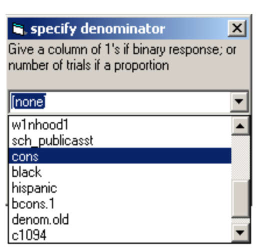
The dependent variable distribution can be changed from the Response Type window by double-clicking the *N* in the Equations window. For the binary model, the distribution function is binomial with a logit link function.



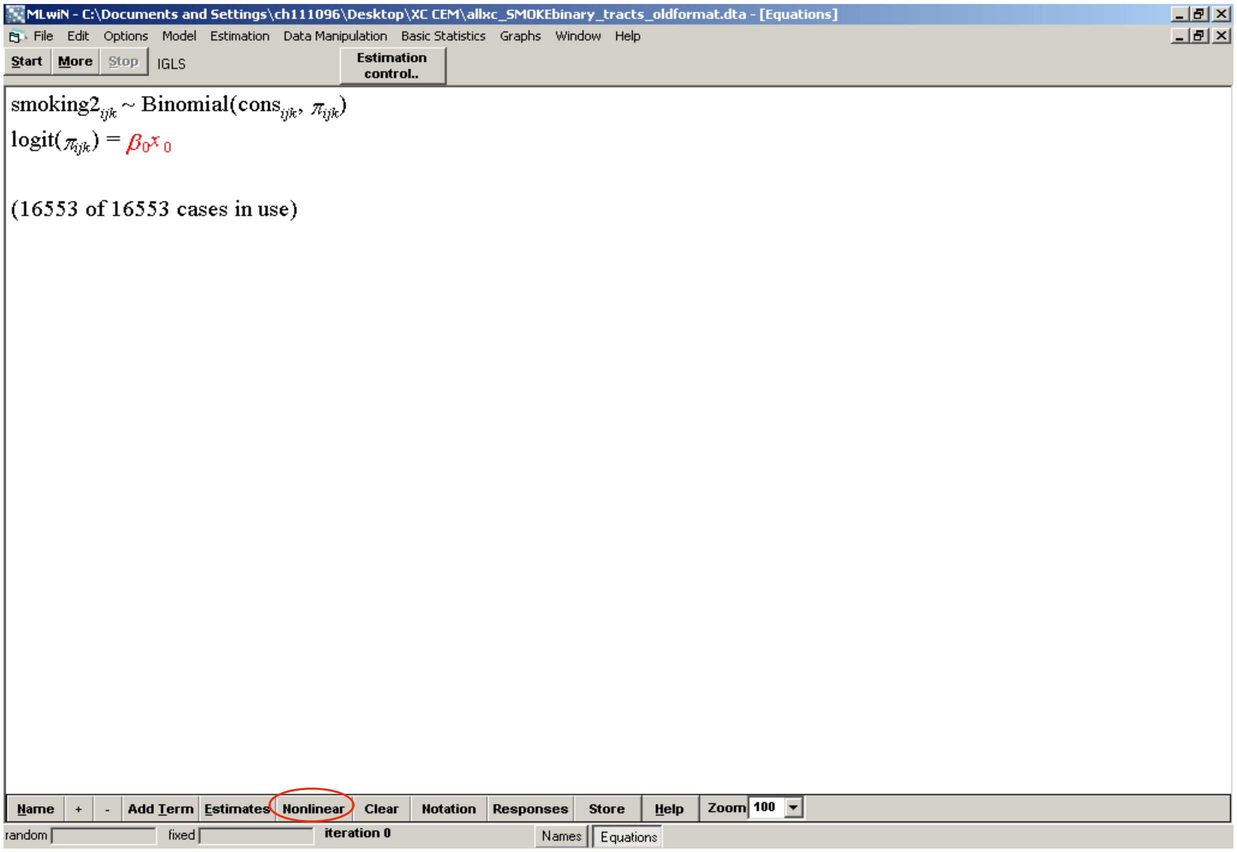
Now instead of the *N* for normal distribution, we see the dependent variable is modeled using a binomial distribution. This indicates that the binary smoking variable follows a binomial distribution with parameters n and π at each of the three levels. The red *n* indicates the denominator, which is a 1 for all observations in the case of binary data (all individuals can take on only one of two values). If the data had been binomial (proportions) the denominator would instead be the total number on which the proportion is based (e.g. modeling proportion of smokers at the school level the denominator would be the total number of students in each school). Double-clicking the red *n* makes the the Specify Denominator window appear.



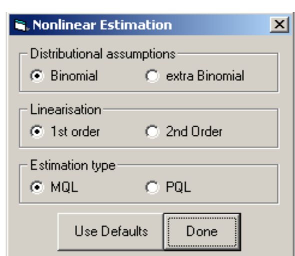
Since we already have a variable which is a 1 for each observation (*cons*), we can specify this variable as the denominator as well.



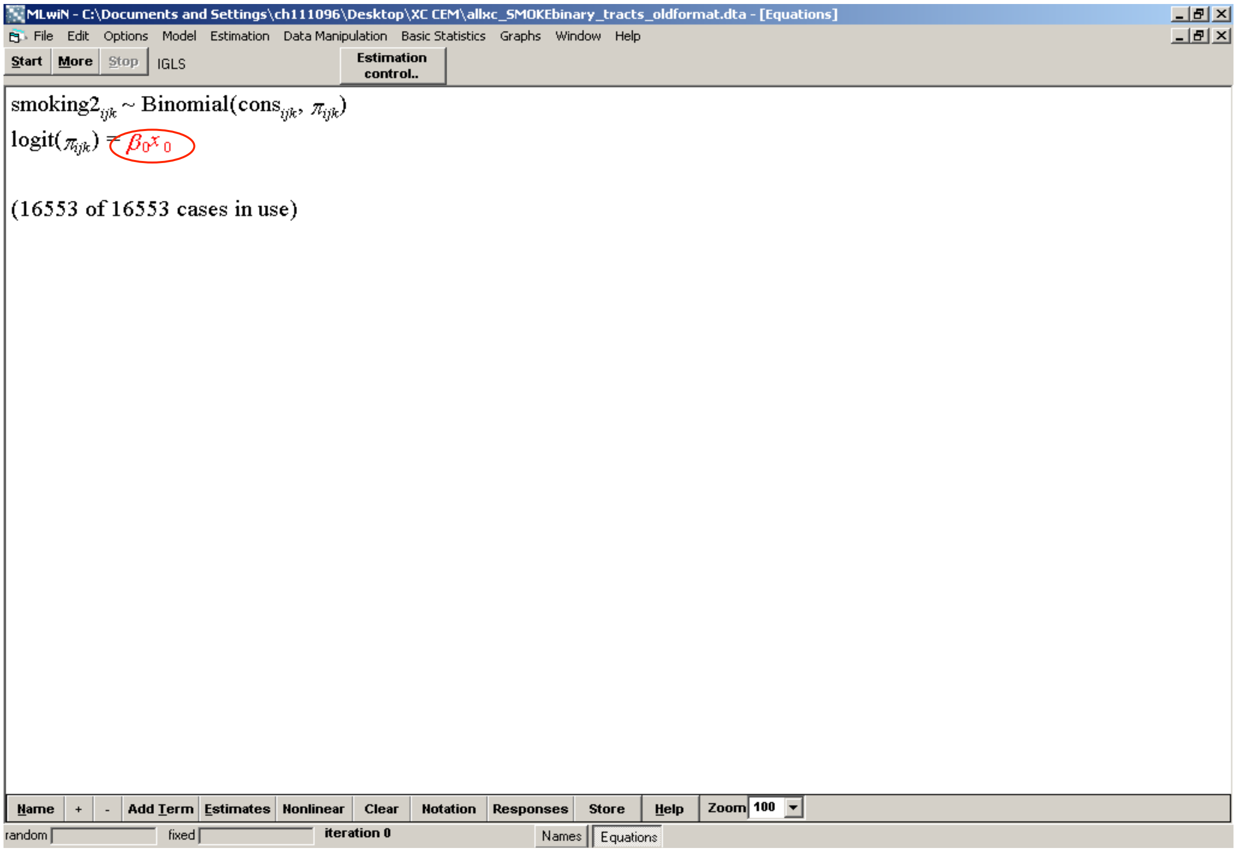
Clicking the “**Nonlinear**” button on the bottom toolbar opens the Nonlinear Estimation window where the default options for assumptions (e.g., binomial, extra binomial), linearization (e.g., first order, second order), and estimation (e.g., MQL, PQL) can be changed.



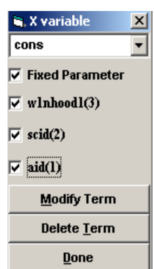
The default distributional assumption is binomial with a 1st order linearization and Marginal Quasi-Likelihood (MQL) estimation. The default settings are sufficient for this model. Once any necessary changes are made, click the “**Done**” button. In this model, we are using the default nonlinear estimation methods. For more information on the other methods, see the MLwiN MCMC manual, Chapter 10 (Browne, 2012).



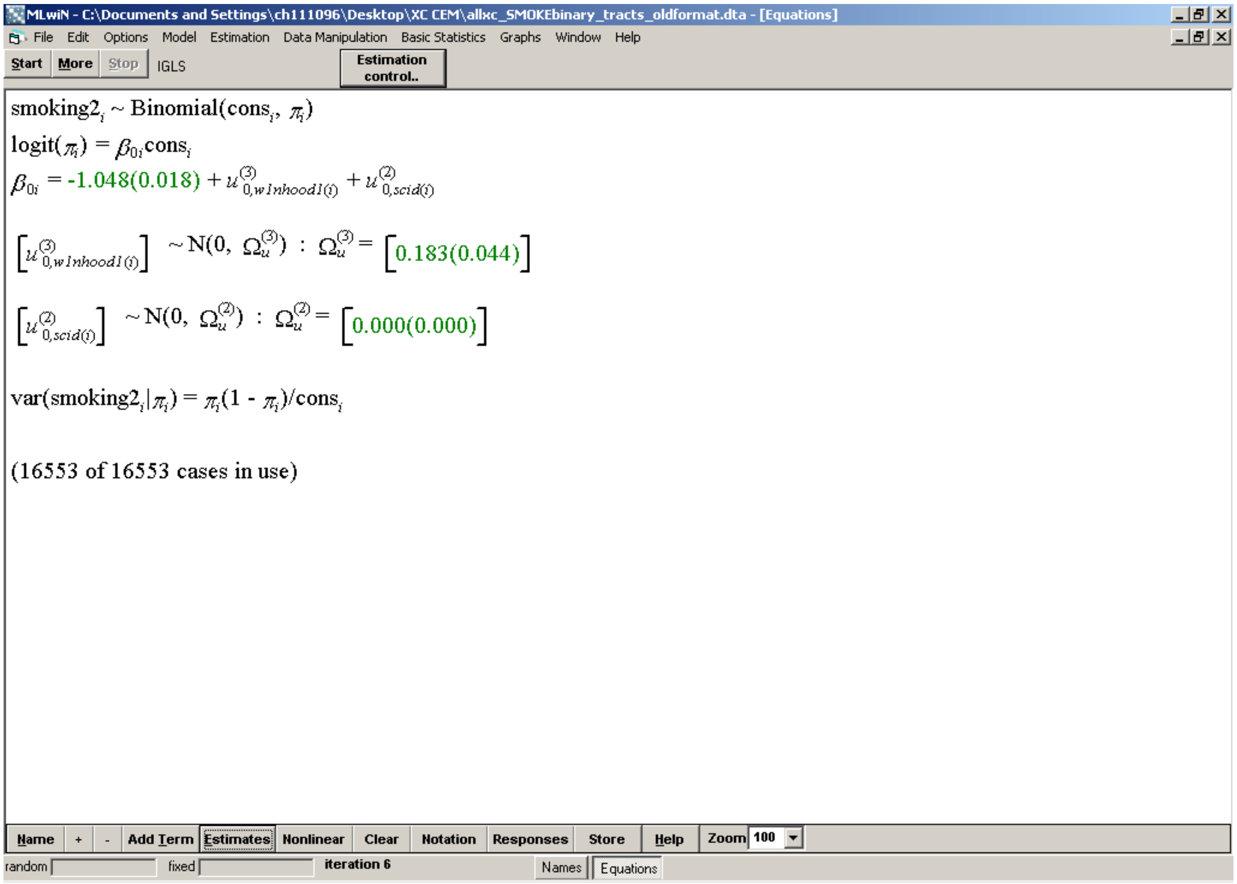
To add a random intercept to the model, double click on the red $\beta_0 x_0$.



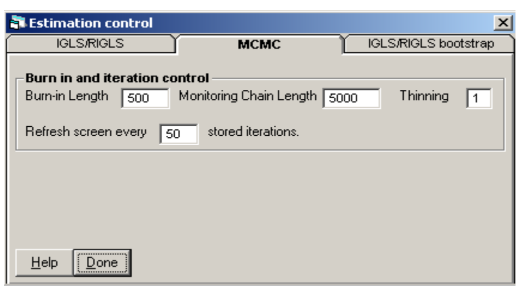
The constant variable is chosen from the drop-down menu as both a fixed effect and a random effect at all three levels. This allows the model to be fit with a random intercept at each level.



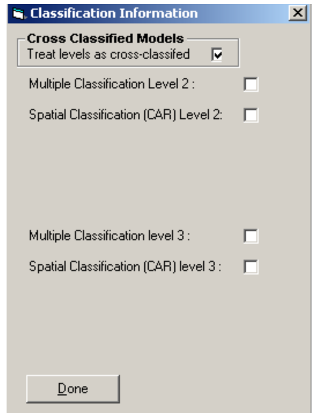
Click the '**Start**" button to fit the model in IGLS. **Note**: Just like with the models estimating a continuous outcome, fitting the model in IGLS assumes a nested structure. These parameter estimates are not interpretable. We will need to refit the model using MCMC re-estimation to fit a cross-classified model.



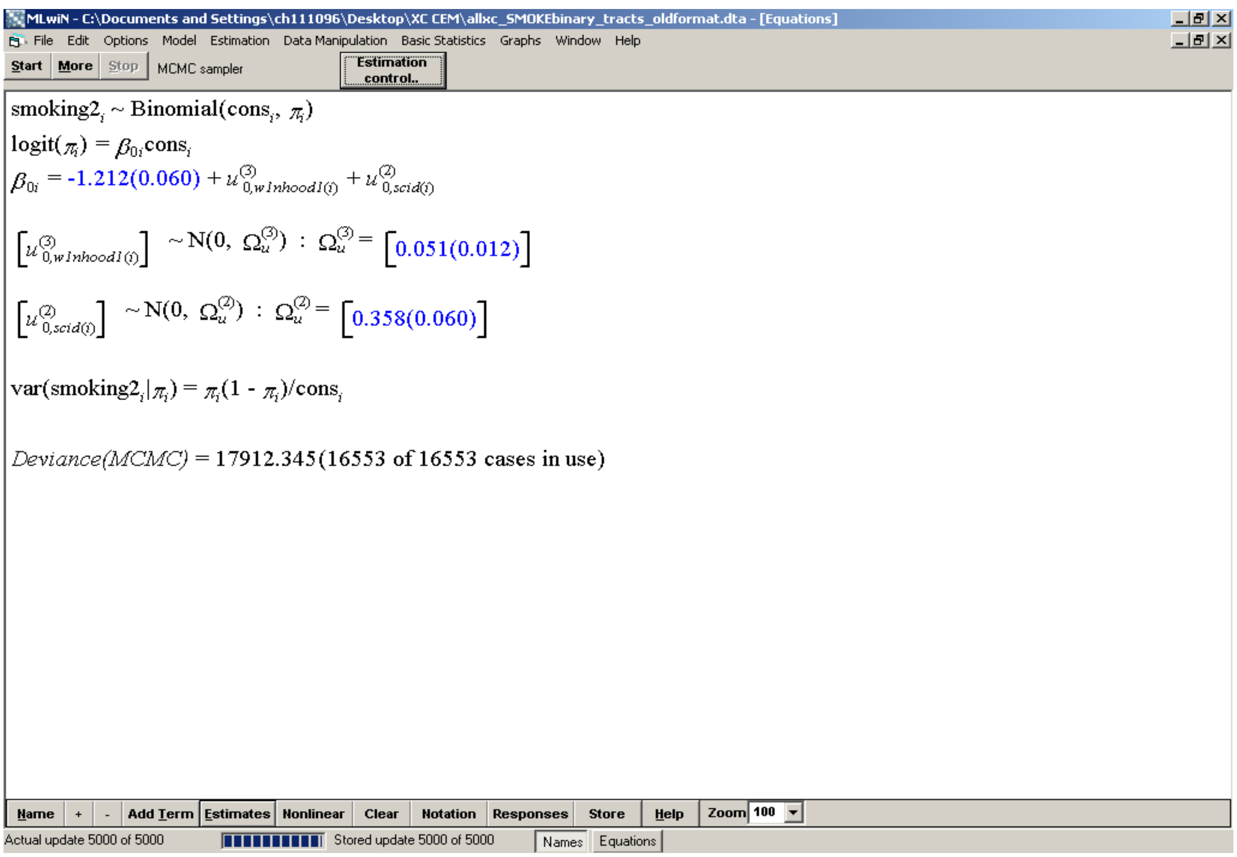
To switch to MCMC estimation, choose *MCMC* from the **Estimation** menu. Make any changes to the burn in length and iteration control in the Estimation Control window and click "**Done**."



Change the model structure to treat the levels as cross-classified by accessing the Classifications Information by choosing *Classifications* from the **Model** menu. Check the box "Treat levels as cross-classified" in the Classification Information window to run the model with a cross-classified structure.



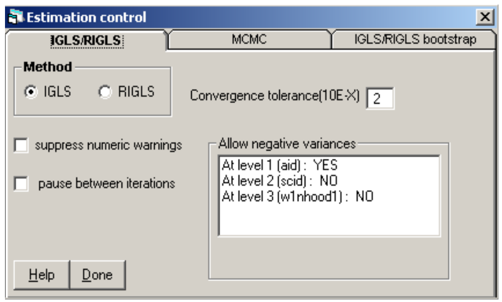
The model can be re-run with the cross-classified structure in MCMC by clicking the "**Start**" button. In a model with a binary outcome, individual-level variance is no longer computed. The estimates given are the parameter estimate (standard error) and variance estimate (standard error) in brackets. The estimates are now interpretable because the levels have been treated as cross-classified. Odds ratios can be computed by exponentiating the parameter estimates. The Deviance statistic is an inadequate measure of model fit for cross-classified models and should be ignored.



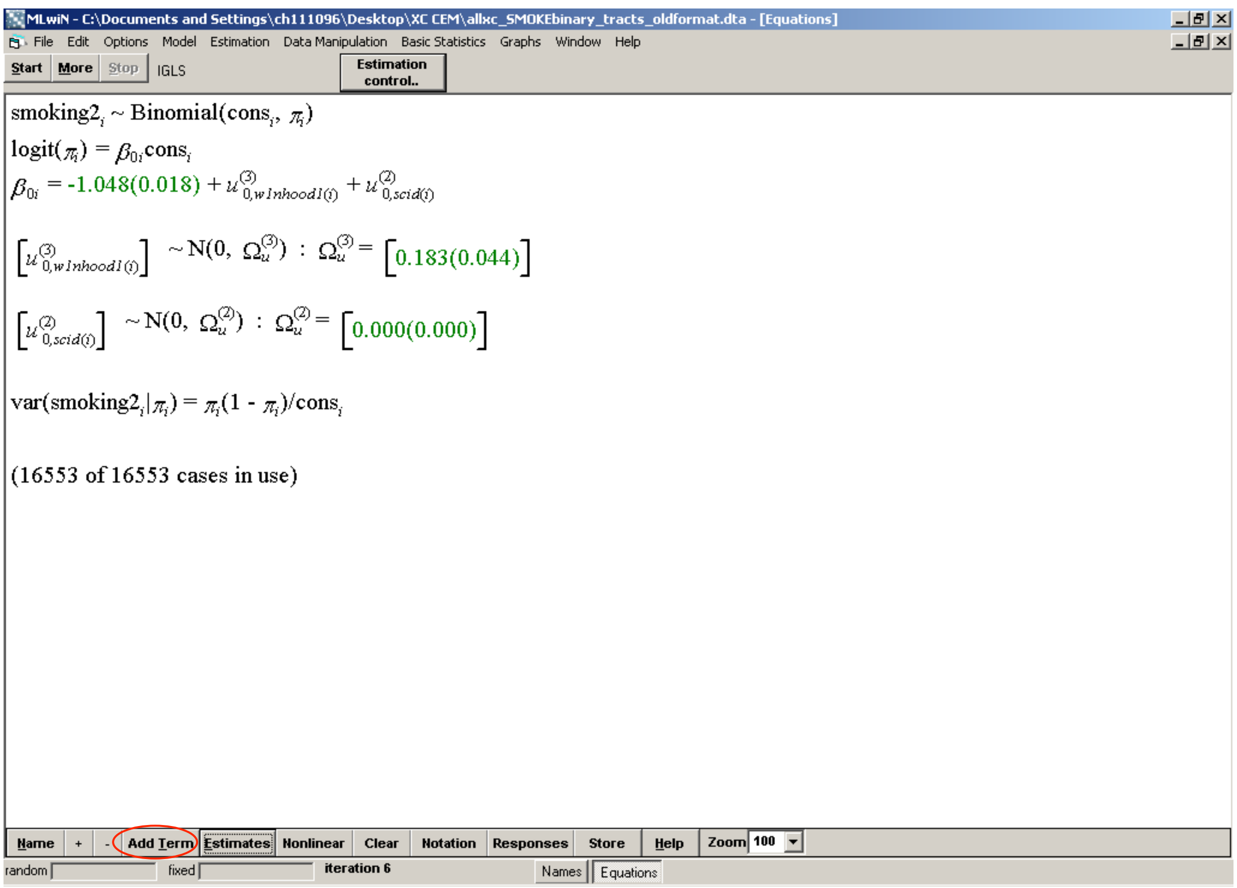
We can request the DIC again from the *MCMC* drop-down menu in the **Model** menu. The DIC for the null random-intercepts binary model is 18111.5.

```
->BDIC
Bayesian Deviance Information Criterion (DIC)
  Dbar    D(thetabar)     pD       DIC
17912.35   17711.23    201.12 18113.46
```

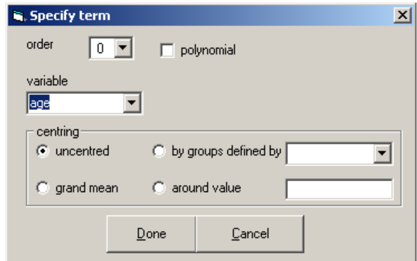
To add more predictors to the model, we again return to IGLS by selecting *IGLS* from the **Estimation** menu. This will make the Estimation Control window appear. You can then make changes to the estimation. To keep defaults, click the "**Done**" button.



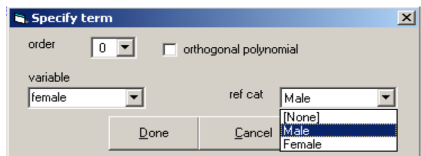
Click the "**Add Term**" button in the bottom toolbar to specify variables to add to the model.



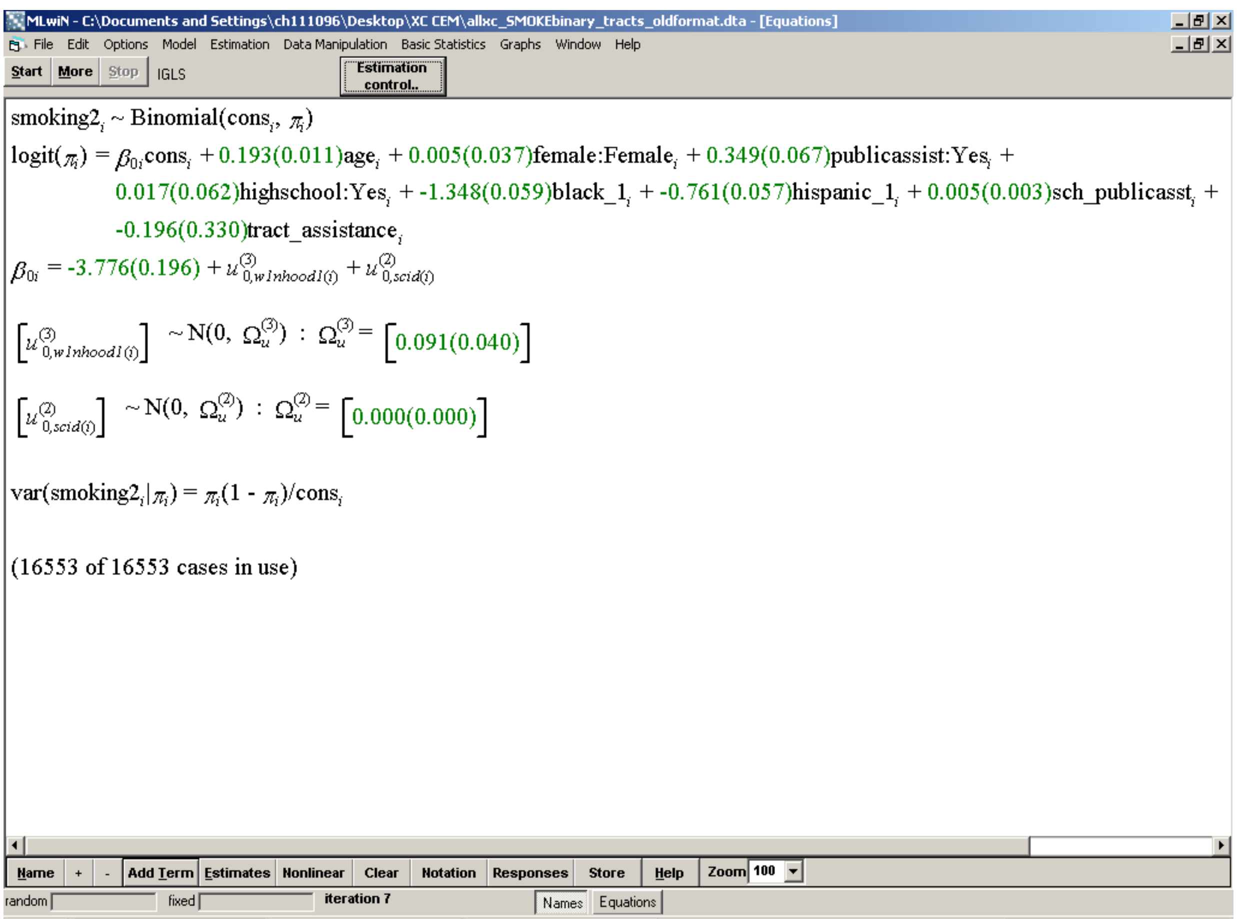
Choose the variable to add from the drop-down menu and if continuous, whether to center and around which value.



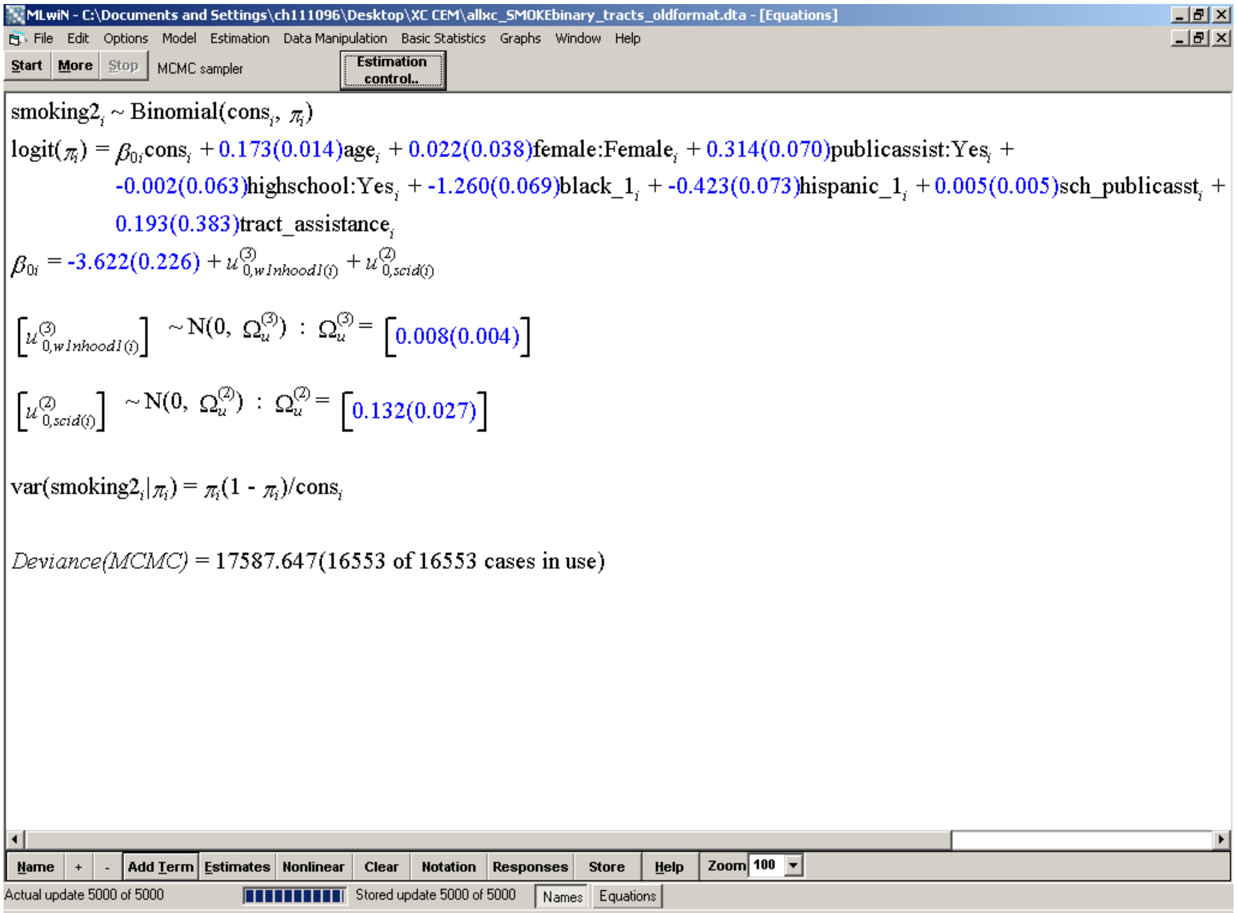
If the predictor being added to the model is categorical, choose the reference category.



Fit the predictors as either fixed effects or random effects by double-clicking on each variable name and choosing the levels at which to model random slopes. The default is to treat all variables as fixed effects. Once any desired changes have been made, run the model by clicking "**Start**."



To run the full model treating the levels as cross-classified, choose *MCMC* from the **Estimation** menu. Make any changes to the burn in length and iteration control in the Estimation Control window and click “**Done.**” Then click the “**Start**” button on the Equations window to re-estimate the model.



The DIC for the full binary model is lower than the null model indicating that the full model is a better fit.

```
->BDIC
Bayesian Deviance Information Criterion (DIC)
  Dbar    D(thetabar)     pD       DIC
17587.65   17474.16    113.48 17701.13
```

References

Software

Browne, W.J. (2009) MCMC Estimation in MLwiN v2.1. Centre for Multilevel Modelling, University of Bristol.

Rasbash, J., Charlton, C., Browne, W.J., Healy, M. and Cameron, B. (2012) *MLwiN Version 2.26*. Centre for Multilevel Modelling, University of Bristol.

Manuals

Browne, W.J. (2012) *MCMC Estimation in MLwiN, v2.26*. Centre for Multilevel Modelling, University of Bristol.

Rasbash, J., Steele, F., Browne, W.J. and Goldstein, H. (2012) *A User's Guide to MLwiN, v2.26*. Centre for Multilevel Modelling, University of Bristol.

Subramanian, S. V. and Jones, K. (2010) *Multilevel Statistical Models: Concepts and Applications*. Boston, MA: Harvard School of Public Health/Bristol, UK: Centre for Multileve Modelling, University of Bristol.