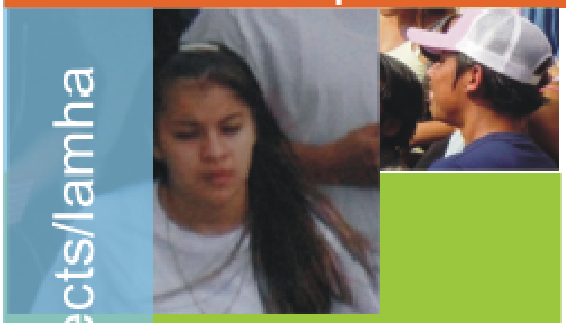
LAMHA

Latino Adolescent Migration, Health, and Adaptation



http://www.cpc.unc.edu/projects/lamha

User's Guide II: Sample Design and Weights

Kim Chantala
Carolina Population Center

Krista M. Perreira
Department of Public Policy

User's Guide II: Sampling Design and Weights

Kim Chantala

Krista M. Perreira

This report was made possible by generous financial support from the William T. Grant Foundation. Additional support was provided by the Z. Smith Reynolds Foundation.

Cover © Tom Swasey

The Carolina Population Center is a community of scholars and professionals collaborating on interdisciplinary research and methods that advance understanding of population issues. Authors are listed alphabetically. The views expressed are those of the authors and should not be attributed to the Carolina Population Center or its funders.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contributions of the project manager, Paula Gildner, for her dedication to managing the data collection and data entry process. We would also like to thank Heidi Bonaduce, Zipatly Mendoza, Helen Cole, Lindsay Stooksberry, Sarah Plastino, Emily Vasquez, and Mercedes Wilson for their assistance with data collection. For their assistance with conducting, transcribing, and translating adolescent qualitative interviews, we are especially gratefully to Sandi Chapman and Tina Siragusa. For her assistance with preparing data collection protocols for parent interviews regarding health beliefs, we thank Gaby Livas-Stein. For their assistance in conducting, transcribing, and translating parent interviews regarding health beliefs, we thank Linda Ko and Stacy Bailey. Finally, we would like to thank Kristin Hoeft for her assistance with data cleaning and analysis.

This research study would not have been possible without the tireless support of staff at the Carolina Population Center including Nancy Dole-Runkle, Lori Dwyer, Pan Riggs, Joyce Tabor, Lance Underwood, and Tom Swasey.

We greatly appreciate the willingness of the ten school districts across North Carolina that helped us to access families.

Most importantly we want to acknowledge and deeply thank the new immigrant youth and parents who told us about their lives and experiences.

SETTING

After growing 394% since 1990, the population of North Carolina was 4.7% Latino in 2000 and most (61%) of these Latinos are first-generation immigrants (Census, 2000). The Latino Adolescent Migration, Health, and Adaptation Study (LAMHA) was designed to obtain data on first-generation, Latino immigrant youth aged 12-18 who are living in high-growth (i.e. ≥ 394% growth between 1990 and 2000) Latino communities in North Carolina with a Latino population of at least 5,000.

The focus on Latino adolescents living in large high-growth communities was both practical and substantive. On a practical level, it greatly facilitated data collection efforts and reduced the cost of the study. The size criterion helped to ensure an adequate sample size while reducing travel costs. The growth criterion facilitated the enrollment of first-generation, immigrant youth into the study. On a substantive level, this design focused our study on communities in North Carolina where mental health services for the Latino community are most needed and least developed

Based on data from the Census 2000, 17% of North Carolina's 100 counties qualified as large high-density Latino communities. Fifty-seven percent (N=217,221) of North Carolina's Latino population lived in these 17 counties and 68% were first-generation immigrants. Latinos in these counties were from several different countries. Most (67%) are from Mexico. Other places of origin included Puerto Rico (6%), Cuba (2%), Central America (9%), and South America (4%). Additionally, about 12% of the Hispanic community did not report a country of origin. Individuals not reporting a place of origin were more likely to be Spaniards, 2nd generation Latinos, or 3rd generation Latinos. Because most are not 1st generation immigrants from Latin America, they were unlikely to meet the criterion for inclusion in this study. Adjusting for this, we expected that 76% of our sample of 1st generation immigrant youth will be of Mexican origin.

SELECTION OF SCHOOLS AND STUDENTS

Selection of High School Clusters

To ensure economic variation in the communities in which Latino youth live, high schools serving large high-growth Latino communities were stratified into two groups – urban and rural. Urban high schools were defined as high schools serving counties where over 50% of the population is living inside an urbanized area or urban cluster. Rural high schools were defined as serving counties where 50% or less of the population is living in an urbanized area or urban cluster. High schools in Duplin, Johnston, Randolph, Robeson, Sampson, and Union counties were included in the rural strata. High schools in Alamance, Cabarrus, Catawba, Durham, Forsyth, Gaston, Guilford, Lee, Mecklenburg, Rowan, and Wake were included in the urban strata. One high school from the urban strata and two high schools from the rural strata were selected with a probability proportional to the size of Latino enrollment in each school. Within these schools, Latino students were randomly selected to be included in our sample.

Figure 1 shows the geographical distribution of the counties in which the eligible high schools are located. Figure 2 shows the distribution of the cities in which eligible high schools are located. As can be easily seen from these figures, North Carolina's Latino population is primarily located along three main interstates – I-95, I-85, I-40. These are common migration pathways with high job growth.

According to the North Carolina Department of Public Instruction, there were 144 high schools serving large, high-growth Latino communities in 2000. These schools enrolled a total of 6,760 Latino students with an average Latino student body of 47 students. There were 52 schools with at least 50 Latino students enrolled. Together these schools enrolled 74% (5,011) of the Latino youth in the 144 high schools serving large, high-growth Latino communities.

Because students in schools with fewer than 50 students would have had small probability weights associated with their selection, we restricted our sampling frame to students in high schools with at least 50 Latino students enrolled. This restriction had little effect on the generalizability of the study but greatly reduced its costs. Sampling from schools with small student bodies would have required us to sample from many more schools, increased the travel costs for interviewers, and increased the administrative costs for the school system. With this sampling structure, we expected to obtain a sufficient sample size by selecting 5 of the 52 eligible high schools.

Selection of Middle School Feeders

Each high school selected was asked to approximate the percentage of the entering Latino class coming from each middle or junior high school in their service area. A feeder school was selected from this list with a probability proportional to the percentage of the high school's entering Latino class that came from the feeder. This ensure that we included younger adolescents aged 12-14 in our sample. There were 205

middle schools serving our 17 high-growth communities. With a minimum of 1 and a maximum of 221, there were an average of 40 Latino students per middle school in 2000 (DPI, 2002). As before, small middle schools with fewer than 25 Latino students were excluded from the sampling frame. This reduced the number of eligible schools to 114. In 2000, these 114 schools served 88% of the Latino students in middle school in large, high-growth Latino communities.

This selection of high school clusters with middle school feeders followed the sampling strategy employed in the National Longitudinal Study of Adolescent Health (Add Health) (CPC, 2002a).

Selection of Students

In an initial screening interview/survey, all Latino students in the selected schools were asked if they were born in the U.S. and how old they were when they moved to the U.S. Research suggests that students arriving to the U.S. at ages younger than 5 are much more like second-generation immigrant youth than first-generation youth. Therefore, in an effort to focus on the effects of migration and acculturation experiences among first-generation youth, we excluded students who were not foreign-born and were not at least 5 years old when they moved to the U.S. This ensured that the youth interviewed were 1.25 or 1.5 generation youth. The screening also ensured that the migration experiences of the youth we survey would be relatively recent (i.e. within the past 15 years).

Of those that were determined to be eligible for inclusion in the study, we randomly selected a number from each school that is proportional to the size of the school's Latino study body. Based on experience from Add Health, a similar survey conducted by the Carolina Population Center, we expected a response rate of approximately 80 percent. To yield a final sample size of 250 students, approximately 325 eligible students were expected to be sampled.

Amendments to Sampling Strategy during Data Collection

Student Eligibility: Early in the data collection process, the eligibility criteria were changed to eliminate "age of arrival in US." Therefore, all first-generation Latino youth regardless of age of migration to the U.S. were included in the study. The updated LAMHA criteria for eligibility were:

Eligibility Requirements for study
1. 12 to 18 years of age on October 1, 2004
2. Enrolled in 6th to 12th grade for the 2004-2005 academic year.
3. Student is of Latino Heritage.

Students who had previously been contacted via phone and who were presumed to be ineligible based on their age of arrival were re-contacted. We also contacted those

individuals who returned letters, reported they wanted to be contacted, and were possibly eligible under the updated eligibility criteria.

School Selection: The original sampling design called for the selection of one high school from the urban strata and two high schools from the rural strata. Due to the relatively low numbers of students meeting our eligibility criteria and available to be interviewed, this was amended. Five high schools were selected from our urban strata and six high schools were selected from our rural strata. Among the 14 middle schools, nine were from the rural strata and five were from the rural strata.

Replacement School Selection: Three of the school districts with high schools originally selected for participation refused to participate. These schools were replaced with schools who were of the same strata (i.e. rural or urban) and who most closely matched the originally selected school in terms of: (1) percent urban, (2) proportion of the student body that identified as Latino, and (3) total school size.

Feeder School Selection: The original sampling design called for the selection of only one feeder school for each high school selected. This design was changed to select ALL feeder schools for each participating high school.

Service Use Data

There are no sampling weights for the service use data. The service use data was collected during the second half of data collection only. Therefore, it cannot be considered a probability sample. **The weights provided for use with the adolescent and adult surveys should not be used with the service use data.**

Summary of Schools Sampled

A total of 10 school districts (6 rural districts and 4 urban districts) participated in the study. Across these school districts, 11 high schools and 14 middle schools participated.

CALCULATION OF POPULATION WEIGHTS

This study is designed to be generalized to all Latino adolescents attending high schools or middle schools in communities that have experienced a rapid growth in their Latino populations.

The computation of sample weights follows the same basic approach used in computing Add Health weights (Tourangeau and Shin, 1999). The weights will be calculated in three separate steps.

Step 1: High School selection probabilities. The selection probability of high school i in strata j will be calculated as:

$$P_{ij} = \frac{T_{i,j}}{T_j / t_j}$$

where $T_{i,j}$ is the total number of Latino adolescents attending high school i in strata j, T_j is the total number of Latino adolescents in all high schools in strata j, and t_j is the number of high schools selected from strata j.

Step 2: Feeder School selection probabilities. For each high school, data on the percentage of adolescents from a given feeder school was not available. Therefore, for each High School i selected in strata j, one feeder school will be selected with probability proportional to the number of Latino adolescents enrolled in the feeder school:

$$P_{Fmj} = P_{ij} \frac{Q_{m,j}}{Q_j / q_j} = P_{ij} \frac{Q_{mj}}{Q_j}.$$

Where $Q_{m,j}$ is the total number of Latino adolescents attending Feeder school m in strata j, Q_j is the total number of Latino adolescents in all feeder schools sending adolescents to high school i in strata j, and q_j is the number of high schools selected from strata j. Because only one feeder school (for each high school) will be selected $q_j = 1$.

Step 3: Adolescent sampling weights. The initial adolescent-level sampling weights will be computed as the inverse of the product of the probability the school was selected and the within-school selection probability. We will stratify adolescents by grade and sex, so all adolescents within the same grade-sex cell for a school will have the same initial sample weight. For adolescents from a high school this will be:

$$W_{ijk} = \frac{N_{ijk}}{P_{ij} n_{ijk}}$$

where N_{ijk} is the number of qualifying Latino adolescents in grade-sex cell k for high school i in strata j, and n_{ijk} represents the number of adolescents selected from that group.

For adolescents from a feeder school the initial sampling weight was computed as

$$W_{mjk} = \frac{N_{mjk}}{P_{Fmj} n_{mjk}}.$$

where N_{mjk} is the number of qualifying Latino adolescents in grade-sex cell k for feeder school m in strata j, and n_{mjk} represents the number of adolescents selected from that group.

Because the number of qualifying Latino adolescents (N_{ijk} and N_{mjk}) in each school was not known, it was estimated from the total Latino enrollment in each High school as follows:

$$N_{ijk} = \frac{R_{ijk} E_{ijk}}{T_{ijk}}$$

Where R_{ijk} is the number of Latino adolescents on the roster, E_{ijk} is the number of screened Latino adolescents who met the eligibility requirements, and T_{ijk} is the total number of Latino adolescents who were screened in high school i, strata j, and grade k. The total number of qualifying Latino adolescents on the roster (N_{mjk}) was estimated in the same manner.

Because some grade-sex cells had no interviewed adolescents, we collapsed cells to do these computations. The cells were collapsed within grade first, then across grade. No additional adjustments such as trimming or post-stratification were made to these weights.

FIGURE 1

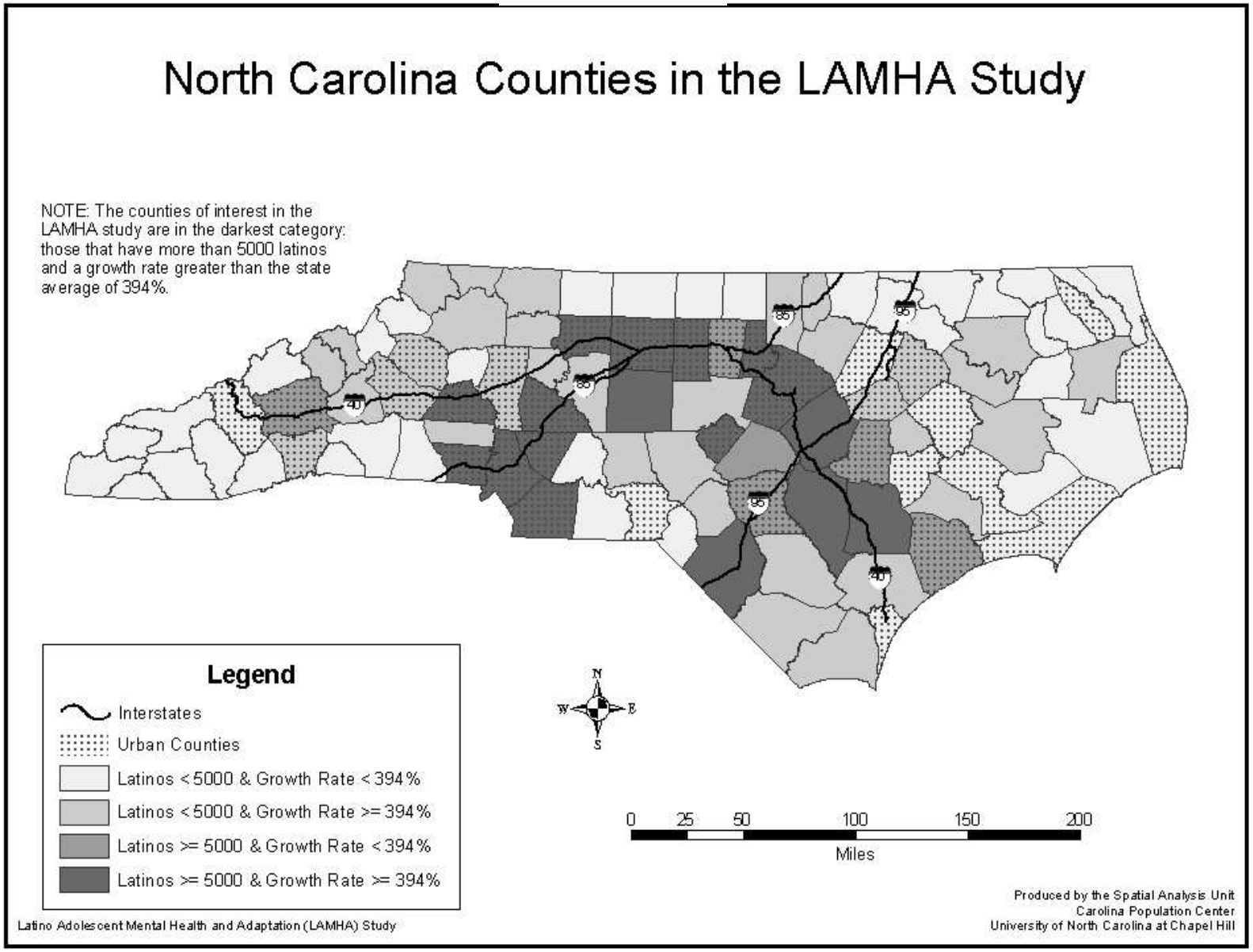
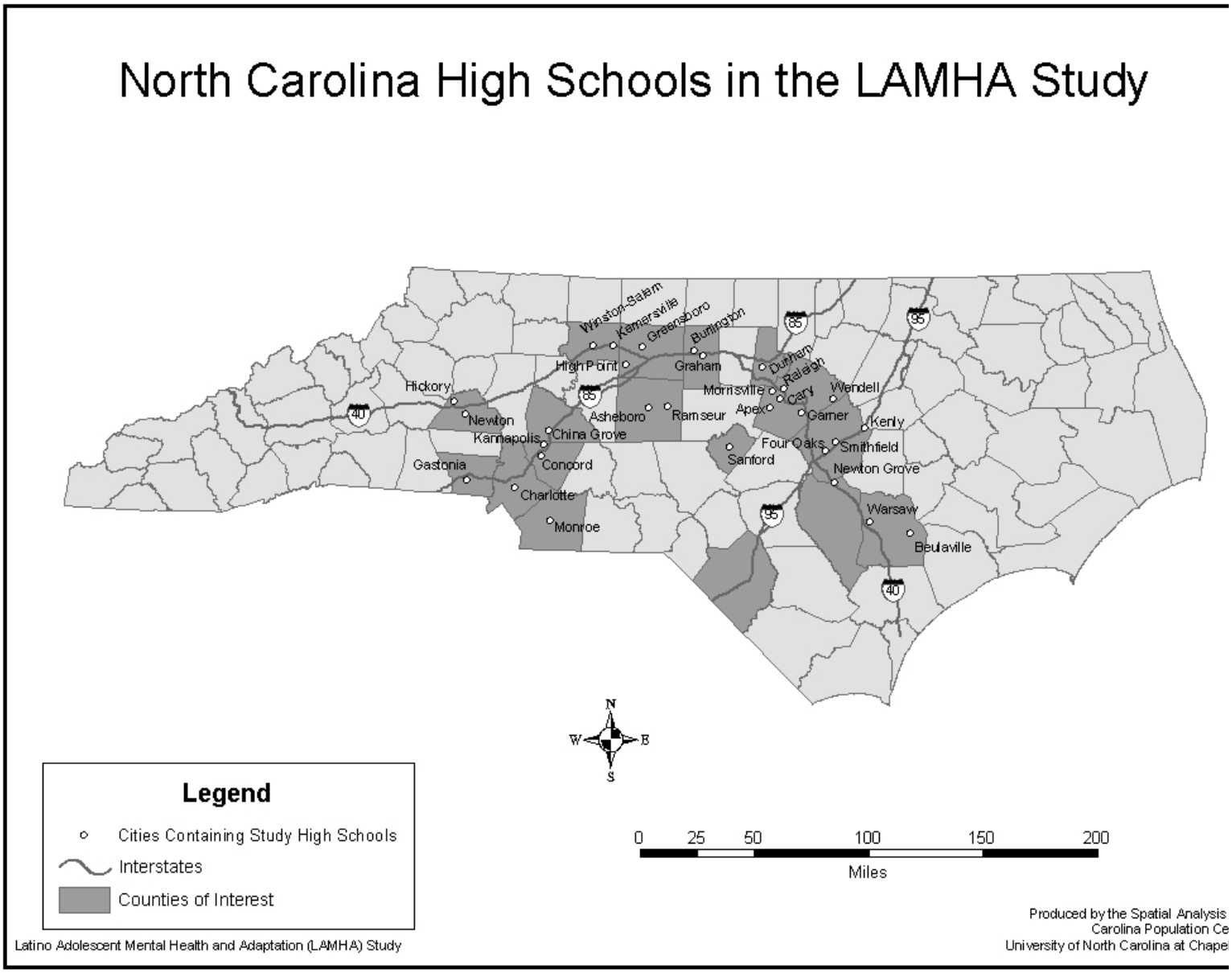

North Carolina Counties in the LAMHA Study
NOTE: The counties of interest in the LAMHA study are in the darkest category: those that have more than 5000 latinos and a growth rate greater than the state average of 394%.
Legend
Interstates
Urban Counties
Latinos < 5000 & Growth Rate < 394%
Latinos < 5000 & Growth Rate >= 394%
Latinos >= 5000 & Growth Rate < 394%
Latinos >= 5000 & Growth Rate >= 394%
N
W
E
S
0 25 50 100 150 200
Miles
Produced by the Spatial Analysis Unit
Carolina Population Center
University of North Carolina at Chapel Hill
Latino Adolescent Mental Health and Adaptation (LAMHA) Study

FIGURE 2


North Carolina High Schools in the LAMHA Study
Winston-Salem
Kernersville
Greensboro
Burlington
High Point
Graham
Durham
Raleigh
Hickory
Morrisville
Cary
Wendell
Newton
Asheboro
Ramseur
Apex
Garner
Kenly
Conover
China Grove
Four Oaks
Smithfield
Sanford
Gastonia
Concord
Newton Grove
Charlotte
Warsaw
Monroe
Beulaville
N
W
E
S
Legend
Cities Containing Study High Schools
Interstates
Counties of Interest
0 25 50 100 150 200
Miles
Latino Adolescent Mental Health and Adaptation (LAMHA) Study
Produced by the Spatial Analysis
Carolina Population Ce
University of North Carolina at Chapel