

UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Topics on Hispanic Demography: Foundations for the Demographic Analysis of the Hispanic Population in the United States

Permalink

<https://escholarship.org/uc/item/7jz408zn>

Author

Kaneshiro, Matheu Shoei

Publication Date

2011

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Topics on Hispanic Demography:
Foundations for the Demographic Analysis of the Hispanic Population in the United States

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Sociology

by

Matheu S. Kaneshiro

August 2011

Dissertation Committee:

Dr. David A. Swanson

Dr. Vanesa Estrada-Correa

Dr. Augustine Kposowa

Copyright by
Matheu S. Kaneshiro
2011

The Dissertation of Matheu S. Kaneshiro is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

I would first like to acknowledge my family for helping me to attain my Ph.D. It is 2011, I am currently 29 years old, and I am finally ready to officially start my career. It would never have been possible without my parents' foundational support in my education. Thanks also go to Veronica, who makes me feel whole and gives my life more meaning and purpose. A "special" thanks to my sister and her family; you guys bring our family closer.

Next, I would like to thank my mentor D. Swanson, without whom I would have either never finished my Ph.D., or would have emerged a much older and more disgruntled doctor. His research support and connections that he has provided me with have been invaluable. His Yoda-like nature also helped me to stay calm throughout the process.

My committee has also been very important in my development. Dr. Estrada-Correa has helped me to refine my methodological skills and academic portfolio, and I am grateful for the amount of time and detail she invested in me. I am also sure she wrote a great letter of recommendation that set me up with my first position outside of academia. Dr. Kposowa has helped me with my Master's thesis and I am thankful for his help on my dissertation. I've also enjoyed speaking on African politics and work at the U.N. (here's to hoping I end up in the bureaucracy).

Thanks and apologies go to my first advisors, Chris Chase-Dunn and Ellen Reese. Thanks for providing me with research and publication opportunities in my early career. You have helped me to round out many of my academic skills and I have grown much from being a part of your teams. I am sorry to abandon Political Economy, but jobs in academia are scarce (and my heart has never really been in academia).

Thanks to humans who have made this process more bearable. Anna and Heather, you guys cool. Talking with you was a good outlet for the pressure-cooker that is academia. Rich is also cool, but he'd better watch out for my strong kung fu. Rotondi, the Loves, Patrick, Megan, Jele, and the other Anthropologists have helped me not to forget to be human. My good students also helped to fuel my soul. Shout outs go to the Exhausted Prayer posse, my SDSU family, and the OG LA family (e.g. Hurtados, Rouras, Sato/Veloz').

Additional shout outs go to the Census peeps. J. Gregg Robinson was a great advisor, Tori Velkoff gave me a great internship, and Tammany Mulder hooked me up with a great gig. Cheers to the 7th floor and related; I'm sorry I couldn't join you in the next phase of my career, but the budget cuts had a great way of circumcising my plans.

Many people have helped me along the way. Some notables are Dr. B. and Dr. Hanneman for methodological advice, and Drs. Aguirre and Sorensen for being part of the prospectus committee. Juliann is also cool; unfortunately we never had the opportunity to co-author anything. Speaking of which, the rest of the WSF crew around the world were also great to know.

Thanks to UCR and the Sociology department for giving me this opportunity. It has been Hell, and I'm not sure it has all been worth it just yet. I am burnt around the edges, but at least I'm leaving with a job, a Ph.D., and a potentially bright future.

ABSTRACT OF THE DISSERTATION

Topics on Hispanic Demography:
Foundations for the Demographic Analysis of the Hispanic Population in the United States

by

Matheu S. Kaneshiro

Doctor of Philosophy, Graduate Program in Sociology
University of California, Riverside, August 2011
Dr. David A. Swanson, Chairperson

The objective of this dissertation is to evaluate the quality of the decennial Census through the use of Demographic Analysis (DA), which is a methodology that estimates population size by using data on births, deaths, in-migration, and out-migration. Using DA, the quality of the 2000 Census will be assessed by using estimates of Census undercount for children aged 0 – 9, as well as measures of relative undercount for the 1990 and 2000 Censuses for Hispanics of older ages. In cleaning the data that measure the components of the Hispanic population, the chapters of this dissertation address topics on the sociology of the Hispanic population.

The first chapter critiques the measurement of the Hispanic population in general, arguing that the Census Bureau's use of Hispanic identifiers is politically charged and ambiguous. Logistic regression is used to demonstrate that the personal identification with the Hispanic ethnicity is affected by dynamics of assimilation, race, and social context. The second chapter argues that the quality of the 1990 Census was compromised by the political atmosphere in which the Census was embedded, while producing alternate estimates of emigration using models informed by recent research. The third chapter re-visits the Hispanic Mortality Paradox

by accounting for its counter-explanations (i.e. death undercount and emigration). Using life tables, it will be shown that the existence of the Mortality Paradox largely depends on the assumptions that one makes regarding the quality of the data used, although it continues to hold when the most plausible data are used. Chapter four demonstrates that interethnic childbirth is largely influenced by a mother's race and ethnicity, although social factors such as education and the marriage market serve to "break down" ethnic divisions and stir the melting pot. Chapter five tests the success rates of a number of imputation methodologies for parental ethnicities and predicts the number of Hispanic births that have occurred throughout the 90s. Putting all these pieces together, the conclusion presents a range of estimates of undercount of Hispanic children aged 0 – 9, as well as relative undercount that compares the counts of the 1990 and 2000 Censuses.

TABLE OF CONTENTS

Introduction	1
Demographic Analysis: the Method Used to Evaluate Census 2000	3
Births	4
Deaths	4
Migration	5
Demographic Analysis for Hispanics	5
Fundamental Issues of this Dissertation: Topics in Hispanic Population Studies.....	9
What is Hispanic?	9
You Can't Count Me!	11
"Hispanic?" Says Who?	11
If I'm Half-Hispanic, Does That Make Me Hispanic?	12
Let's Start Counting.....	13
References	15

Chapter 1

The Ambiguous Ethnicity: Factors Contributing to the (Non)Identification with the Hispanic Ethnicity.....	18
Spanish/Hispanic/Latino (Origin or Descent): The Political/Historical Context.....	19
"Hispanic" Becomes an Ethnicity (Not a Race)	21
Moving Up and Out: Becoming White	22
Moving Sideways? Seeking an Alternate Theoretical Approach to Hispanic Identification.....	24
Hypotheses	26
Methods.....	27
Data.....	28
Descriptive Statistics of Variables by Racial/Ethnic Group	30
Social Mobility and the Racial Hierarchy, Multivariate Models.....	32
Moving Sideways: The Social Contextual Factors of the Hispanic Identity	33

Conclusion.....	34
Discussion	36
References	39
Tables	42

Chapter 2

Emigration and Undercount of the Hispanic Population, 1990 – 2000	46
Residual Estimates of Emigration and their Limitations.....	47
A New Generation of Emigration Estimates	50
The Current Study	52
Data.....	53
Methods 1: Residual Estimates of Emigration.....	55
Methods 2: Survey-Based Estimates of Emigration.....	56
Results.....	61
Estimates of Emigration.....	61
Explaining Error: Age, Year of Immigration, and Political Context	63
Conclusion.....	66
References	69
Tables	72

Chapter 3

Confronting the Challenges of the Hispanic Mortality Paradox: Emigration and Underreporting on Death Certificates.....	78
The Hispanic Mortality Paradox/Advantage.....	79
Hypotheses	84
Methods.....	85
Data.....	87
Results.....	90
Life Expectancies of Hispanics and Non Hispanic Whites, Before and After Death Data Adjustment	90

A Qualification of the Salmon Bias.....	91
Conclusion.....	95
Discussion	96
References	99
Tables	102
Figure	107
Appendix Tables.....	108

Chapter 4

Predicting Interethnic Childbirth by the Characteristics of Mothers: A Peek into the Melting Pot	110
Hispanic Family Formation.....	111
Interethnic Relationships.....	113
Interethnic Children	114
Data and Methods	116
Results.....	117
Discussion: What the Statistics Tell Us about Interethnic Parentage.....	119
Conclusion.....	121
References	124
Tables	127
Appendix Table	129

Chapter 5

Imputation Procedures for Birth Records:.....	130
A Continuation of Chapter 4	130
Common Imputation Procedures	131
Data.....	135
Methods.....	137
Results of Imputation Procedures	139
Conclusion, Part 1: Which Imputation Procedure to Use	142

Conclusion, Part 2: Estimated Counts of Hispanic Children Born in the United States	145
References	148
Tables	149
Conclusion	
A Demographic Analysis of the Hispanic Population	155
Onward: The Demographic Analysis of the Hispanic Population, Ages 0 – 9	166
Relative Undercount of the Hispanic Population, Ages 10 and Up	168
Conclusion of the Conclusion.....	171
References	174
Tables	176
Figure	178
Appendix Tables.....	179

LIST OF TABLES

Table 1.1: Select Descriptive Statistics by Hispanic Tri-Racial Classification Group	42
Table 1.2: Predictors of Socioeconomic Status Indicators for Hispanics, Multivariate Regression	44
Table 1.3: Multivariate Models Predicting the Identification with the Hispanic Ethnicity (odds ratios displayed)	45
Table 2.1: Emigration Rates by Study	72
Table 2.2: Hispanic emigration for those residing in the U.S. in 1990: 1990-2000	73
Table 2.3: Collective Error ¹ by Age.....	74
Table 2.4: Correlates of Collective Error, Correlation Matrix	75
Table 2.5: Count differences of Foreign-Born Hispanics across Censuses by year of entry	76
Table 3.1: Hispanic Count Adjusted for Undercount	102
Table 3.2: Hispanic Mortality Rates and Life Expectancies, with and without Death Adjustments: 1990.....	103
Table 3.3: Non Hispanic White Mortality Rates and Life Expectancies, with and without Death Adjustments: 1990	104
Table 3.4: Emigration counts for Hispanics before and after undercount Adjustment, 1990	105
Table 3.5: Life Expectancies of Non-Hispanic Whites and Hispanics Using Varying Data Adjustments	106
Table 3.6: Life Table for Hispanics, 1990, Unadjusted Data	108
Table 3.7: Life Table for Non Hispanic Whites, 1990, Unadjusted Data.....	109
Table 4.1: Percent of Births that are Interethnic, by Ethnicity and Marital Status of the Mother, 1990 - 2000.....	127
Table 4.2: Interethnic Childbirth by Characteristics of the Mother: Logistic Regression Odds Ratios.....	128
Table 4.3: Alternate Models Predicting the Ethnicities of Fathers. Logistic Regression Odds Ratios	129
Table 5.1: Models Used for Imputation of Father Ethnicities: Odds Ratios	149
Table 5.2: Results of Imputation Procedures on Missing Father Ethnicities	150
Table 5.3: Counts of Hispanic Children Aged 0 - 10 by Imputation Method	151

Table 5.4: Predicted Counts of Total Hispanic Children by Imputation Rule and Age.....	152
Table 5.5: Regression Models Predicting Hispanic Births for States Missing a Hispanic Item	153
Table 5.6: Predicted Counts of Native-Born Hispanic Children, Complete Data Set	154
Table 6.1: Demographic Analysis Estimate of Hispanic Children.....	176
Table 6.2: Pseudo DA and Relative Undercount for Hispanics, 1990 (adjusted) vs. 2000.....	177
Table 6.3: Data Components for Demographic Analysis, 2000	179
Table 6.4: Data Components for Pseudo-Demographic Analysis, 2000	180

LIST OF FIGURES

Figure 2.1: Hispanic Emigration for those Residing in the U.S. in 1990: 1990-2000	77
Figure 3.1: Emigration Rates for Hispanics after Undercount Adjustment, 1990	107
Figure 6.1: Relative Undercount, 1990 (Undercount-Adjusted) vs. 2000 Census	178

INTRODUCTION

Hispanics or Latinos (hereafter referred to as “Hispanics”) are arguably the hands of the United States. The unseen janitor that works the night shift, the butcher that packages and displays meat in pristine containers, the picker of the tomatoes that end up on one’s sandwich, and the chef that prepares one’s food from an isolated kitchen are all disproportionately Hispanics. Millions of Hispanics are also in professions of education and art, health care, and management and business, representing over one million of each of the aforementioned professionals (PEW Hispanic Center 2010). Representing 15.7% of the population of U.S. (derived from the 2009 American Community Survey), Hispanics are many of the people who are teaching, studying, working, living, and dying in the U.S.

Hispanics are rejuvenating the U.S. population. The non-Hispanic population has grown around 4.9% from 2000 to 2010 (based on Decennial Census data), which may pose problems in replacing the workers that increasingly retire and die from the Baby Boom generation. Fortunately, over the same decade, Hispanics have been growing at around 43%, rejuvenating the population with higher birth rates and providing labor through international migration. Despite all the issues regarding the measurement of this ambiguously-defined population, it is clear that the Hispanic population is growing. The estimate of a “Spanish origin” population however defined started at between 9 and 9.6 million in 1970 (based on a 5% Spanish origin item and 15% Spanish mother tongue item in 1970, respectively), to around 14.6 million in 1980 (being the first Census to include some form of Spanish/Hispanic/Latino self-identifier), 22.4 million in 1990, 35.3 million in 2000, and 50.5 million in 2010 (Gibson and Jung 2002; Passel, Cohn, and Lopez 2011).

For numerous reasons, it is important to have an idea of the size of the Hispanic population. From the public-policy perspective, it is important to know how public funds should be allocated to properly provide for the schools, roads, and services for communities. From a social justice perspective, proof of discrimination depends on accurate statistics on the size and the demographic characteristics of the population. From the political perspective, an understanding of the voting potential of Hispanics can help politicians to further the advancement of the U.S.

The problem with data on the size of the Hispanic population is that we are simply not sure how accurate the data are. The most reliable data that exist come from the U.S. Bureau of the Census (hereafter referred to as the “Census”) although the few studies on the accuracy of the Hispanic count on the Census have demonstrated undercounts of the population.¹ One of the two methods that the Census uses to measure error in the Census is Demographic Analysis (DA), having been used as a standard method of evaluation dating back to the 1960 Census. Unfortunately, DA has only historically examined “black” and “non-black” populations in attempting to understand Decennial Census errors by race and ethnicity.

The objective of this dissertation is to conduct DA for Hispanics specifically, taking into account the problems with data that lie along the way. With the growth of the Hispanic population and its current status as the largest minority group in the United States, the Census Bureau has an interest in understanding more about the Hispanic population and its size, and this research will serve to contribute to the development of DA for Hispanics by filling in some of the missing gaps in the current status of the field. And, considering the resounding success in

¹ Examples of Demographic Analyses (DA) for Hispanics include Robinson, Adlakha, and West (2002), Fernandez (1995), and Adlakha et al. (2003). Unfortunately, the “odd findings” (mainly undercount) that resulted from the use of unclear data were not followed up on, leaving Hispanic DA underdeveloped.

the accuracy of the DA releases that predated the 2010 Census (Cohn 2011), DA is sure to receive much attention in the coming years.

DA is an analytic technique that is used to produce estimates of the United States' population through the use of administrative records (primarily birth and death records). To measure the size of the population, DA utilizes the Fundamental Demographic Equation that sums the components of population size: namely births, deaths, and migration. In order to produce reliable estimates of the population based on these components, the data underlying these components must be heavily cleaned. There are a number of imperfections with the data as they exist including data that are missing, data that are misclassified, data that do not align across data sets, and data that do not validly measure a concept. The process of dealing with these problems is referred to as "cleaning data." This cleaning will be done en route to estimating the Hispanic population in the U.S., laying the foundations for DA methodologies for future research. To add flesh to this dissertation, the chapters will address specific topics in Hispanic population studies, uncovering larger dynamics of race and ethnicity while also cleaning the data to be used.

DEMOGRAPHIC ANALYSIS: THE METHOD USED TO EVALUATE CENSUS 2000

DA is an analytic technique that is used to produce estimates of the United States' population, largely relying on the use of administrative records as data inputs (primarily birth and death records). The purest form of DA utilizes the Fundamental Demographic Equation (FDE, Equation 1), whose core components are births, deaths, and migration.

$$\text{Population} = \text{Births} - \text{Deaths} + \text{Immigration} - \text{Emigration} \quad [1]$$

At its face, the FDE appears to be an extremely simple equation that would lend toward the production of a simple dissertation without the need of any in-depth analyses. However, the

process of performing DA is deceptively complex, as the estimation of each component of the FDE is a monstrous problem in itself since all data available are rife with error. In other words, for an accurate DA to be performed, each component of the FDE must be investigated and thoroughly cleaned. What follows in this section is a description of some of the data components and their attendant problems when used to estimate the size of the population in general. That is, this section describes the least of DA's worries and does not touch problems of ethnic identification, which will be reserved for later. The sources of data described below were used by the Census Bureau in its 2000 DA estimate.

Births

Births are the driving force behind DA and population sizes in general, and the data source that provides birth inputs come from vital registries of birth records (certificates). One problem with relying on birth record data is accounting for the completeness of birth registration. Birth records for 1940, for example, were estimated to be 92.5% percent complete overall, although 81.9% complete for births to blacks (Passel 1992). Birth registration completeness has steadily improved, reaching over 99% by the 1960s (McDevitt, O'Connell, and Joyce 2001). DA for older persons that relies on birth certificate data thus calls for "corrective" measures for imperfect birth registration. Due to the inconsistencies in older birth records, the Census Bureau also uses Medicare registration (and surveys of non-registration) to estimate the population 65 and older.

Deaths

As with births, deaths are similarly recorded by the vital registration system. There are few means with which to empirically test the completeness of the death count, although there are similarly few means to die in the U.S. without being assigned a death certificate. Due to these factors, all deaths after 1960 are assumed to be complete (U.S. Bureau of the Census 1988).

Migration

Migration is, quite simply, the component of DA that is the most difficult to estimate (Robinson 2010). Information on migration is drawn from a smattering of sources, each used to measure an equally smattered segment of the migrant population. Native-born net emigration is estimated using foreign Censuses and State Department data from foreign Consulate data, although the results are highly questionable (Gibbs et al. 2001; Robinson 2010). Armed forces overseas data are gathered from the Department of Defense. Emigration estimates for foreign-born persons in the 1990s are extrapolated from the estimates of such emigration for the 1980s. The emigration estimates of the 1980s are estimated using a cohort component residual technique that compares counts of the foreign-born across Censuses.

The main source of immigration statistics is based on legal permanent residents from the Immigration and Naturalization Service for 2000 (Perry et al. 2001). Estimates of temporary migrants (i.e. students and workers) were based on the Census' Supplemental Survey, using algorithms that rely on indicators of temporary residence in the U.S. (Cassidy and Pearson 2001). Migrants from Puerto Rico are estimated by using a residual survival method for the population residing in Puerto Rico (Christenson 2001). And if all of these sources are not adequate to capture migration, a "residual foreign-born migration" component is also included which is based on the difference between Census counts of the foreign born and the sum of the above sources of international immigration (Robinson 2010).

DEMOGRAPHIC ANALYSIS FOR HISPANICS

Demographic Analysis (DA) has traditionally been formally conducted by the U.S. Census Bureau for the entire United States population by age, sex, and the limited race categories of black/non-black (Robinson 2010). DA is particularly useful in measuring the accuracy of the decennial

Census as well as other surveys (Robinson, Adlakha, and West 2002; Velkoff and Devine 2009). It will not be until the coming 2010 Decennial Census results are released that DA will be conducted for the largest minority population in the United States (i.e. Hispanics), which continues to grow at a rate far exceeding the non-Hispanic population (Fry 2008; Guzman and McConnel 2002). These preliminary numbers introduce new, as-of-yet unchallenged estimates that will be highly scrutinized by policy makers and civil society alike. As DA for Hispanics is new, all of the unique challenges that the DA of Hispanics faces are still being freshly confronted, and it is important to enter the new decade with ever-improving estimates of the Hispanic population.

DA by Hispanic/non-Hispanic has traditionally taken a back-seat to DA by black/non-black populations for the U.S. Census Bureau, involving lesser-developed methods and receiving less treatment than the black/non-black analyses. Robinson, Adlakha, and West (2002) performed a variant of DA by using the 1990 Census count of Hispanics to estimate the number of Hispanics expected to be found in 2000, adjusting the 1990 counts by using data on births, deaths, and migration. When comparing these estimates to the Census counts in 2000, the authors found very large differences (labeled “error of closure of Census 2000”), including over 30% errors for males aged 18-29, and nearly 10% errors for those aged 0-19. These authors argue that these discrepancies are largely due to the understatement of immigration components for the Hispanic population. Working on the assumptions of previous research, Robinson, Adlakha, and West (2002) find that doubling the previous estimate of “residual immigration” produced a reasonable upper DA estimate of Hispanics.

Using a birth-driven estimate of Hispanic children, Fernandez (1995) estimated the Hispanic population under 10 years of age for 1990. In his analysis, Fernandez notes that

estimates of Hispanics using births is highly sensitive to the way in which one classifies a birth as Hispanic; using the “mother rule” (a classification rule which assigns a child a Hispanic ethnicity if his/her mother is Hispanic – also the “official” assignment rule used by the National Center for Health Statistics in summary statistics) leads to unrealistically low estimates of the numbers of Hispanic children vis-à-vis Census 1990. Conversely, coding births as Hispanic when either the mother or father is also listed as Hispanic leads to “more plausible” figures when comparing to the Census. Additionally, estimates of the Hispanic population are highly sensitive to how missing data on ethnicity are imputed.

Adlakha et al. (2003) also estimated the Hispanic population under 10 for Census 2000 using birth records, life table survival ratios, and estimates of net migration. These authors revisit the dilemma concerning the classification of children as Hispanic by first noting that the classification of children as Hispanic using the mother rule produced results that were “clearly inconsistent” with other studies, as such an estimate suggested that the undercount of Hispanic children was at an abnormally low .4% nationally. Classifying children as Hispanic if either parent was Hispanic produced undercount estimates of 12%.

The previous DA attempts for the Hispanic population were all very important in highlighting areas of research that would assist in the production of reliable estimates of the Hispanic population. Unfortunately, these studies were not followed by subsequent attempts to alleviate the problems highlighted, leaving Hispanic DA underdeveloped. Robinson et al. (2002) demonstrated that immigration components fall short when measuring the residual migration component (possibly by half). Fernandez (1995) raises the issue of identifying Hispanics on birth certificates which becomes particularly tricky when attempting to identify how many interethnic children will identify as Hispanic (Lee and Edmonston 2006). This issue is revisited by Adlakha et

al. (2003) who, like Fernandez (1995), argued that classifying children as Hispanic when either parent is Hispanic produces results that are more consistent with previous research. Fernandez (1995) also notes that the imputation procedures performed to deal with missing data can have large effects on the count of Hispanics.

In sum, all of the essential data components of the Hispanic population are in need of serious “cleaning” if one is to conduct the DA of Hispanics. The previous variants of Hispanic DA highlight problems in birth classification, immigration, and imputation procedures in specific. Emigration estimates are also imperfect, as previous attempts to nail down reliable rates of emigration have fallen short when using DA techniques (Mulder, Guzmán, and Brittingham 2002). Death data also has its own host of errors of estimation for Hispanics. Race and ethnicity on death certificates are recorded by a funeral director instructed to gather information from an informant or from his/her own observation. Both of these methods can differ from one’s self-identification (and thus one’s identification in the Census), as verified in numerous studies (Arias et al. 2008; Rosenberg et al. 1999).

In light of all of these problems, this research will attempt to provide “clean” data inputs in order to produce alternate population counts of the Hispanic population, evaluating the quality of the Decennial Censuses in the process. The chapters of this dissertation address topics in Hispanic population studies (see next section), but they also provide insight on dealing with problems in the data. Chapter 2 will use new literature to produce emigration estimates for the 1990-2000 period while providing assumptions regarding the age groups that are most undercounted in the 1990 Census. Chapter 3 will apply undercount to the 1990 Census baseline population while also adjusting data on death counts in order to account for ethnic

misclassification. Chapters 4 and 5 will develop imputation procedures for missing data on death certificates and use the latest literature to assign ethnicities to births.

FUNDAMENTAL ISSUES OF THIS DISSERTATION: TOPICS IN HISPANIC POPULATION STUDIES

While the ultimate goal of this dissertation is to measure the size of the Hispanic population, the chapters of the dissertation navigate through many topics in Hispanic population studies as well as in ethnic studies in general. Chapter 1, for example, comments on assimilation and other mechanisms that lead to the formation of one's ethnic identity. Similar to Chapter 1, Chapter 2 highlights the political undertones that are inherent in data-collection, as the political context is argued to be a strong influence in "coloring" the data. Chapter 3 touches on the Hispanic Mortality Paradox and finds that, while the results heavily depend on the assumptions that one makes, Hispanics retain a mortality advantage over Non Hispanic whites even when controlling for the counter-explanations of the Mortality Paradox. Chapter 4 demonstrates that interethnic births are partially the result of one's demographic characteristics and the social context in which one lives.

The chapters of the dissertation thus touch upon varying topics that are relevant for population and ethnic studies. In order to measure Hispanics, there are many issues that need to be addressed. A summary of some of the issues and questions tackled in the dissertation follows.

What is Hispanic?

Before even considering the concept of measuring a Hispanic population, one needs to first ask the obvious but deceptively complex question: "well, who qualifies as Hispanic?" Due to the politics of the 1970s and the timing of the Census, the "Hispanic" category became an ethnicity

on the Decennial Census which was distinct from a race. That is, persons were asked to identify *both* racially and ethnically. This distinction between race and ethnicity has become confusing to many residents of the U.S., leading many to simply not answer the Hispanic item or to identify as “some other race” when ethnically identifying as Hispanic.

The conflation of the race and Hispanic concepts influences a person’s identification with the Hispanic item. Chapter 1 will demonstrate that persons with a racial identity that is distinct from the “racial Hispanic” concept are more often likely to not identify as an ethnic Hispanic, despite having been born in or having ancestry from an American Spanish-speaking country. That is, racial blacks and Asians from Latin America have higher rates of non-identification with the Hispanic ethnicity than do their white and “some other race” counterparts. Additionally, the social environment in which one lives influences the identification with the Hispanic ethnicity, as these “Hispanics of heritage” vary in identification with the ethnicity based on the ethnicity of one’s spouse and the presence of fellow Hispanics in one’s community. Assimilation and acculturation also influence the identification with the Hispanic ethnicity. Advancing in assimilation and acculturation indicators, such as years of residence in the U.S. and income, leads to higher rates of non-identification with the Hispanic ethnicity.

Chapter 1 argues that among Hispanics are populations of “ambiguous ethnics” that are in between ethnic and racial identities. For these persons, the social environment in which one lives can either push persons from, or pull them into identification as Hispanic. This process of identity-formation can also theoretically apply to all groups, wherein the definitions that are laid before a person, whether from the Census or from the population at large, influences the formation of one’s own identity.

You Can't Count Me!

Even aside from the identities of Hispanics, the actual capturing of populations becomes an issue if persons avoid to be counted in the Census. Chapter 2 will demonstrate that the estimation of emigration becomes, frankly, impossible when surviving the 1990 Census counts to 2000 and assuming that the differences between the two comes from emigration. Alternate estimates of emigration are in order, and Chapter 2 produces these emigration estimates using three recent studies that produce emigration rates. Chapter 2 will also be used to produce crude measures of undercount for the Hispanic population counted in the 1990 Census and suggestions toward alleviating the data-complications that arise from the undercount of the 1990 Census.

Chapter 2 also comments on data-collection within political contexts. It is argued that the political context of the 1990 Census had a tremendous influence on the suspicion of the government for Hispanics. With the passage of the Immigration and Reform Act of 1986, many Hispanics may have felt threatened by the government, possibly viewing the Census as an instrument with which the government would be better able to closely monitor and sanction the Hispanic population. This finding reminds the data-user that data are rarely apolitical, as politics is intimately involved with the quality of the data that one collects.

"Hispanic?" Says Who?

The data that are used to identify the Hispanic population are collected through different means. Unfortunately, this leads to inconsistent rates of identification with the Hispanic population. Death data, for example, is recorded by a funeral director, who ideally would use an informant but would otherwise use independent judgment. In contrast, individuals themselves identify themselves on Census instruments. This means that data on Hispanics does not always

come from the same source, which would ideally be the individual. On the one hand, someone else writes one's ethnicity on death certificates. On the other hand, the person him/herself identifies herself as Hispanic on Census items. This produces errors of count that make data-comparisons questionable.

Chapter 3 addresses this issue by providing adjustments to death data to account for the differing rates of identification with the Hispanic ethnicity that stem from the method of data collection. Death rates need to be adjusted upward in order to account for the deflated counts of deaths that occur to persons who were not identified by others as Hispanics, although they would otherwise identify themselves as Hispanic.

Accounting for the misclassification of death certificates, Chapter 3 revisits the Hispanic Mortality Paradox which is characterized by the high health outcomes for Hispanics despite their tendency to have lower socioeconomic statuses (which is associated with low health outcomes). Accounting for death misclassification, this research demonstrates that the Mortality Paradox still holds, as Hispanics continue to demonstrate a mortality advantage over Non Hispanic whites. Additionally, Chapter 3 applies the lessons of Chapter 2 in applying rates of population undercount and emigration (which also must be adjusted for undercount) to the data in order to determine the health outcomes of Hispanics, continuing to demonstrate that Hispanics enjoy a mortality advantage to Non Hispanic whites. However, the health outcomes of Hispanics heavily depend on the assumptions that one makes regarding the quality of the data.

If I'm Half-Hispanic, Does That Make Me Hispanic?

Chapters 4 and 5 address the role that interethnicity plays on Hispanic population counts. As with interracial children, interethnic children will continue to change the ethnic landscape of the U.S. and pose interesting dilemmas to population researchers who wish to identify "clearly

defined” populations. With the blurring of the boundaries between races and ethnicities, the task of clearly defining populations will become increasingly difficult.

Chapter 4 tests models that predict the birth of an interethnic child on the basis of a mother’s demographic characteristics. It will be shown that Hispanic mothers who have interethnic children have very different demographic profiles than do their Non Hispanic counterparts. Specifically, Hispanic mothers with interethnic children tend to be older and more highly educated, whereas Non Hispanic mothers tend to be younger and less educated. Hispanic mothers with interethnic children also tend to have fewer previous children than their Non Hispanic counterparts. Time and marriage market indicators also have an influence on having interethnic children. In general, this research suggests that the boundaries between ethnicities in mating patterns are being neutralized as the social environment in which one lives, including neighborhood ethnic composition and educational networks, are factors that influence mating patterns. In other words, demographic characteristics, time, and social context are contributing to the “melting pot” effect.

Chapter 5 applies the findings of Chapter 4 for the purposes of DA. The models that are used to predict the ethnicity of one’s mating partner are used to impute missing parental ethnicity on birth certificates (which is around 14% for fathers). Additionally, Chapter 5 applies rates of identification as “Hispanic” for interethnic children to predict the number of Hispanic children that would be identified as Hispanic on the Census.

Let’s Start Counting

Having cleaned the necessary data components, the Conclusion of the dissertation will be poised to perform a DA for the Hispanic population under 10. Births would have identified the number of Hispanic children, deaths would have been cleaned, and assumptions regarding

migration would have been made. With all of these data components, this research will be in a position to estimate the Hispanic population under 10 and produce rates of undercount of the 2000 Census by age.

Due to the relative recency of the inclusion of the Hispanic ethnicity item on birth certificates, the Hispanic population aged 10 and up will not be able to be measured using a pure DA method. Instead, a modified or “pseudo DA” will be conducted for the 10+ Hispanic population. This method will take the same cleaned death and migration components, but will also use an adjusted 1990 Census count to provide a baseline estimate of the Hispanic population. With these scores, a “relative undercount” that compares the undercount rates for the 1990 and 2000 Censuses will be made.

REFERENCES

- Adlakha, Arjun, J. Robinson, Kirsten West, and Antonio Bruce. 2003. "Assessment of Consistency of Census Data with Demographic Benchmarks at the Subnational Level." *Census 2000 Evaluation O.20*. Washington, D.C: U.S. Census Bureau.
- Arias, Elizabeth., William Schauman, Karl Eschbach, Paul Sorlie, and Eric Backlund. 2008. "The Validity of Race and Hispanic Origin Reporting on Death Certificates in the United States." *Vital Health Stat 2*:1-24.
- Cassidy, Rachen and Lucinda Pearson. 2001. Evaluating Components of International Migration: Temporary (Legal) Migrants. *Population Division Working Paper Series* No. 60. Washington, D.C.: U.S. Census Bureau.
- Christenson, Matthew. 2001. Evaluating Components of International Migration: Migration between Puerto Rico and the United States. *Population Division Working Paper Series* No. 64. Washington, D.C.: U.S. Census Bureau.
- Cohn, D'Vera. 2011. "State Population Estimates and Census 2010 Counts: Did They Match?" *PEW Social & Demographic Trends*. Retrieved March 1, 2011 (<http://census.pewsocialtrends.org/2011/state-population-estimates-and-census-2010-counts-did-they-match>).
- Fernandez, Edward W. 1995. "Using Analytic Techniques to Evaluate the 1990 Census Coverage of Young Hispanics." *Population Division Working Paper Series* No. 11. Washington, D.C.: U.S. Census Bureau.
- Freire, Paulo. 1993. *Pedagogy of the Oppressed*. New York, NY: Continuum.
- Fry, Richard. 2008. "Latino Settlement in the New Century." *PEW Hispanic Center*. Retrieved March 1, 2011 (<http://pewhispanic.org/files/reports/96.pdf>).
- Gibbs, Jim, Gregory Harper, Mark Rubin, and Hyon Shin. 2003. "Evaluating Components of International Migration: Native-Born Emigrants." *Population Division Working Paper Series* No. 11. Washington, D.C.: U.S. Census Bureau.
- Gibson, Campbell and Kay Jung. 2002. "Historical Census Statistics on Population Totals By Race, 1790 to 1990, and By Hispanic Origin, 1970 to 1990, For The United States, Regions, Divisions, and States." *Population Division Working Paper Series* No. 56. Washington, D.C.: U.S. Census Bureau.
- Guzmán, Betsy and Eileen McConnell. 2002. "The Hispanic Population: 1990-2000 Growth and Change." *Population Research and Policy Review* 21:109-128.

- Lee, Sharon and Barry Edmonston. 2006. "Hispanic Intermarriage, Identification, and U.S. Latino Population Change." *Social Science Quarterly* 87(5):1263-1279.
- McDevitt, Tom, Martin O'Connell, and Colleen Joyce. 2001. *Evaluating the Components of Births and Deaths used in Demographic Analysis*. Washington, DC: Population Division, U.S. Census Bureau.
- Memmi, Albert. 1991. *The Colonizer and the Colonized*. Boston, MA: Beacon Press.
- Mulder, Tammany J., Betsy Guzmán, and Angela Brittingham. 2002. "Evaluating Components of International Migration: Foreign-born Emigrants." *Population Division Working Paper Series* No. 62. Washington, DC: U.S. Census Bureau.
- Passel, Jeffrey. 1992. *Alternative Estimates of Black Births Corrected for Underregistration: 1935-1980*. Washington, DC: The Urban Institute.
- Passel, Jeffrey S., D'Vera Cohn, Mark Hugo Lopez. 2011. "Hispanics Account for More than Half of Nation's Growth in Past Decade." *PEW Hispanic Center*. Retrieved July 21, 2011 (<http://pewhispanic.org/reports/report.php?ReportID=140>).
- Perry, Marc J., Barbara J. Van der Vate, Lea Auman, and Kathleen Morris. 2001. "Evaluating Components of International Migration: Legal Migrants." *Population Division Working Paper Series* No. 59. Washington, DC: U.S. Census Bureau.
- PEW Hispanic Center. 2010. "Statistical Portrait of Hispanics in the United States, 2008." *PEW Hispanic Center*. Retrieved March 1, 2011 (<http://pewhispanic.org/factsheets/factsheet.php?FactsheetID=58>).
- Robinson, J. Gregory. 2010. "Coverage of Population in Census 2000 Based on Demographic Analysis: The History Behind the Numbers." Presented at the U.S. Census Bureau Workshop: 2010 Demographic Analysis Technical Review, January 8, Suitland, MA.
- Robinson, J., Arjun Adlakha, and Kirsten West. 2002. "Coverage of Population in Census 2000: Results from Demographic Analysis." Presented at the Annual Meeting of the Population Association of America, May 8-11, Atlanta, GA.
- Rosenberg, Harry, Jeffrey Maurer, Paul Sorlie, Norman Johnson, Marian MacDorman, Donna Hoyert, James Spitler, and Chester Scott. 1999. "Quality of Death Rates by Race and Hispanic Origin: A Summary of Current Research, 1999." *Department of Health and Human Services Publications* No PHSP 99-132B.
- U.S. Census Bureau. 1988. "Methodology for Developing Estimates of Coverage in the 1980 Census Based on Demographic Analysis: Birth and Death Statistics, 1935-1980." *1980 Census Preliminary Evaluation Results Memorandum Series 112*. Washington, DC: U.S. Census Bureau.

Velkoff, Victoria and Jason Devine. 2009. "The 2010 Census Demographic Analysis Program."
Presented at the Quarterly Meeting of the Council of Professional Associations on Federal
Statistics, September 11, Alexandria, VA.

CHAPTER 1

THE AMBIGUOUS ETHNICITY: FACTORS CONTRIBUTING TO THE
(NON)IDENTIFICATION WITH THE HISPANIC ETHNICITY

America is not a country. America is comprised of three continents, each with a rich historical legacy that involves migration, dispossession, slavery, and ethnic mixing. While the U.S. has had interest in measuring the sizes of the non-North Americans, the task of identifying these “other” Americans has been a tricky endeavor.

There exist classic images of the “other Americans.” These are the persons who are often viewed racially as Hispanic or Latino/a and feature “indigenous” characteristics, most notably with darker skin and hair than that which is found on the classic images of whites. On the other hand, there are other “other Americans” who are as white as any European, or as black as any African. Although these other Americans have many diverse features, the U.S. Americans attempt to put all of them in a single category.

The diversity of the other Americans is captured in the “Hispanic” ethnicity (which does not include Canadians and other Americans from non-Spanish speaking countries). It is a neat category that subsumes the majority of the other Americans, as the majority of the other Americans use Spanish as their mother tongue. Inevitably, many of these Hispanics so defined do not fit into the classic image of the “other” American. In these cases, problems of identification arise wherein persons conflate the Hispanic ethnic category based on their “mother tongue” (perhaps even forgotten after generations) with the Hispanic racial identity.

This chapter seeks to explore some of the dynamics that underlie the identification with the Hispanic ethnicity in the Census. Critically viewing the Hispanic ethnicity, it demonstrates

the political background, racial undertones, racial relations, and contextual nature that make identification with the Hispanic ethnicity fluid and dependent upon numerous factors. For Hispanics, a “triracial hierarchy” is demonstrated to be a function of assimilation, influencing those most assimilated to move progressively toward whiteness and potentially abandoning the Hispanic ethnicity altogether. Additionally, this chapter highlights the ambiguous nature of the Hispanic ethnicity in general, highlighting how race and social context have the strongest influence on the identification as Hispanic.

SPANISH/HISPANIC/LATINO (ORIGIN OR DESCENT): THE POLITICAL/HISTORICAL CONTEXT

The history of the identification of a “Spanish speaking” population is inevitably tied to the political context in which each age is embedded. At all times, political interests had been involved, creating a non-white population to be identified (whether Mexican or Spanish-speaking) and counted.

The historical context of the 1920s was particularly sensitive to “foreigners” and immigration, as the U.S. government was undergoing adjustment in the wake of World War I. European migration was limited, prompting the need to reach out to Mexican workers. This was followed by the Great Depression and massive deportations (Durand et al. 2001). It was in 1930 that the “Mexican” category made its (first and final) appearance in the U.S. Census as a race. Mexican-Americans and the Mexican government opposed having Mexican as a racial category, demanding it to be removed from the racial options on the Census (Siegel and Passel 1979:7). Chapa (2000) argues that a major factor in eliminating the “Mexican” category stemmed from the many rights that were only available to “whites,” not the least of which being the right to become a citizen. Ten years later, during the rumblings of the Second World War and the height

of racial tensions, Census enumerators for the 1940 Census were told to classify Mexicans as “white.”

Aside from the 1930 Mexican racial category, no items were included on the Census that could directly measure any Hispanic population until 1970. Indirect measures of a Hispanic population were found in the 1940 “persons of Spanish mother tongue” item, as well as the 1950 and 1960 “persons of Spanish surname” item (although sampling and coding issues make consistent measurement across Censuses questionable, Bean and Tienda 1987).

The relevance and necessity of including race on the Decennial Census was called into question in the 1960s, leading to the serious consideration of the removal of racial identifiers for the 1970 Census. However, in the midst of the civil rights movements of the 1960s, minorities gained leeway in public policy through such vehicles as the Voting Rights Act of 1965 which banned racial discriminatory impediments to voting (Farley 1991; Snipp and Lott 2009). At this point in time, proliferating forms of federal aid became contingent upon population size and proof of discrimination. From this turn of events, the social climate made it imperative that race (and later, ethnicity) would be included in the Census for collecting data on minorities – a social imperative that became law through OMB Directive Number 15 in 1977.

Sensing a need to be accurately counted, Mexican-American advocacy groups demanded a self-identification item on the 1970 Census, insisting that the surname method greatly underestimated the size of the Hispanic population. The Census Bureau rejected the inclusion of a direct Hispanic identifier due to the lack of time available for testing and because millions of surveys have already been printed. The White House then intervened, leading to the appearance of a Hispanic ethnic identifier on the 5% long-form sample of the Census survey (Choldin 1986).

“HISPANIC” BECOMES AN ETHNICITY (NOT A RACE)

It is in this historical context that the Hispanic item was included in the Census. A “color or race” item had already been included in the 1970 Census surveys which included the options of Indian (American), white, Negro or Black, Japanese, Chinese, Filipino, Hawaiian, Korean, and Other. For 5% of the Census surveys the questionnaire included, many items after the race item, “is this person’s origin or descent...” with the options of Mexican, Puerto Rican, Cuban, Central or South American, Other Spanish, and No, none of these. As is expected from the inertia of institutions, this distinction between the race and origin (later termed “ethnicity”) would not be changed in subsequent Census items. To complicate matters further, the 2000 Census item uses neither “origin” nor “descent” to solicit identification with the Hispanic ethnicity in asking “is this person Spanish/Hispanic Latino”, changing the framing of the ethnicity from an ancestral ethnicity (as implied from the words “origin” and “descent”) to a personal identity.

The distinction between race and (Hispanic) ethnicity has been preserved at least in part due to the demands of the Hispanic community itself. The National Council of La Raza decided to support the combination of the race and ethnic categories only if doing so led to a “greater and more accurate response rate” (National Council of La Raza 1995:8). Tests of the combined question indicated that a lower response rate would result, prompting strong resistance to enact such proposals (Rodriguez 2000).² At all times, however, the Census has been beholden to the final say of the Office of Management and Budget (OMB) who defines the racial and ethnic categories that are to be present in the Census.

² Subsequent studies, however, suggest that a combined race and ethnicity item will not drastically alter counts of Hispanics when respondents are allowed to choose more than one race (Hirschman, Alba, and Farley 2000).

There is evidence, even from the Hispanic community, that the separation of race and ethnicity is confusing. Identification with a race does not appear to be very clear, as evidenced in the over 40% of Hispanics who do not identify as an official race in choosing “some other race” (Greico and Cassidy 2001; Navarro 2007). And, in selecting identification with some other race, many Latinos write in a national country of origin or panethnic identifier (i.e. “Latino”) in specifying the other race with which they identify (Rodriguez 2000). Indeed, most Latinos would select a Hispanic race if given the opportunity to do so (Campbell and Rogalin 2006; Hirschman, Alba, and Farley 2000). In a society where racial options only include white, black, Asian, and some form of Native American/Islander, many ethnic Hispanics find that their race is not represented and subsequently identify as “other.”

MOVING UP AND OUT: BECOMING WHITE

One view on the Hispanic ethnicity is that is a reflection of a minority status – which is a view that historically held true in the decades in which rights were only afforded to whites. From this perspective, theory on the “colonized mentality” can be useful in explaining some of the patterns seen in the self-identification with the Hispanic ethnicity. This theory posits that colonized cultures and histories become lost to children as they become educated by the dominant culture in its norms, perspectives, and histories. The dominated class progressively accepts the portraits painted of them, viewing themselves and their roles in the same light as their dominators (Freire 1993; Memmi 1991).

Using theory on the colonized mentality, the Hispanic identity can be seen as the label that is used by the dominant culture to describe the minority group. While Hispanics may initially prefer to identify with specific nations of ancestry, years of living in the United States may lead these persons to accept panethnic labels (Mendieta 2000; Portes and Rumbaut 2001).

These colonized persons arguably accept the label that is used to describe them due to discrimination, which imparts a minority identity that sets them apart from whites (Golash-Boza 2006; Portes and Rumbaut 2001). On the other hand, Mexicans of higher socioeconomic status have been demonstrated to identify as “Hispanic” less frequently than their lower-status counterparts is related to (Duncan and Trejo 2007, 2008; Alba and Islam 2009). Persons who claim ancestry or birth from a Spanish speaking (i.e. “Hispanic”) country but do not identify as “Hispanic” have been called “Hispanic non- Hispanics” (Emeka 2008).

In rejecting one’s self and one’s culture, the theory on the colonized mentality argues that persons can become (or attempt to become) assimilated into the dominant culture. Classical assimilation posits that persons can become integrated into the larger society through socio-psychological means by such processes as cultivating friendships, eliminating prejudices, identifying with the larger society, and marrying with a member of the dominant group (Gordon 1964; Park 1950). Other literature on assimilation uses socioeconomic indicators to measure incorporation into society, viewing socioeconomic advancement (particularly increases in income and education) as indicators of assimilation (Neidert and Farley 1985).

The tri-racial hierarchy theoretical perspective argues that race relations in the United States are moving away from a simple binary (i.e. “black/white”) and moving toward a tri-racial structure. Those at the very bottom are classified under the “collective black” category (e.g. dark-skinned Hispanics) while “whites” (e.g. whites and assimilated Latinos) occupy the top. In the middle are “honorary whites” (e.g. light-skinned Latinos) who are higher socioeconomically than the “collective black” and who buffer ethnic tensions between the two poles (Bonilla-Silva 2004; Bonilla-Silva and Glover 2006).

In line with the tri-racial hierarchy, the empirical literature does suggest that race is important in explaining social variation among Hispanics. It is suggested that shades of whiteness for Hispanics can be attained through nativity, education, and income, as these three variables are positively related to the identification with the “white” race versus “some other race” for Hispanics (Lee and Tafoya 2006; Tafoya 2004). Leaving the Hispanic category altogether and entering the “white” category is also a function of race and interethnicity, as interethnic children are more frequently identified as “white” when the Hispanic parent is identified as Hispanic-white than when he/she is Hispanic-non-white (Brunsma 2005; Qian 2004).

A politically-influenced interpretation of the adoption of the Hispanic label as a specifically minority label may draw from the use of minority racial and ethnic statistics for political purposes. That is, as the usage of data collected for minorities vis-à-vis whites is used as evidence of racial/ethnic inequality and discrimination, the identification with a minority group for persons of fewer socio-economic resources may be a subconscious way to signal to the public that inequality continues to exist. That is, the bias for “Hispanics” of fewer resources to identify as Hispanic may be a means to highlight the struggles that Hispanics face, or at least to set one’s self apart from a socially/politically dominant group.

MOVING SIDEWAYS? SEEKING AN ALTERNATE THEORETICAL APPROACH TO HISPANIC IDENTIFICATION

What the above literatures lack is a third option out of a simple ladder-like advancement that moves persons from oppressed to majority cultures. For Hispanics, the previous theories offer explanations of assimilation into non-Hispanic “whiteness,” although this approach does not apply to non whites who do not identify as Hispanic, despite being born in or having ancestry

from a Spanish-speaking country. A theoretical approach is needed to explain Hispanic identification in general that is not limited to discourses of assimilation and socioeconomic advancement.

In seeking an explanation for identification with the Hispanic ethnicity in general, the history of the Hispanic term in the Census provides some insight. It is clear that the inclusion of the Hispanic item on the Census was not entirely an “organic” process in which an identity arose from the general society to reflect a self-identified population. Instead, political interests maneuvered to mold an item that could be conveniently placed in the Census survey. In many ways, asking a person to identify as “Spanish/Hispanic/Latino” is an ambiguous, government-defined ethnicity that mashes potential linguistic, racial, national, and ancestral interpretations.

This research will proceed with the proposition that governmental and societal racial/ethnic categories do not always align with a person’s identity. As such, there would inevitably be the presence of *ambiguous ethnics* who are caught in the borderlines of racial/ethnic categories or who can be placed in more than one category. These ambiguous ethnics may hold multiple (some relatively dormant) ethnic identities throughout their lives, receiving direct and contrasting messages from both society and public agencies regarding what identities they should adopt. Treatment by society and/or public institutions can “push” them to hold on more firmly to a primary salient identity and leave behind another.

The term “Hispanic” is sticky in that it can either be a racial identity or a linguistic/ancestral/national category. For persons identifying with a Hispanic ethnicity as well as “some other race” (nearly half of Hispanics), the non-selection of any of the available race categories suggest that their racial identity *is* Hispanic. Persons who are of countries or ancestries of Spanish speaking countries are linguistically “Hispanic” due to OMB’s categorical

definition. However, when a linguistic Hispanic's racial identity is firmly rooted in a racial category that is not the "classic" Hispanic racial identity (e.g. "black"), that person's racial and linguistic identities are disconnected, leading the "Hispanic" identity to be ambiguous and tenuous since they do not identify as a racial Hispanic. For these persons, one's social context – friends, family, and treatment by society – can lead to the "pushing" toward a particular racial/ethnic identity at the expense of another.

There are thus two dynamics that apply for Hispanics in the United States using the ambiguous ethnic hypothesis. The first dynamic is the racialization of the term "Hispanic," which is most evidenced in Hispanics' racial identification with "some other race." The second dynamic is the volatility of the linguistic "Hispanic" category for those who do not racially identify as Hispanic. The linguistic Hispanic ethnic identity can be activated or deactivated by the persons with whom one interacts, such as one's family and community. The messages that one receives by those surrounding her can serve to "push" her to identify one way or another.

HYPOTHESES

All of the following hypotheses apply to persons who either identify as Hispanic or have been born in, or claim ancestry from, Spanish-speaking countries (excluding Spain). In bringing the assimilation and racial-hierarchy literatures together, this research hypothesizes that advancement up the racial hierarchy is associated with advanced levels of the indicators of assimilation. That is, advancement up the racial hierarchy is associated with greater levels of out-marriage with non-Hispanic whites, for those who are married (H1). Additionally, indicators of socioeconomic assimilation are hypothesized to be related with the racial hierarchy; advancement up the racial hierarchy is associated with higher levels of education, income, socioeconomic status, and occupational prestige (H2). Other indicators of acculturation include

nativity and speaking English at home, which are hypothesized to be associated with both the racial hierarchy (H3) and the indicators of socioeconomic assimilation (H4).

A second objective will be to seek explanations for the identification with the Hispanic ethnicity in general. Drawing on the ideas proposed in the “moving up” section, it will be hypothesized that the non-identification with the Hispanic ethnicity will be related to higher levels of income (H5) and education (H6). In line with the ideas behind the “ambiguous ethnics,” this research will test the effects of the race and social context on the Hispanic ethnicity. It will be hypothesized that those identifying as “some other race” are more likely to identify as Hispanic (H7), while those who identify as black or Asian are less likely to identify as Hispanic (H8). Additionally, this research will hypothesize that speaking English at home and having a non-Hispanic white spouse deactivate the Hispanic identity (H9 and H10, respectively), while marrying a Hispanic is hypothesized to activate the Hispanic identity (H11).

METHODS

This research will use bivariate statistics, namely percentages, means, and medians by racial/ethnic group in simple explorations of the data. This will tell us the general characteristics of the populations under study and their variation by the racial groups of white, honorary white, collective black, and ambiguous ethnic. Multivariate linear regression will be used primarily to test whether the racial hierarchy is related to the indicators of socioeconomic assimilation. Multivariate logistic regression will lastly be used to predict identification with the Hispanic ethnicity.

DATA

All of the data used will come from the Integrated Public Use Sample of the 2000 Census, conveniently and wonderfully packaged by the Minnesota Population Center (Ruggles et al. 2010). Heads of household were exclusively used in this analysis, as these were the persons who (theoretically) filled the Census forms themselves, making all data gathered by self-identification. All self-identified Hispanics were included in this analysis. Additionally, all those claiming ancestry (the first ancestry listed in the Census) or birth in Spanish-speaking countries in the Americas were included (that is, Brazil, Spain, and Portugal were notably excluded in this analysis). Thus, all persons self-identifying (heads of household) as Hispanic or linked to a country (via ancestry and/or birth) that is considered “Hispanic” are included. With this sample, we are able to isolate all that would be considered Hispanic, whether or not they self-identity as Hispanic.

The socioeconomic variables used in this analysis are income, socioeconomic status, occupational prestige, and education. Each of these variables measures conceptually-distinct indicators of social “advancement” that are tied to assimilation and integration into society, although they are highly correlated. Personal income is provided in the IPUMS dataset. The IPUMS dataset also provides an indicator of socioeconomic status called the “Duncan Socioeconomic Index.” Socioeconomic status (SES) is calculated through a weighted sum of occupational education and occupational income with the weights derived through perceptions of occupational prestige (see Duncan 1961). Additionally, IPUMS provides an indicator of occupational prestige (based on Nakao and Treas 1994), which is a weighted average of ratings

of various occupations based on GSS survey data.³ Each of these three measures is logged to deal with skewed data. Education is transformed into an ordinal variable, with less than high school, high school or equivalent, some college, bachelor's degree, or higher than bachelor's degree as the categories.

Percent Hispanic in the area is calculated using the percentage of heads of householders that are Hispanic within the respective Public Use Microdata Area (PUMA). "White spouse" was positively coded if a head of householder's spouse was identified as white, non-Hispanic. Sex is coded 1 for males when used as a dummy variable.

Racial/Ethnic Classifications

The "honorary white" category is coded 1 for those who are "White, Hispanic." These are the persons who are Hispanic, but may be light-skinned or are closer to advancing into the white category. The "collective black" category comprises those who identify as "Non-White, Hispanic." These persons have neither left the Hispanic label nor entered into a whiter category.

"Hispanic non-Hispanics" – i.e. those who do not self-identify as Hispanic but would be so classified by OMB – were persons who claimed first ancestry (hereafter "first ancestry" will be referred to as "ancestry") and/or birth in a Spanish-speaking country (excluding Spain) without checking the Hispanic item on the Census. Additionally, the following persons were included, although their person-weights were divided by 2 under the assumption that only half of them (a low assumption) represented our target ethnicity: those who claimed birth, but not ancestry, in a Spanish-speaking country; those who were native-born but claimed said ancestry;

³ For more information on the Duncan Socioeconomic Index, one may refer to <http://usa.ipums.org/usa-action/variableDescription.do?mnemonic=SEI>. More information on the Nakao and Treas occupational prestige score can be found at <http://usa.ipums.org/usa-action/variableDescription.do?mnemonic=PRENT>.

and those who did not list an ancestry but were born in said countries (again, Spaniards were excluded).

Those who were identified as “Hispanic non-Hispanics” and white (effectively the White, Non-Hispanic persons in our sample) were classified as “white” on the triracial hierarchy. These are the persons who left the Hispanic category and assimilated into whiteness, despite having origins in Latin America. The “Hispanic non-Hispanic” non-whites, on the other hand, will be classified as “ambiguous ethnics” for Table 1.1.

DESCRIPTIVE STATISTICS OF VARIABLES BY RACIAL/ETHNIC GROUP

[TABLE 1.1]

Table 1.1 provides summary statistics of the variables included in this study by the racial category. As displayed, all of the indicators of socioeconomic advancement/assimilation are related to the racial hierarchy. As persons advance up the hierarchy, the scores of the socioeconomic variables similarly increase (supporting H2). Income, for example, averages at \$23,823 for collective blacks, goes up to \$28,892 for honorary whites, and jumps up to \$39,847 for whites. Education displays a similar pattern, as 49.3% of collective blacks do not have a high school education, while the corresponding figure for honorary whites is 41.6% and for whites is 21.7%. Post-high school education also increases across the racial categories. The same story can be told for socioeconomic status as well as occupational prestige as both variables’ scores increase across the racial hierarchy.

The ambiguous ethnic group has consistently higher averages on the socioeconomic variables mentioned than does the collective black (supporting H5). For the bulk of the socioeconomic variables, the scores of the ambiguous ethnic group even surpass those of the honorary white group. It is only in personal income that the selective ambiguous ethnic group

trails behind honorary white group. On all of these measures, whites score higher than the rest of the groups, further affirming their placement on the racial hierarchy.

Other variables also have the expected relationship with the racial hierarchy: lower rates of foreign-born persons, higher rates of English speakers at home (H3), and higher numbers of years in the country (for the foreign-born) are all associated with advancement up the racial hierarchy. The idea that foreign-born persons arrive in the United States with national-based identities and subsequently learn to identify as Hispanic is not confirmed in this analysis, as the Hispanic non-Hispanics average higher lengths of stay in the United States.

The idea that the Hispanic identity is a linguistic category is supported by the figures on language spoken at home. Those identifying as Hispanic are far less likely to speak English at home, with 80% of honorary whites and 84% of the collective black not speaking English at home. In contrast, 60% of the white group and 56% of the ambiguous ethnic group speak English at home.

Social context/community is shown to have a strong relationship with ethnic identity. For example, persons identifying as Hispanic live in areas with higher Hispanic concentration (around 30% on average), while the Hispanic non-Hispanics live in areas of less than 20% Hispanic on average (H7). The “Hispanic” identity may, thus, also be a product of a cultural base that is cultivated through networks, grocery stores, and restaurants that reaffirm a collective “Latino” feeling of “us.” Perhaps the most striking finding across the groups deals with marriage, wherein the white group overwhelmingly marries fellow non-Hispanic whites (82%) compared to the other categories with said marriage rates between 10 and 17% (H1). Similarly, over 80% of Hispanics marry other Hispanics, whereas over 84.9% of non-Hispanic whites and 73.3% of non-Hispanic non whites do not marry other Hispanics. As argued by Lieberman and Waters

(1993), married persons often simplify their ancestries to match up with those of their partners.

This finding suggests that marriage with non-Hispanic whites may be one mechanism for assimilating out of the Hispanic category and into the white category.

SOCIAL MOBILITY AND THE RACIAL HIERARCHY, MULTIVARIATE MODELS

[TABLE 1.2]

Table 1.2 displays the results of models that predict socioeconomic advancement. In all multivariate models, the triracial hierarchy retains the expected effects on the indicators of socioeconomic status (H2). Being part of the white category is associated with universally higher scores on all four dependent variables when compared with membership in any of the other racial categories. Being part of the honorary white category tends to lead to higher scores than does being part of the (omitted) collective black category.

Indicators of acculturation also have expected relationships with the dependent variables (H4). Specifically, being native born and speaking English at home both lead to high socioeconomic outcomes. Educational attainment is positively related to income. Tautologically, higher levels of education are also related to having higher scores on socioeconomic status and occupational prestige, although in these cases education served as a control variable since these variables are partially functions of education. Regarding the other control variables, age is shown to have almost a negligent (yet significant) effect on all dependent variables, although age has a stronger negative effect on education. Males are favored in terms of income and occupational prestige (mildly), although females are favored in socioeconomic status and education.

MOVING SIDEWAYS: THE SOCIAL CONTEXTUAL FACTORS OF THE HISPANIC IDENTITY

[TABLE 1.3]

Model 1 on Table 1.3 revisits the socioeconomic explanation for identification with the Hispanic ethnicity, controlling for number of years in the country, age, and sex. All socioeconomic variables are demonstrated to be negatively related to identification with the Hispanic ethnicity, decreasing probabilities of identification as income (H5) and educational attainment (H6) increase. Years of residence in the country is also negatively related with Hispanic identification, as more recent arrivals are more likely to identify as Hispanic.

Model 2 introduces racial dummy variables and two social context variables in English spoken at home and percent Hispanic in PUMA. The socioeconomic status variables decrease their predictive power but remain in the same direction. The social context variables added are also significant, as increasing the Hispanic population in the area is positively related to identifying as Hispanic. Speaking English at home has a large effect in suppressing identification as Hispanic (H9), suggesting that the linguistic character of the Hispanic category can be activated through language use in the household. Strikingly, racial variables have tremendously large effects on the identification with the Hispanic ethnicity. Identifying as “some other race” increases the odds of identifying as Hispanic by 35.78 times (H7), while identifying as blacks or Asians decrease such odds (at a factor of .33 and .28, respectively, H8). This suggests that “some other race” is indeed a proxy for the identification as a “racial Hispanic” and that the Hispanic ethnic identity is ambiguous for blacks and Asians.

Model 3 is a full model that includes all of the variables. The interpretation of the relationships for the revisited variables is tricky, as one must keep in mind that race, spousal ethnicity, and years of residence are all controlled for. This caveat said, the full model reveals a

reversed positive relationship wherein the income and education increase identification with the Hispanic ethnicity when controlling for all newly introduced variables (adding context to hypothesis H5). Under the same controls, the foreign born are demonstrated to be more than half as likely to identify as Hispanic as seen in the numbers of years in the U.S. variable.

What Model 3 highlights is the importance of two factors that influence identification as Hispanic: race and social context. By far, race continues to have the strongest effect on Hispanic identification. Most of the social context variables also have strong effects on ethnic identification, most notably in spousal ethnicity/race where being married to a Hispanic increases the odds of identifying as Hispanic by a factor of 17, and having a white non-Hispanic spouse reduces such by a factor of .79 (H11 and H10, respectively). Speaking English at home also decreases the odds of identifying as Hispanic by one-half. The full model thus qualifies the effects that socioeconomic status and nativity have on the Hispanic ethnicity, while demonstrating that the ideas underlying the ambiguous ethnicity (i.e. race and social context) have very strong and robust effects on identifying as Hispanic.

CONCLUSION

This chapter served to critically explore the Hispanic ethnicity as defined by the OMB. The history of the ethnicity is rife with politics, and in the end the Hispanic ethnicity was deemed to be distinct from the existing racial categories. Multivariate analyses demonstrated that the Hispanic ethnicity is largely a racial identity for persons not identifying as white. Those identifying as “some other race” were much more likely to identify as Hispanic than were those identifying with a more “salient” racial category (i.e. Asian and black), suggesting that “some other race” is often a proxy for “Hispanic race.”

Multivariate analyses lend support to the theory on the “triracial hierarchy” for Hispanics, demonstrating that indicators of socioeconomic assimilation are related to advancement of a racial/ethnic hierarchy. Bivariate statistics also demonstrate that other indicators of assimilation including having a white (non-Hispanic) spouse, speaking English at home, having nativity, and residing in the United States for longer periods of time are also related to the racial hierarchy. Furthermore, multivariate analyses suggest that assimilating from a Hispanic category to the non-Hispanic white category is also associated with the above-mentioned variables. This dynamic may also be (consciously or not) politically motivated, as the identification with a minority label may be a means to signal to the public that inequality continues to exist for Hispanics. Identification as Hispanic to highlight inequalities may be used to further governmental support for minority groups, or just to set one’s self apart from the dominant group.

While the relationships posited under the tri-racial hierarchy are supported, the identification as Hispanic in general is more strongly supported by racial and social dynamics. For Hispanics in general (i.e. both whites and non whites), the idea of “assimilating” out of the Hispanic ethnicity through income and education is only partially supported, depending on the control variables included. Hispanic identification is more powerfully predicted by race and social context. Specifically, “some other race” is shown to be a proxy for the image of the “classic,” racial Hispanic, as they are the most likely to identify as Hispanic. Blacks and Asians, on the other hand, are very likely to *not* identify as Hispanic despite their place of birth and ancestry. Additionally, the social context in which one lives, in this case exemplified by the ethnicity of one’s spouse and the language that one is surrounded by, has the greatest effect on the identification with the Hispanic ethnicity.

DISCUSSION

This chapter would like to propose that ethnicity is often ambiguous and fluid, being defined, pushed, and pulled from many directions by several forces. When caught within two ethnicities, social forces can lead to the abandonment of one ethnicity in favor of another, particularly when one ethnicity is ambiguously defined and not in alignment with the organically emerging identities of the community. In this case, a public institution defines Hispanic as a linguistic/ancestral ethnicity, although in doing so it uses a term that is often racialized. While persons who view themselves racially as Hispanic (i.e. those identifying as “some other race”) have no trouble identifying as Hispanic, those who view themselves as a race distinct from Hispanic receive contradictory messages from society that tell them that they do not fit the mold of the “classic” racial Hispanic. Tipping points exist that influence these ambiguous ethnics to abandon the Hispanic ethnicity, such as marriage with a non-Hispanic. The subsequent family that is created changes one’s own self-perception, highlighting one’s Hispanic origins when married to a fellow Hispanic, or suppressing one’s Hispanic identity when not married to a Hispanic.

This case on the Hispanic ethnicity highlights a social dynamic wherein a government-defined ethnicity is suppressed or highlighted due to one’s social context. The Dominican with African roots or the Peruvian with Asian roots may identify as Hispanic when speaking Spanish at home, or simply as black or Asian when treated as such in social settings. Similarly, light-skinned Argentinians may identify as white if raising a family with a white spouse, or as Hispanic if routinely eating empanadas with the family. The opposite dynamic may also occur wherein persons with a categorical ethnicity that lies dormant in normal day-to-day functions may suddenly have their categorical ethnicity activated due to governmental pressures. Yugoslavians

who suddenly become Serbian through nationalist politics, Rwandans who suddenly become Tutsis through ethnic mobilization, or even Californians that become Japanese through internment policies are examples in which political action can serve as ethnic “tipping points” that define a person’s identity.

Perhaps even more generally, the idea of the ambiguous ethnicity can help to elucidate paths toward identity formation. Clearly, every person is born without a concept of one’s own race and ethnicity. He/she learns her race/ethnicity – not to mention the definitions of and distinctions between race and ethnicity – through social processes. For every person, classic racial definitions in the culture, ethnic definitions of the governmental institutions, and one’s family’s identity are all influences in one’s self-perception. This specific case demonstrates the power of one’s family and culture in serving as “identity tipping points,” although in other cases a government’s policies can surely a stronger influence, or multiple salient identities can be held simultaneously under the right conditions.

Given the political motivations for including the Hispanic item on the Census as well as the political implications of identifying as Hispanic, this governmentally-defined ethnicity can be brought to life when it highlights the struggles and experiences that a group faces. Thus, a government’s incentives alongside a mobilized public can serve to construct an identity. The political history of the construction of an ethnic definition, as well as the political uses of such a definition, is therefore another factor that influences identities. One must not forget that without the specific political history, context, and implications that the U.S. underwent in collecting racial/ethnic data, perhaps the notion of race and ethnicity would be a different animal in the U.S. (which would also mean that we would not even be discussing DA by racial/ethnic groups). A contrasting case-study to be reserved for future research may compare

and contrast the political histories of race/ethnicity in Brazil and the U.S. to explain the contrasting conception of race/ethnicities of the two countries (let alone any other country).

Painting portraits of the diverse Hispanic population can lead to biases when one considers the sample of Hispanics who are (not) captured in the Census. Duncan and Trejo (2007, 2008) emphasize that the socioeconomic achievement of Mexicans may be downplayed due to the high-performance of persons who do not identify as “Hispanic, Mexican,” despite having their ancestry from Mexico. This research re-emphasizes this point, as socioeconomic status often leads white Hispanics to rescind identification with their Hispanic ancestral roots. This research also highlights the racial diversity of Hispanics that is missed in the Census, as Hispanics who are neither “some other race” nor white often do not identify as Hispanic. Generations of intermarriage, socioeconomic advancement, and acculturation will make Hispanics an increasingly difficult population to define, trace, and count.

REFERENCES

- Alba, Richard and Tariqul Islam. 2009. "The Case of Disappearing Mexican Americans: An Ethnic-Identity Mystery." *Population Research Policy Review* 28:109-121.
- Bean, Frank and Marta Tienda. 1987. *The Hispanic Population of the United States*. New York NY: Russel Sage Foundation.
- Bonilla-Silva, Eduardo. 2004. "From Bi-Racial to Tri-Racial: Towards a New System of Racial Stratification in the USA." *Ethnic and Racial Studies* 27:931-950.
- Bonilla-Silva, Eduardo and Karen Glover. 2006. "'We Are All Americans': The Latin Americanization of Race Relations in the United States." Pp. 149-186 in *The Changing Terrain of Race and Ethnicity*, edited by Maria Krysan and Amanda Lewis. New York, NY: Russell Sage Foundation.
- Brunsma, David. 2005. "Interracial Families and the Racial Identification of Mixed-Race Children: Evidence from the Early Childhood Longitudinal Study." *Social Forces* 84:1131-1157.
- Campbell, Mary and Christabel Rogalin. 2006. "Categorical Imperatives: The Interaction of Latino and Racial Identification." *Social Science Quarterly* 87:1030-1052.
- Chapa, Jorge. 2000. "Hispanic/Latino Ethnicity and Identifiers." Pp. 45-48 in *Encyclopedia of the US Census*, edited by Margo Anderson. Washington, DC: CQ Press.
- Choldin, Harvey. 1986. "Statistics and Politics: The 'Hispanic Issue' in the 1980 Census." *Demography* 23:403-418.
- Duncan, Brian, and Stephen J. Trejo. 2007. "Ethnic Identification, Intermarriage, and Unmeasured Progress by Mexican Americans." Pp. 227-269 in *Mexican Immigration to the United States*, edited by George J. Borjas. Chicago, IL: University of Chicago Press.
- Duncan, Brian and Stephen Trejo. 2008. *Ancestry versus Ethnicity: The Complexity and Selectivity of Mexican Identification in the United States*. Discussion Paper No. 3552, Forschungs institut zur Zukunft der Arbeit Institute for the Study of Labor, Bonn, Germany.
- Duncan, Otis D. 1961. "A Socioeconomic Index for All Occupations." Pp. 109-138 in *Occupations and Social Status*, edited by A. Reiss. New York, NY: Free Press.
- Durand, Jorge, Douglas Massey, and René Zenteno. 2001. "Mexican Immigration to the United States: Continuities and Change." *Latin American Research Review* 36(1):2001.
- Emeka, Amon. 2008. *Who are the Hispanic Non-Hispanics and Why Do They Matter? Toward a More Inclusive Definition of the Hispanic Population*. Presented at the Annual Meeting of the American Sociological Association, August 3, Boston, MA.

- Farley, Reynolds. 1991. "The New Census Question about Ancestry: What Did It Tell Us?" *Demography* 28(3):411-429.
- Freire, Paulo. 1993. *Pedagogy of the Oppressed*. New York: Continuum Books.
- Golash-Boza, Tanya. "Dropping the Hyphen? Becoming Latino(a)-American through Racialized Assimilation." *Social Forces* 85:27-55.
- Gordon, Milton. 1964. *Assimilation in American life: The Role of Race, Religion and National Origins*. New York NY: Oxford University Press.
- Greico, Elizabeth and Rachel Cassidy. 2001. *Overview of Race and Hispanic Origin, Census 2000 Brief*. Washington, DC: U.S. Census Bureau.
- Hirschman, Charles, Richard Alba, Reynolds Farley. 2000. "The Meaning and Measurement of Race in the U.S. Census: Glimpses into the Future." *Demography* 37(3):381-393.
- Lee, Sharon and Sonya Tafoya. 2006. "Rethinking US Census Racial and Ethnic Categories for the 21st Century." *Journal of Economic and Social Measurement* 31:233-252.
- Lieberson, Stanley and Mary Waters. 1993. "The Ethnic Responses of Whites: What Causes Their Instability, Simplification, and Inconsistencies?" *Social Forces* 72:421-451.
- Memmi, Albert. 1991. *The Colonizer and the Colonized*. Boston, MA: Beacon Press.
- Mendieta, Eduardo. 2000. "The Making of New Peoples: Hispanicizing Race." Pp. 45-60 in *Hispanics/Latinos in the United States: Ethnicity, Race, and Rights*, edited by Jorge Gracia and Pablo de Greiff. New York NY: Routledge.
- Nakao, Keiko and Judith Treas. 1994. "Updating Occupational Prestige and Socioeconomic Scores: How the New Measures Measure Up." *Sociological Methodology* 24:1-72.
- National Council of La Raza. 1995. "Accurate Racial/Ethnic Data Should Drive Category Review." *Poverty and Race* 4:7-9.
- Navarro, Mireya. 2007. "Going Beyond Black and White, Hispanics in Census Pick 'Other'." Pp. 214-217 in *Race, Class, and Gender in the United States*, edited by P. Rothenberg. New York, NY: Worth Publishers.
- Neidert, Lisa and Reynolds Farley. 1985. "Assimilation in the United States: An Analysis of Ethnic and Generation Differences in Status and Achievement." *American Sociological Review* 50:840-850.
- Park, Robert. 1950. *Race and Culture*. Glencoe, IL: The Free Press.
- Portes, Alejandro and Ruben Rumbaut. 2001. *Legacies: The Story of the Immigrant Second Generation*. Berkeley, CA: University of California Press and Russell Sage Foundation.

Qian, Zhenchao. 2004. "Options: Racial/Ethnic Identification of Children of Intermarried Couples." *Social Science Quarterly* 85:746-766.

Rodríguez, Clara. 2000. *Changing Race: Latinos, the Census, and the History of Ethnicity in the United States*. New York, NY: New York University Press.

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. *Integrated Public Use Microdata Series: Version 5.0*. Minneapolis, MN: University of Minnesota, 2010.

Siegel, Jacob and Jeffrey Passel. 1979. "Coverage of the Hispanic Population of the United States in the 1970 Census." *Current Population Reports: Special Studies: Series P-23*; no. 82, Washington, D. C.: U.S. Bureau of the Census.

Snipp, Matthew and Juanita Tamayo Lott. 2009. "Population Shifts and Demographic Methods." *Journal of Official Statistics* 25(1):99–124.

Tafoya, Sonya. 2004. *Shades of Belonging*. Washington, DC: Pew Hispanic Center.

TABLES

Table 1.1: Select Descriptive Statistics by Hispanic Tri-Racial Classification Group

		NonHispanic White (White)	Hispanic, white (honorary white)	Hispanic, NonWhite (collective black)	NonHispanic, NonWhite (Ambiguous Ethnics)
Personal Income	Mean	39,847	28,892	23,823	27,652
	Median	27,700	20,000	18,900	20,000
Education, %	no high school	21.7	41.6	49.3	32.7
	high school	22.4	21.3	22.2	23.1
	some college	30.7	22.6	20.7	28.2
	Bachelors	15.7	8.6	5.4	10.5
	> bachelors	9.4	5.9	2.6	5.6
Socioeconomic Status	Mean	38.6	30.3	27.2	31.7
	Median	44.0	19.0	18.0	24.0
Occupational Status	Mean	38.1	33.6	32.6	34.0
	Median	39.6	34.7	33.9	35.3
Foreign Born, %	Yes	44.0	57.2	63.2	61.4
	No	56.0	42.8	36.8	38.6
English Spoken at Home, %	Yes	59.6	19.9	16.3	44.2
	No	40.4	80.1	83.7	55.8
Spouse, % (sample: married persons)	Nonwhite	17.9	83.4	90.2	84.9
	White	82.1	16.6	9.8	15.1
Spouse, %	Non Hispanic	15.1	81.9	86.9	26.7
	Hispanic	84.9	18.1	13.1	73.3
Years in Country (sample: foreign- born)	Mean	28.7	21.2	18.7	21.9
	Median	27.0	19.0	17.0	20.0
Age	Mean	43.3	44.6	39.9	43.3
	Median	40.0	42.0	38.0	40.0

Sex	Male	66.6	68.3	66.6	56.7
	Female	33.4	31.7	33.4	43.3
Percent Hispanic in PUMA	Mean	17.7	31.9	28.4	19.0
	Median	9.4	24.0	23.0	12.3
racesimplified	White	100.0	100.0	0.0	0.0
	Black	0.0	0.0	4.1	49.4
	AI/AN*	0.0	0.0	2.0	9.3
	Asian	0.0	0.0	0.5	9.0
	Pacific Islander	0.0	0.0	0.2	1.2
	Multiracial	0.0	0.0	11.8	25.2
	Some other race	0.0	0.0	81.4	5.9
TOTAL		187,833	4,086,943	4,152,565	54,881

*AI/AN is short for American Indian or Pacific Islander

Table 1.2: Predictors of Socioeconomic Status Indicators for Hispanics, Multivariate Regression

	Linear Regression			Ordinal Logistic Regression, odds ratios
	Dependent Variables			
	Income (logged)	Socioeconomic Status (logged)	Occupational Prestige (logged)	Education
<i>Racial Category (excluded category "collective black")</i>				
White	0.145	0.136	0.048	2.557
honorary white	0.041	0.060	0.023	1.502
<i>Acculturation</i>				
Native	0.052	0.174	0.066	2.266
English	0.067	0.021	0.006	1.514
<i>Education</i>				
Highschool	0.317	0.276	0.085	
Somecollege	0.555	0.592	0.216	
Bachelors	0.875	0.968	0.410	
>bachelors	1.079	1.135	0.538	
<i>Control</i>				
age/10	0.002	-0.006	0.001	0.851
Male	0.572	-0.119	0.024	0.960

All variables are significant at the .000 level.

Table 1.3: Multivariate Models Predicting the Identification with the Hispanic Ethnicity (odds ratios displayed)

	Model 1	Model 2	Model 3
<i>Socioeconomic Status</i>			
Loginc	0.89	0.95	1.03
Highschool	0.55	0.77	1.07
Somecollege	0.45	0.70	1.30
Bachelors	0.31	0.57	1.13
> bachelors	0.32	0.59	1.25
<i>Race (omitted: white)</i>			
other race		35.78	29.26
Black		0.33	0.34
Asian		0.28	0.35
<i>Social Context</i>			
spouse Hispanic			17.08
spouse white			0.79
percent Hispanic in PUMA		1.18	1.01
English at home		0.28	0.49
<i>Acculturation (number of years in country, omitted: native-born)</i>			
0-12	1.77		0.43
13-24	1.65		0.46
25+	0.87		0.43
<i>Control</i>			
Age	0.97	0.95	1.03
Male	1.14	1.03	1.03

*all variables are significant at the .000 level except for "male" in the last model (.003)

**percent Hispanic is divided by 10

CHAPTER 2

EMIGRATION AND UNDERCOUNT OF THE HISPANIC POPULATION, 1990 – 2000

In a world of perfect data, the estimation of the population size is simple and can be pinpointed with the use of the Fundamental Demographic Equation (FDE), wherein the population size equals births, minus deaths, plus immigration, and minus emigration. The FDE could also be rearranged to estimate the size of the components of the FDE, as emigration can be the calculated as the difference between an initial population size and a survived population (as well as net immigration, if one would wish to account for new arrivals). This method of estimating emigration is referred to as the “residual method” and has been a popular method for estimating emigration in previous decades (see Ahmed and Robinson 1994; Mulder, Guzmán, and Brittingham 2002; Warren and Peck 1980).

Unfortunately, the residual method for estimating counts of emigrants relies on perfect data that, frankly, do not exist. Imperfect data has resulted in the estimation of emigration through the use of residual methodologies that are implausible. Research conducted at the Census Bureau produces estimates of population components that are often considered to be the “gold standard,” particularly due to the detailed data and investigative resources to which the Bureau has access. However, many of these residual estimates of emigration calculated by the Census Bureau have produced questionable results, including many “negative emigration” counts (Ahmed and Robinson 1994; Mulder et al. 2002). The clear culprit that prevents the residual method from producing reliable emigration estimates results from the undercount of the Hispanic population. In this case, Hispanics in 1990 were estimated to be undercounted at higher proportions than were Hispanics in 2000 (Mulry 2006).

Fortunately, recent scholarship has estimated emigration rates by using sample-based methodologies. Less affected by undercount strictly-defined, the sample-based methodologies estimate emigration rates based on longitudinal data that collect data on the same individuals (or households) over time. The first objective of this research is dedicated toward using these studies in order to estimate emigration counts for Hispanics by age across 1990 – 2000. The second objective of this research is to uncover sources of error for emigration estimates, paying particular attention to the undercount of Hispanics in the 1990 Census. It is argued that age, year of immigration, and the political context in which the 1990 Census was embedded have strong effects on the Census Bureau's ability to accurately count Hispanics.

RESIDUAL ESTIMATES OF EMIGRATION AND THEIR LIMITATIONS

An early use of residual methods in estimation emigration for the U.S. was performed by Warren and Peck (1980), who estimated emigration counts over 1960 – 1970 by age and sex for the foreign born. Although these scholars did not estimate emigration rates by racial or ethnic detail, the residual method performed reasonably well and produced overall emigration estimates of 5.2% for the stock of foreign-born persons in the U.S. in 1960 for the decade, as well as an 18% estimated emigration rate for those who entered the U.S. between 1960 and 1970. Despite these emigration estimates, emigration was found to be negative for the youngest age group.

Emigration for the documented foreign-born population has been estimated in earlier decades by a few scholars who recognized the need to modify (i.e. "clean") the available data. Using INS and Census data, Borjas and Bratsberg (1996) estimate emigration rates for the documented immigrant population across 1970 – 1980 using a modified residual method. Adjusting the data for temporary and undocumented migrants, the authors estimate emigration

rates for the documented populations from the Americas at around 25 – 35% (depending on the region of origin and year of entry). Additionally, Warren and Passel (1987) adjust Census data by dealing with misreporting of country of birth, citizenship, and nativity status, and also adjust INS data by underreporting. These authors estimate yearly emigration for the foreign-born at around 100,000.

Later emigration estimates would attempt to introduce greater demographic detail, revealing problems in the data components that make the use of the residual method untenable for producing detailed emigration estimates. Ahmed and Robinson (1994) used the residual method to produce emigration rates by race and ethnicity for the foreign-born through 1980 – 1990. In doing so, the authors were restricted in the samples that could be used to estimate emigration, as cohorts from several countries produced negative emigration counts. Due to this problem, emigration was only estimated for “selected countries,” and emigration rates for the Hispanic non-selected countries was assumed to be half of the emigration rates for the Hispanic selected countries. Over the decade, the emigration rate for Hispanics was estimated to be 7%.

Mulder et al. (2002) also attempted to estimate emigration by demographic detail for the foreign-born by using the residual method for 1990 – 2000. Like Ahmed and Robinson (1994), this study produced many negative emigrants that rendered the authors to state that “the foreign-born emigration statistics and estimates that exist are of questionable quality.” These authors were unable to produce emigration counts in which they had confidence within the time allotted to them.

There are a number of problems with using a residual method for estimating emigration, all of which stemming from an underlying fundamental flaw: data are imperfect. Two underlying assumption with the data used (as well as basically all data) are that the

sampling error (or Census error, in this case) and non-sampling error are minimal-to-nonexistent (Judson and Swanson 2011). These assumptions become challenged as differences in Census coverage and changing questionnaire items influence the quality of the data. On a related note, the estimates of the various components of population change are assumed to be accurate, essentially lodging different data sources together into the same puzzle; each data source with attendant differences in data collection processes (Judson 2006; Judson and Swanson 2011). In this particular case, the most egregious source of error stems from the undercount of persons at the initial time point, which ends up producing errors of large enough magnitude to produce large counts of negative emigration.

This research also suggests that the political climate and respondent fears have influenced the undocumented populations to be more wary of being counted in the Census prior to the passage of the Immigration Reform and Control Act of 1987 (IRCA), which would have granted many undocumented immigrants legal status by the time of the 1990 Census (but not in time for previous Censuses). This idea is supported by the fact that the persons from the countries who have had the highest rates of legal-status adjustment through IRCA were the same persons who were more highly counted in 1990 as opposed to 1980 (Ahmed and Robinson 1994). While the Census undercount of the undocumented immigrants is uncertain, it is very likely to be much higher than their documented counterparts (Warren and Passel 1987).

When constructing emigration rates by demographic detail, even larger problems emerge apart from coverage error. In identifying emigration exclusively for “Hispanics,” for example, one would run into problems regarding how people are identified as “Hispanic.” In earlier decades, the methods for identifying Hispanic populations on the U.S. Decennial Census have ranged from a “Mexican” category to linguistic affiliation, before being included in a self-

identifying item for a 5% sample in 1970. Furthermore, the differences between the Hispanic self-identifier across the Censuses have likely produced large differences in response rates for Hispanics. For example, the Hispanic item in 1990 appeared three items *after* the race item, whereas in 2000 the Hispanic item directly preceded the race item. This resulted in a much lower non-response rate for the Hispanic item in 2000, although evidence suggests that this did not lower the proportions of persons reporting as Hispanic (Martin 2002).

A NEW GENERATION OF EMIGRATION ESTIMATES

In light of the problems inherent in Census data, it is evident that the residual method will be insufficient in estimating emigration precisely. Although other methods have been used to estimate emigration rates, many of these studies are not able to be reproduced for the 1990s. Fernandez (1995), for example, used foreign Census and U.S. State Department data to estimate emigration rates for the native-born population, producing a total emigration count of 48,000 per year. Subsequent attempts to replicate this methodology raised serious questions regarding the accuracy of any resulting emigration estimates, cautiously producing a count of 18,000 native-born emigrants per year through the 1990s (Gibbs et al. 2003).

Massey, Durand, and Malone (2002) use their own Mexican Migration Project data, which is collected from semi-structured interviews of select populations of a snowball sample, taking a slightly anthropological life-course approach to collecting data. These authors find very high rates of emigration for Mexicans that vary greatly by year of entry, documentation status, and sex, producing two-year rates of emigration that range from 20% and 60% over 1966 to 1985. The in-depth approach of the Massey et al. (2002) study produces what is perhaps the most rich and valid data available, although it is limited in its sample (i.e. restricted to Mexicans within their sampling frame).

Woodrow-Lafield (1995) used select Current Population Survey (CPS) data from 1987, 1988, and 1989 to estimate ranges of emigration by using an item that asks whether any immediate relatives emigrated from the U.S. Using this familial-network approach to estimating emigration, Woodrow-Lafield (1995) estimates “stock” emigration by region of birth, which includes a high estimate for Mexico (as an example) at 43.8% in 1988 and a low estimate of 21.9% in 1989. Unfortunately, such CPS items on emigration have not been available for the 1990s.

The previous emigration estimates are notable in their novelty and their methodologies that do not require the use of the Decennial Census. However, for numerous reasons, such methodologies are not immediately replicable for more recent years, let alone for the Hispanic population in specific. Fortunately, recent scholarship has produced new sample-based estimates of rates of emigration for the 1990s. Two of these alternate emigration estimates utilized matched CPS data to produce emigration rates for the foreign-born population. Van Hook et al. (2006) used a complex methodology to estimate emigration which is briefly described as follows. For matched households that were not followed-up in subsequent years, probabilities are assigned to those who migrated internally, emigrated from the country, died, or were not followed up for other reasons.⁴ Proportions emigrating were calculated as a function of the other components, and these proportions were applied to the weighted CPS sample to estimate rates and numbers of emigrants. Return migration was estimated using

⁴ Proportions of persons not followed-up that were assumed to be internal migrants were estimated using the “residence one year ago” item, which counts the numbers of persons who have internally migrated in the past year. Deaths were estimated using the National Health Interview Survey. Persons not followed up for other reasons were taken from native born rates of “other” non follow up reasons. By assuming that emigration for the native-born was 0, rates of non-follow-up for “other” reasons were estimated using residual methods from the same equation used to divide the components of non-follow-up. In other words, non follow-up (NFU) for “other reasons” was equal to the Total NFU minus NFU for deaths and internal migration.

Massey et al. (2002), and was subtracted from the total estimated emigration in order to calculate net emigration in a manner that accounts for circular migration. Based on Fernandez (1995) and Gibbs et al. (2003), emigration for native-born Hispanics was assumed to be negligible (0). Using a similar approach for estimating emigration rates, Passel et al. (2006) accounted for non-responses on follow-up CPS surveys by death (using life tables), internal mobility (using estimates provided by the CPS), matching-processing error (derived from a multivariate model), and emigration (the remaining probability).

Schwabish (2009) used three linked administrative data sources provided by the Social Security Administration (i.e. the Detailed Earnings Records, Numerical Identification System, and Master Beneficiary Record) to identify foreign-born work histories. Workers who report earnings for at least one year and subsequently have at least two years of unemployment are considered to have emigrated from the country. However, due to the sample used by Schwabish, emigration rates are unable to be estimated for the undocumented population. To simplify references to these sources of emigration, future references to these estimates will also be collectively referred to as the “alternate estimates of emigration,” or individually as the “Van Hook,” “Passel,” and “Schwabish” estimates, respectively.

THE CURRENT STUDY

The objectives of this study are twofold. First, emigration by age for Hispanics will be estimated for the 1990s using a basic residual methodology as well as by using the emigration rates calculated by the three studies described above, namely Van Hook et al. (2006), Passel et al. (2006), and Schwabish (2009). Creating emigration through the use of the residual methodology will highlight the problems that undercount plays in making residual estimates of emigration

unreliable. Second, undercount will be estimated by age, and explanations will be offered to explain the high undercount of Hispanics for the 1990 Census.

This study will argue that an important factor that has negatively affected the response rates for the Census stems from the historical political context. The year 1990 was situated in a metaphorical political battle wherein governmental policies directly compromised the welfare of a large portion of the immigrant population. The Immigration Reform and Control Act (IRCA) of 1986 increased the policing of the border while placing restrictions on the ability for the foreign-born to work, requiring documentation of eligibility to work while imposing sanctions on employers who hire undocumented workers. Persons who have been working in the U.S. for four years without documentation were able to apply for amnesty, but workers who arrived more recently were effectively criminalized. In light of this historical event, it is argued that the quality of the Census in 1990 had suffered, as many foreign-born persons would have felt threatened by what could have been interpreted as a means by which the government can identify the “illegals.”

DATA

The Census Bureau’s American Factfinder tool was used to provide total counts of Hispanics for 1990 and 2000. Within the age groups provided by American Factfinder, the counts of Hispanics for single-age groups were distributed from the corresponding proportions found in the Integrated Public Use Microdata Sample (IPUMS, Ruggles et al. 2010). The IPUMS files were also used to provide counts by year of entry. For the 2000 IPUMS data, person-weights for those who reported immigrating into the U.S. in 1990 were multiplied by one-fourths, as it would be assumed that one-fourth of the reported 1990 immigrants would have resided in the U.S. on April 1, 1990 (which is the official date of the Census and one-fourth into the 1990 year). These

calculated immigrants would be subtracted from American Factfinder's Census 2000 count of total Hispanics (distributed to single-years of age through the use of IPUMS age-distributions) to produce 2000 counts of the cohorts that were present in the U.S. for the 1990 Census.

Deaths, used to create survivorship ratios, are gathered from the National Center for Health Statistics' (NCHS) Multiple Cause of Death micro-data file. Year of death has been re-coded to Census years beginning on April 1. Deaths were adjusted for ethnic misclassification using the "misclassification ratios" estimated by Arias et al. (2008)⁵. Deaths to persons reporting residence outside of the U.S. will not be included.

Death data have not been collected for all states throughout all of the 1990s. In 1990, data on Hispanic origin were not collected by Louisiana, New Hampshire, and Oklahoma. Oklahoma and New Hampshire continued to lack data on Hispanic origin for 1991 and 1992, and Oklahoma did not include data on Hispanic origin until 1997. While these states only house a small proportion of Hispanics, it is nonetheless advisable to account for this gap in the data. It will be assumed that the mortality rates for Hispanics in states with a missing Hispanic origin item on the death certificate mirror the mortality rates for Hispanics in states where a Hispanic origin item was included on the death certificate. Using mortality rates for states that include a Hispanic death certificate item, Hispanic deaths will be proportionally allocated to states that lack a Hispanic death certificate item thusly:

$$H.Deaths_{missing,year,age} = \frac{H.Deaths_{nonmissing,year,age}}{H.Count_{nonmissing,year,age}} \times H.Count_{missing,year,age} \quad [2.1]$$

⁵ While Arias et al. (2008) estimated death *overcount* for Hispanics under the age of 25, this research will not downwardly adjust the death counts for these cohorts.

For 1990, population counts by age, ethnicity, and state will be gathered from the Decennial Census. For 1991 – 1999, population counts by age, ethnicity, and state will be gathered from the U.S. Census Bureau Population Estimates.

[TABLE 2.1]

The three studies used to create the alternate estimates of emigration each list rates of emigration by select demographic characteristics. Table 2.1 displays the figures that are taken from these studies which are used in this analysis. The figures are displayed in percentages of the foreign-born who emigrate.

METHODS 1: RESIDUAL ESTIMATES OF EMIGRATION

The simplest and most popular way to estimate emigration would involve the use of residual methods (e.g. Mulder 2002). This method simply survives a base population and subtracts the count of the same population at a second point of time (in this case persons residing in the U.S. by the 1990 Census). Populations will be survived through the use of the survivorship ratios created from the life table. The basic steps used to create the survivorship ratios are below (see also Kintner 2004 for details on the life table). The life table used in this analysis will be based on single-year age intervals. In this section, “*a*” will represent age.

$$m_a = \text{deaths}_a / \text{base population}_a \quad [2.2a]$$

$$q_a = m_a / (1 + m_a) \quad [2.2b]$$

$$l_a = l_{a-1} (1 - q_{a-1}) \quad [2.2c]$$

$$d_a = l_a * q_a \quad [2.2d]$$

$$L_a = l_a + .5 * d_{a-1} \quad [2.2e]$$

$$T_a = T_{a+1} + L_a \quad [2.2f]$$

The survivorship ratios are used to produce predicted counts of persons that will survive “t” years into the future (which can also translate to “the year 2000” in many cases). These survivorship ratios can be interpreted as the ratio of persons out of a base population that will be expected to survive into the future. The survivorship ratios are calculated using the results from the life table, as described below.

$$s_a^t = L_{a+t} / L_a, \text{ or} \quad [2.2g]$$

$s_a^t = T_{a+t} / T_a$, as applied to the population that will be survived into (and within) the last age interval. [2.2h]

Finally, the survived population values are calculated using the following formula

$$\hat{p}_{a+t,t} = p_{a,0} \prod_{i=a}^{a+t} s_i \quad [2.3]$$

Where

$\hat{p}_{a+t,t}$ = The expected number of persons aged $a+t$ at time t , and

$p_{a,0}$ = The population of age a at time 0.

Having survived the base population, emigration will be estimated as follows.

$$Emig_{a,t} = \hat{p}_{a+t,t} - p_{a+t,t} \quad [2.4]$$

Where

$p_{a,t}$ is the population counted at time t (in this case Census “year 2000”) who are native-born or who have been reported to reside in the U.S. in 1990, and

$Emig$ = Emigration count

METHODS 2: SURVEY-BASED ESTIMATES OF EMIGRATION

Survived Hispanics

Emigration per year would be applied to Hispanics who have survived from the previous period and who remain in the country. Equation 2.3 will be used for each year to survive the base

population of foreign-born Hispanics which have emigration estimated from the previous year (described in the subsequent paragraphs) subtracted from the count, producing $H.FBCount_{a,t}$ (Hispanic foreign-born count for age a at time t). Sums of all the age groups for each year will produce the total survived Hispanics, or $H.FBCount_t$.

Total Hispanic Emigration Rates

This first equation will use the different rates of emigration for persons of varying places of origin (i.e. usually Mexico, Carribbean, and Central/South America) and weigh them according to their representation in the IPUMS database.

$$H.ERate_{total} = \sum_{c=1}^n (ERate_c \times \frac{Emig_c}{Emig_{total}}) \quad [2.5]$$

Where

$H.$ = Hispanic,

$ERate$ = Rate of emigration (proportion), and

c = Country or region (i.e. Mexico, Central/South America).

Where $Emig$ is not available, it is calculated as

$$Emig_{c,t} = ERate_c \times H.FBCount_{c,t} \quad [2.6]$$

Equation 2.6 is also used to estimate Hispanic emigration in subsequent phases after emigration rates for total Hispanics have been solved for. In these cases, c = Hispanic.

The rates of emigration calculated by Swabish (2009) only pertain to the documented foreign-born. As such, it would be prudent to adjust the rates upward by accounting for the size of the undocumented population and its concomitant rates of emigration. This task will require using emigration rates of Passel et al. (2006), who estimate separate rates of emigration for the undocumented population (2.2%) as well as for the total population (1.5%). By accounting for

these rates and assuming that the differences between the rates apply specifically to the Hispanic population, the total rate of emigration based on Swabish will be as follows.

$$H.ERate_{total} = \left(\frac{ERate_{total,P} \times H.FBCount_{total} - ERate_{u,P} \times H.FBCount_u}{H.FBCount_d} \right)^{-1} \times ERate_{total,P} \times H.ERate_{d,Sw} \quad [2.7]$$

Where

P = Passel et al. (2006) estimated figure,

u = undocumented status, and

d = documented status.

Sw = Swabish (2009) estimated figure, which only refers to documented immigrants.

The formula within the parentheses produce an estimate of the emigration rate for Hispanic documented immigrants using Passel's estimate. This is divided into the total emigration rate of Passel to estimate a multiplicative weight to multiply by the documented emigration rate of Swabish (2009) in order to estimate total emigration from Swabish's figures. Conceptually, the formula reads "total emigration divided by domestic emigration, times domestic emigration, equals total emigration," reconciling estimates from two sources. Counts of Hispanics by legal status works under the assumption that there are 2,176,000 undocumented immigrants captured in the 1990 Census (based on the average of Woodrow-Lafield's 1995 estimates), and that Hispanics represent 79% of undocumented immigrants (based on Passel, Van Hook, and Bean 2004).

Emigration for the Hispanic Foreign-Born Residing in the U.S. for Ten or More Years

Rates of emigration vary greatly by the length of stay in the U.S. For this reason, emigration rates for those residing in the U.S. for ten or more years will be estimated separately from the

total emigration rate. Emigration for those who have resided in the country for ten years or more will be used as the emigration rate for 1999, as the stock of persons remaining in the U.S. in 1990 would have collectively been residing in the U.S. for ten or more years.

$$H.ERate_{10+} = (H.FBCount_{total} \times H.ERate_{total}) \times \left(\frac{\sum_{x=1}^n (H.FBCount_x \times ERate_x)}{\sum_{y=1}^n (H.FBCount_y \times ERate_y)} \right) \times (H.FBCount_{10+})^{-1} \quad [2.8]$$

Where

10+ = total of those residing in the U.S. for ten years or more,

x = all available values corresponding to 10+ years of residence in the U.S., and

y = all available values corresponding to years of residence in the U.S, 10+ years and otherwise.

To break down equation 2.7, the formula in the first set of parentheses estimates the total emigration for Hispanics. The formula in the second set of parentheses estimates the proportion of Hispanics who emigrate, for those who have resided in the U.S. for ten years or more.⁶ The formula in the last set of parentheses is simply the inverse of the count of all foreign-born Hispanics who have been in the country for ten or more years. Putting these pieces together, this formula assumes that Hispanics emigrate by length of stay in the U.S. at the same rates as does the entire U.S. population (which is a reasonable assumption considering that Hispanics comprise the majority of the foreign born). The rate of emigration for Hispanics residing in the U.S. for ten years or more is simply the calculated number of Hispanic emigrants who have resided in the country for ten years or more – calculated as the total number of Hispanic emigrants times the proportion of the total Hispanic emigrants that have resided in the

⁶ Van Hook et al. (2006) provide their own counts of emigrants by years of residence in the U.S. These counts were used in lieu of the estimated counts of emigrants by years of residence in the U.S.

U.S. for ten years or more – divided by all the Hispanic foreign-born that have resided in the U.S. for ten years or more.

Yearly Rate of Emigration

Having calculated both total emigration rates for Hispanics, as well as rates of emigration for Hispanics who have resided in the U.S. for ten years or more, the next step will be to calculate the rates of emigration per year, as our cohort progressively moves from a “total” rate of emigration to one for which a “ten or more years” rate of emigration is more applicable. This research will work under the assumption that the rates of emigration decline linearly from a “total” rate to a “ten or more years” rate. As such, the yearly rates will be calculated using the following formula. The initial baseline t is 0 (which corresponds to 1990).

$$H.ERate_t = H.ERate_{total} - \left(\frac{H.ERate_{total} - H.ERate_{10+}}{10} \right) \times t \quad [2.9]$$

Distributing Emigrants into Age Groups

Total migrants per year would need to be distributed in two steps, as all three sets of estimates have larger age groupings than the desired five-year age groups. The first step distributes total migrants by the age groups included in the previous authors’ studies.⁷ The formula in the parentheses calculates the proportion of emigrants that come from a particular age group, which is then multiplied by the total emigrants to distribute total emigrants to the age group in question.

$$H.Emig_{agegrp,t} = \left(\frac{H.FBCount_{agegrp} \times ERate_{agegrp}}{\sum_{agegrp=1}^6 (H.FBCount_{agegrp} \times ERate_{agegrp})} \right) \times H.Emig_{total,t} \quad [2.10]$$

⁷ Swabish’s age groupings differed from the other authors in two age groupings, 16-24 (as opposed to 15-24) and 45-62 (as opposed to 45-64). These groups were treated as if they aligned with the other age groupings grouped within age categories that are multiples of five. Rates of emigration for those under 16 were assumed to be the same as for those from 35-44 (their presumed parents), and those over 65 were assumed to have the same emigration rates as those aged 45-62. These assumptions are supported by the fact that the other two studies find these corresponding emigration rates to be similar.

Where

$agegrp$ = the age groupings used by the three studies, and

$H.Emig_{total,t}$ is derived from a modified version of Equation 2.5 that uses Hispanic-specific inputs for time t .

Secondly, the counts distributed into the larger age groups ($agegrp$) would be distributed into single-years of age by using simple proportions.

$$H.Emig_{a,t} = \frac{H.FBCount_{a,t}}{H.FBCount_{agegrp,t}} \times H.Emig_{agegrp,t} \quad [2.11]$$

Where $agegrp$ is specific to the age group to which a corresponds.

Total Emigration for Age Groups

The total emigration for each age cohort will be produced by summing each age-specific emigration count rate for each time point.

$$H.Emig_{a,0-9} = \sum_{t=0}^9 H.Emig_{a+t,t} \quad [2.12]$$

RESULTS

Estimates of Emigration

[TABLE 2.2]

The most peculiar of the migration estimates below is the residual estimate of migration.

Incredibly, the total emigration calculated from the residual method is -482,871, suggesting significant net “negative emigration.” Assuming the perfect accuracy of the data used to make these calculations, namely the 1990 and 2000 Censuses, the only conceptual way for this to be the case would be the exit of these hundreds of thousands of Hispanics before the 1990 Census, who then return to be counted in the 2000 Census. A more likely explanation of this phenomenon is the high rate of undercount in the 1990 Census, which has been estimated as

4.99% for Hispanics (Mulry 2006). As Hispanics in the 1990 Census were likely undercounted at much higher rates than Hispanics in the 2000 Census, residual estimates would deflate estimates of emigration – in this case to such an extent that residual methods would produce negative figures. Despite the flaws present in the data sources used, the residual method of emigration estimation was still able to capture emigration for persons aged 30 and over.

Of the three alternate estimates of emigration, the Van Hook rates of emigration produce, by far, the highest estimate of emigration. Using these rates of emigration, total ten-year emigration is estimated at nearly two-and-a-half million, which is over twice as high as the next highest estimate in Passel. The estimate based on Schwabish's rates of emigration is around 20% lower than Passel's estimate, and thus can serve as a lower-bound estimate of the sample-based estimates of emigration. However, Schwabish's estimate still produces a total count of emigration that is much higher than that which is produced by using the residual method, which can serve as the lowest estimate of emigration.

[FIGURE 2.1]

Figure 1 graphically displays the emigration estimates by age. Before the age of 45, the residual estimate clearly stands out as the black-sheep of the emigration estimates with counts that are consistently around 300,000 lower than the Van Hook estimates for Hispanics under 40 years of age in 1990. From 45 – 65, however, the estimates based on Passel stands out as the underestimate. For the first half of age groups, the Van Hook estimate leads the pack in estimated counts of emigration and stand out as a positive extreme, particularly for the in the 10-20 and 30-40 age intervals.

Despite these differences in the pattern of migration, there seems to be agreement among the alternate estimates of emigration that the 20-40 year cohorts tend to emigrate at

the highest absolute levels. There additionally is agreement across all estimates that the highest age group emigrates at levels that are relatively high. The high absolute numbers of emigrants for persons of “working age” (20-44) potentially raise issues regarding the “circular migration” of Hispanics (Massey et al. 2002; Rendall and Torr 2008). That is, high counts of emigration at persons of working age may be a function of Hispanics who routinely emigrate from, and immigrate into the U.S., which may potentially raise the “absolute counts” of emigration. However, two of the estimates account for circular migration. Van Hook et al. (2006) use estimates of circular migration to adjust emigration estimates, and Schwabish (2009) relies on data that would not allow for multiple counts of exits for the same person. Although the Van Hook estimates account for circular migration, these emigration estimates still far outpace the other estimates, suggesting that circular migration will not play a factor in inflating emigration estimates at rates that exceed these estimates displayed.

Explaining Error: Age, Year of Immigration, and Political Context

Since all estimates of emigration intend to measure the same concept, the alternate estimates of emigration can be used to evaluate the residual estimate and thereby make comments on its data sources. The relationships between the emigration estimates and sources of error are assumed to be as follows:

$$H.Emig_{a,alternate} + Error_{a,alternate} = H.Emig_{a,residual} + Undercount_{a,1990} - Undercount_{a,2000} - Undercount_{mortality} \quad [2.13]$$

These variables can be re-arranged to isolate error.

$$H.Emig_{a,alternate} - H.Emig_{a,residual} = Undercount_{a,1990} - Undercount_{a,2000} - Error_{a,alternate} - Undercount_{mortality} \quad [2.14]$$

It will be assumed that the average of the alternate estimates of emigration is a “truer” approximation of $H.Emig_a$, and thus the average will be used for said component. It can also be assumed that $Undercount_{mortality}$ is 0, since this was already accounted for in the study using Arias et al. (2008). The quality of the 2000 Census is far superior to the quality of the 1990 Census for Hispanics (Mulry 2006), leading to the overshadowing of $Undercount_{2000}$ by $Undercount_{1990}$. Thus, the majority of the “collective error” will be the differential between $Undercount_{1990}$ and $Error_{alternate}$. The collective error is displayed in absolute terms and as a percentage of the total survived Hispanic population on Table 2.3.

[TABLE 2.3]

Table 2.3 displays a linear relationship between age and collective error. Error for the youngest age interval is estimated at 11.7%, while the error for the oldest age interval has a very high negative value of 13%. Literature that comments on the undercount of children (O’Hare 2009; West and Robinson 1999) would suggest that the majority of the error found for the younger intervals is due to undercount for the 1990 Census. On the other hand, literature on the salmon bias, or the emigration of older Hispanics (see Chapter 3, Abraido-Lanza et al. 1999; Elo and Turra 2008; Palloni and Arias 2004), would suggest that the alternate estimates of emigration underestimated the emigration of older Hispanics. The fact that the residual estimates of emigration, with all of its problems of undercount, still estimated higher counts of emigration for older Hispanics (over age 65) supports the idea that alternate estimates of emigration may have been understated at higher ages. The degree to which undercount and error in emigration estimation apply to these age groups (let alone the other age groups) is undetermined and will be reserved for future research.

[TABLE 2.4]

A closer look at the data demonstrates that age may not necessarily be the main factor that predicts collective error. One would be able to test the relationships that various factors play on collective error by creating a dataset that includes single years of age as the rows (“cases”) and corresponding collective error percentages in one column. Assigning the mean year of immigration for each age (using Census 2000 data) and regressing age and year of immigration on error, one finds that the year of immigration has a stronger effect on predicting collective error than does age, and even reverses age’s relationship with error when placed in the same model (regressions not shown). Table 2.4 displays the correlation matrix of possible predictors of collective error. While age is very highly correlated with year of entry, one finds that average year of entry per age has a slightly stronger correlation on error than does age. Table 2.4 includes three variables that represent the proportions of persons by age that arrive in a particular time-period with respect to IRCA legislation. Those arriving in time for IRCA amnesty (pre-1982) tend to have lower scores on collective error. On the other hand, those who were ineligible for IRCA amnesty are more likely to have higher scores on collective error.

[TABLE 2.5]

If one assumes that undercount of the 1990 Census comprises the bulk of error, the relationship between year of entry and error can offer clues regarding the dynamics that contribute to undercount. It is clear that year of entry has a strong linear effect on count. However, additional analyses also suggest that the political social context (i.e. IRCA legislation) also affect the quality of the Census. Table 2.5 displays simple counts of Hispanics by year of entry for 1990 and 2000, providing a “quick-and-dirty” glance at potential emigration minus 1990 undercount rates. Although emigration rates are higher for more recent entrants (and, hence, the numbers of more recent entrants should theoretically be higher for 1990 and 2000),

Table 2.5 shows that the most recent entrants are counted at the lowest rates in 1990 vis-à-vis the 2000 Census. Additionally, there are clear cleavages across the three eras, as entrants in the post-IRCA period tend to have much lower counts in 1990 than in 2000. Entrants who are ineligible for amnesty but have entered prior to the passage of IRCA are similarly counted at lower rates in 1990 than in 2000. Those who would qualify for amnesty, however, are more highly counted in 1990 than in 2000.

Table 2.5 also demonstrates that there is no clear linear pattern between year of entry and count discrepancies within these year ranges provided, which would otherwise explain count discrepancies as a linear “year of entry” effect. Entrants in 1982-1984 are missed in the 1990 Census at higher rates than entrants in 1985-1986, and entrants during 1980-1981 are counted at far higher rates than entrants during 1975-1979. In other words, count discrepancies are not solely a function of years of residence, which suggests that the phases of IRCA implementation have real effects on count in the 1990 Census.

CONCLUSION

This chapter has served to produce estimates of the number of Hispanic emigrants between the 1990 and 2000 Censuses. Residual methods of estimating emigration are the most conceptually-simple, yet it is evident that they are insufficient in producing accurate estimates of emigration. Fortunately, alternate methods of estimating emigration rates based on samples have been produced in the recent years. The resulting estimates produce higher counts of emigration, ranging from just under 1,000,000 to just under 2,500,000 emigrants total over 1990-2000.

The alternate estimates of emigration generally follow a similar pattern. This pattern sees high absolute counts of emigration for Hispanics aged 20-44, which seem to stabilize to a relatively consistent count in later ages. Emigration for cohorts younger than 20 rises relatively

steadily and peak around age 30. Following age 45, the residual method produces counts of emigration that generally align with (and often surpass) the alternate estimates.

Although the alternate estimates of emigration account for the undercount and non-sampling errors that are attendant upon using Decennial Census data, the alternate estimates of emigration are not free from error. The Passel and Van Hook algorithms are founded on the assumption that the rates of non-response, internal migration, death, and “other reasons” are accurate. The accuracy of mortality, for example, rest on the consistency of CPS respondents and the National Health Re-interview Survey. Non-response for “other reasons” is assumed to be the same for the native-born as for the foreign-born. And, as with all survey-based estimates, sampling and non-sampling error is assumed to be minimal. For Schwabish estimates, not filing a tax return for two years is assumed to be an indicator of death or emigration, and the undocumented population is notoriously absent in the sample. All of these aforementioned assumptions can be challenged. In short, many of the issues raised by Judson (2006) and Judson and Swanson (2011) continue to apply to these estimates.

Delving deeper into the discrepancies between the residual and alternate estimates of emigration, the data allow for some explorations regarding explanations behind the undercount found on the 1990 Census and error in the alternate emigration estimates. It is shown that age tends to have a negative relationship on collective error. However, year of entry is shown to have a stronger negative influence on collective error than does age. Additionally, differences in the counts of Hispanics are evident across categories of post-IRCA, pre-IRCA without amnesty eligibility, and pre-IRCA with amnesty eligibility where those who are eligible for amnesty are much more likely to be counted in 1990 versus the other categories. The relationship between

the differences in the count and the various waves of IRCA legislation also show sharp, non-linear cleavages.

Larger implications of the last finding point to the relevance of political climate on the trust of the government, let alone of the Census. As groups of residents feel negatively targeted, the suspicion level of those targeted are inevitably raised regarding their perceptions of the intentions of the government. The quality of the 1990 Census therefore suffered considerably due to the political climate in which the 1990 Census was situated. This underscores the necessity of the Census and its surveys to remain neutral and non-threatening in the information that it gathers.

Data quality is subject to numerous sources of error, and statisticians are readily aware of many of the limits inherent in the data that are available. Demographers in particular who deal with Census data confront a source of error which deals with the public receptiveness of the collectors of data. Suspicion of governmental institutions affects the data collected by governmental institutions, as evidenced in this case dealing with Hispanics. The degree to which this suspicion remains and affects subsequent Censuses is uncertain, although it would be safe to assume that those who feel discriminated against or targeted – including Hispanics, blacks, South Asians, and Middle Easterners (although some of the latter two categories of South Asians would be prompted to identify as “white”) – continue to be undercounted.

REFERENCES

- Abraido-Lanza, Ana, Bruce Dohrenwend, Daisy Ng-Mak, and Blake Turner. 1999. "The Latino Mortality Paradox: A Test of the 'Salmon Bias' and Healthy Migrant Hypothesis." *American Journal of Public Health* 89(10):1543-1548.
- Ahmed, Bashir, and J. Gregory Robinson. 1994. "Estimates of the Foreign-born Population: 1980-1990." *Population Division Working Paper Series* No. 9. Washington, D.C.: U.S. Census Bureau.
- Arias, Elizabeth., William Schauman, Karl Eschbach, Paul Sorlie, and Eric Backlund. 2008. "The Validity of Race and Hispanic Origin Reporting on Death Certificates in the United States." *Vital Health Stat* 2:1-24.
- Borjas, George J. and Bernt Bratsberg. 1996. "Who Leaves? The Outmigration of the Foreign-Born." *The Review of Economics and Statistics* 78(1):165-176.
- Elo, Cassio, and Irma Turra. 2008. "The Impact of Salmon Bias on the Hispanic Mortality Advantage: New Evidence from Social Security Data." *Population Research Policy Review* 27:515-530.
- Fernandez, Edward W. 1995. "Census Bureau Estimation of the Annual Emigration of U.S. Born Persons by Using Foreign Censuses and Selected Administrative Data: Circa 1980." *Population Division Working Paper Series* No. 10. Washington, DC: U.S. Census Bureau.
- Gibbs, Jim, Gregory Harper, Mark Rubin, and Hyon Shin. 2003. "Evaluating Components of International Migration: Native-Born Emigrants." *Population Division Working Paper Series* No. 63. Washington, D.C.: U.S. Census Bureau.
- Judson, Dean. 2006. "Memorandum for E. Lamas (21 September): Subject: Preliminary Results and Evaluation of Experimental Estimates of the Foreign-Born Population by Legal Status." Retrieved March, 2008 (<http://wikileaks.org/wiki/USCensusBureauEstimatesofUnauthorizedPersons>).
- Judson, Dean and David Swanson. 2011. "Estimating Characteristics of the Foreign-Born by Legal Status: An Evaluation of Data and Methods." *Springer Briefs in Population Studies, Volume 2*. Dordrecht, The Netherlands: Springer, B.V. Press.
- Kintner, Hallie. 2004. "The Life Table." Pp. 301-340 in *The Methods and Materials of Demography*, edited by Jacob Siegel and David Swanson. Boston, MA: Elsevier Academic Press.
- Martin, Elizabeth. 2002. "Questionnaire Effects on Reporting of Race and Hispanic Origin: Results of a Replication of the 1990 Mail Short Form in Census 2000." *Census 2000 Alternative Questionnaire Experiment*. Washington, DC: U.S. Census Bureau.

Massey, Douglas. Jorge Durand, and Nolan Malone. 2002. *Beyond Smoke and Mirrors: Mexican Immigration in an Era of Economic Integration*. New York, NY: Russell Sage Foundation.

Mulder, Tammany J., Betsy Guzmán, and Angela Brittingham. 2002. "Evaluating Components of International Migration: Foreign-born Emigrants." *Population Division Working Paper Series* No. 62. Washington, D.C.: U.S. Census Bureau.

Mulry, Mary. 2006. *Summary of Accuracy and Coverage Evaluation for Census 2000*. Washington, D.C.: U.S. Census Bureau.

O'Hare, William. 2009. *Why are Young Children Missed So Often in the Census?* Kids Count Working Paper. Annie E. Casey Foundation, Baltimore, MD.

Palloni, Alberto and Elizabeth Arias. 2004. "Paradox Lost: Explaining the Hispanic Adult Mortality Advantage." *Demography* 41:385-415.

Passel, Jeffrey S., Jennifer Van Hook, and Frank D. Bean. 2004. "Estimate of the Legal and Unauthorized Foreign-Born Population for the United States and Selected States, Based on census 2000." Retrieved April 1, 2011 (http://www.copafs.org/UserFiles/file/seminars/methodology_and_data_quality/Estimates%20of%20the%20Legal%20and%20Unauthorized%20Foreign-Born%20Population%20for%20the%20United%20States%20and%20Selected%20States,%20Based%20on%20Census%202000.pdf).

Passel, Jeffrey, Jennifer Van Hook, Frank Bean, and Weiwei Zhang. 2006. "Emigration: Measurement Issues and New Estimates." Paper presented at Congress and Tomorrow's Foreign-Born Population: Evaluating Immigration Projections, September 26-27, Washington, DC.

Rendall, Michael and Bema Torr. 2008. "Emigration and Schooling among Second-Generation Mexican-American Children." *International Migration Review* 42:729-738.

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. *Integrated Public Use Microdata Series: Version 5.0*. Minneapolis, MN: University of Minnesota, 2010.

Schwabish, Jonathan. 2009. "Identifying Rates of Emigration in the United States using Administrative Earnings Records." Working Paper, Congressional Budget Office, Washington, DC.

Van Hook, Jennifer, Weiwei Zhang, Frank Bean, and Jeffrey Passel. 2006. "Foreign-Born Emigration: A New Approach and Estimates Based on Matched CPS Files." *Demography* 43:361-382.

Warren, Robert and Jeffrey S. Passel. 1987. "A Count of the Uncountable: Estimates of Undocumented Aliens Counted in the 1980 United States Census." *Demography* 24(3):375-393.

Warren, Robert, and Jennifer Marks Peck. 1980. "Foreign-Born Emigration from the United States: 1960 to 1970." *Demography* 17(1): 71-84.

West, Kirsten and J. Gregory Robinson. 1999. "What Do We Know About The Undercount of Children?" *U.S. Bureau of the Census Population Division Working Paper* No. 39, Washington, D.C: U.S. Census Bureau.

Woodrow-Lafield, Karen A. 1995. "An Analysis of Net Immigration in Census Coverage Evaluation." *Population Research and Policy Review* 14:173-204.

TABLES

Table 2.1: Emigration Rates by Study

	Van Hook et al. (2006)	Passel et al. (2006)	Swabish (2009)*
Total	2.9	1.5	1.31
"Hispanic" areas			
Mexico	4.3	1.8	1.33
Caribbean	1.8	1.2	1.27
Central and South America	2.4	0.9	
Central America			0.95
South America			1.21
Years in the U.S.			
0-4	5		2.72
5-9	3.6		1.78
10+	2	1.8	
10-15			1.27
16-20			0.98
21+			0.22
Age			
0-14	4.9	2.4	
15-24	3.6	1.2	0.69
25-34	2.7	2.7	1.25
35-44	4.6	2.6	1.49
45-64	1.3	0.1	2.31
65+	1.2	1.1	

*Swabish's estimates only apply to the documented population.

Table 2.2: Hispanic emigration for those residing in the U.S.
in 1990: 1990-2000

Age in 1990	Residual	VanHook	Passel	Schwabish
0	-231,091	86,389	36,561	19,926
5	-209,231	133,636	53,189	29,658
10	-145,406	160,280	52,972	29,174
15	-137,443	214,083	80,257	35,394
20	-80,606	302,156	182,315	74,392
25	-5,345	352,794	244,396	104,454
30	41,515	404,265	219,959	103,674
35	51,582	333,778	147,425	96,354
40	20,153	170,255	51,090	93,307
45	37,930	74,813	4,563	83,600
50	17,874	58,365	3,558	65,292
55	21,035	44,573	7,772	50,568
60	24,451	33,623	16,270	39,679
65	33,623	24,230	16,855	29,490
70	31,519	15,549	10,791	19,028
75+	46,569	22,532	15,487	28,221
Total	-482,871	2,431,321	1,143,460	902,211

Table 2.3: Collective Error¹ by Age

Age in 1990	Residual	Average Alternate Emigration	Collective Error	Collective Error Percent ²
0	-231,091	47,625	278,716	11.7
5	-209,231	72,161	281,392	12.9
10	-145,406	80,809	226,215	11.4
15	-137,443	109,911	247,354	12.2
20	-80,606	186,288	266,894	11.7
25	-5,345	233,881	239,226	10.4
30	41,515	242,633	201,118	10.0
35	51,582	192,519	140,937	8.7
40	20,153	104,884	84,731	6.8
45	37,930	54,325	16,395	1.8
50	17,874	42,405	24,531	3.5
55	21,035	34,304	13,269	2.3
60	24,451	29,857	5,406	1.2
65	33,623	23,525	-10,098	-3.0
70	31,519	15,123	-16,396	-8.3
75+	46,569	22,080	-24,489	-13.0
Total	-482,871	1,492,331	1,975,202	9.2

¹Collective Error represents 1990 Undercount minus 2000 Undercount minus the error of the average alternate estimate counts of emigration

²The denominator for Collective Error Percent is the survived 1990 Census count of Hispanics.

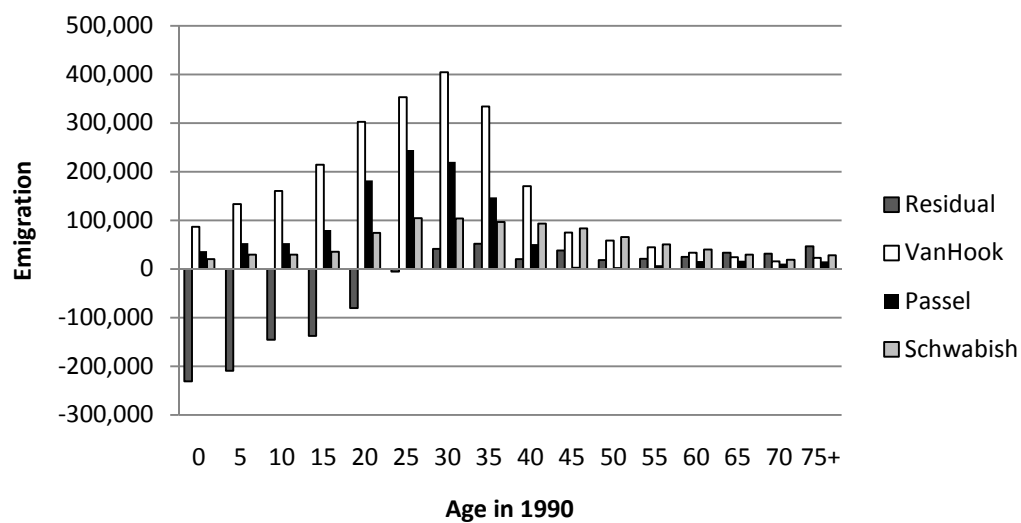
Table 2.4: Correlates of Collective Error, Correlation Matrix

	Collective Error	Age	Year	Amnesty	Pre-IRCA, No Amnesty	Post- IRCA
Age	-0.7966	1				
Year of Immigration	-0.8078	0.9952	1			
IRCA Amnesty Eligibility	-0.6455	0.9256	-0.914	1		
Entry Before IRCA, No Amnesty Eligibility	0.5602	-0.6697	0.6915	-0.662	1	
Entry Post-IRCA	0.5469	-0.8447	0.8203	-0.9405	0.3679	1

Table 2.5: Count differences of Foreign-Born Hispanics across Censuses by year of entry

Year of Entry	Year of Entry	1990 Count	2000 Count	Net Percent Difference
Post-IRCA	1987-1990	1,621,640	1,807,877	-11.48
Pre-IRCA, no amnesty eligibility	1985-1986	955,960	992,860	-3.86
	1982-1984	865,918	909,820	-5.07
Pre-IRCA, amnesty eligibility	1980-1981	1,070,591	968,067	9.58
	1975-1979	1,311,803	1,271,206	3.09

Figure 2.1: Hispanic emigration for those residing in the U.S. in 1990: 1990-2000



CHAPTER 3

CONFRONTING THE CHALLENGES OF THE HISPANIC MORTALITY PARADOX: EMIGRATION AND UNDERREPORTING ON DEATH CERTIFICATES

The high health outcomes of Hispanics have been referred to as a “paradox,” as their generally lower socioeconomic status would presumably be tied to negative health outcomes (Markides and Coreil 1986). This paradox is not without its critics who largely argue that the paradox is a result of faulty data. Pieces of counter-explanations point to the roles that underreporting of death certificates and emigration of Hispanics play in lowering mortality rates. This research aims to account for all of these counter-explanations of the paradox which would serve to increase the mortality rates of Hispanics and decrease the mortality rates for Non Hispanic whites. This research will compare the resultant low estimates of Hispanic life expectancies with the high estimates of Non Hispanic white life expectancies, thereby testing for the continued existence of the Hispanic Mortality Paradox after accounting for data artifacts. Additionally, population undercount will also be considered as a factor that alters the health outcomes of both populations. In sum, this research proposes to use data that are as clean as possible in assessing the health outcomes of Hispanics and Non Hispanic whites. This research will demonstrate that adjusting all sources of data to produce the most plausible estimates of the population, emigration, and death (i.e. adjusting for death misclassification, emigration, and population undercount), the Hispanic Mortality Paradox continues to hold. However, the mortality advantage that Hispanics have over Non Hispanic whites largely depends on the assumptions that one makes regarding the accuracy of the data, as the life expectancies of Non

Hispanic whites are calculated to be higher than those of Hispanics under half of these estimates if population undercount is not considered.

THE HISPANIC MORTALITY PARADOX/ADVANTAGE

In 1986, Markides and Coreil noted that Hispanics, while tending to have a low socioeconomic status similar to those of blacks, paradoxically tended to have health outcomes similar to Non-Hispanic whites (NH whites). This finding is particularly paradoxical when considering that low socioeconomic status is associated with negative health outcomes (see Feinstein 1993 for a literature review). For this reason, the high health outcomes of Hispanics have been termed a “Hispanic Mortality Paradox” or “Advantage” (hereafter referred to interchangeably with “Paradox”).

Since the 1986 article, much literature has been written on the mortality advantage of Hispanics. In multivariate models controlling for socioeconomic characteristics, Hispanics have been demonstrated to have favorable mortality rates (see Liao et. al 1998; Sorlie et al. 1993). The voluminous work of Hummer and his associates adds much to this body of research through the use of linked National Health Interview Survey (NHIS)-Multiple Cause of Death data, demonstrating that foreign-born Mexicans and Other Hispanics have lower mortality for ages 45-64 and 65+ (Hummer et al. 1999) and continuing to find similar patterns using updated data sets (e.g. Hummer et al. 2000; Hummer, Benjamins, and Rogers 2004). Despite the positive health outcomes for Hispanics, self-reporting of health problems tend to be higher for older Mexican Americans than for Non-Hispanic whites (Hummer et al. 2004; Markides et al. 1997).

Palloni and Arias (2004) group the explanations of the Paradox under three general schools of thought. One explanation is based on cultural effects, which argues that Hispanic migrants tend to in-migrate (hereafter “immigrate”) into the U.S. with different lifestyles, family

structures, and social networks that lend toward positive health outcomes. A second posits that the Paradox can be explained away through data artifacts, most notably in the misclassification of ethnicities by data collectors (i.e. death certificates). A third explanation views migration as a factor in explaining health disparities, as migration may select a healthier stock of persons than those who do not migrate. Out-migration (hereafter “emigration”) for less healthy persons who return to their original country to receive care or to die (known as the “salmon bias”) may also decrease the number of deaths reported for migrants relative to the numbers of migrants that are recorded on Censuses and other data counts. These explanations will be further explained below.

Cultural Effects

Hispanic migrants that enter the country may have healthier lifestyles stemming from their culture of origin. Original cultural ideas on diet, smoking, and preventative care may differ from the McDonalds and health-oblivious-until-problems-arise culture of the United States (particularly in areas of lower socioeconomic status). Additionally, having a stronger family-oriented culture decreases the probability of living alone and increases the probability of drawing on social networks and social support, meeting socio-psychological needs that buffer against illness (Abraido-Lanza et. al. 1999; Arias 1998). Palloni and Arias (2004) argue that there is no statistical basis for cultural effects, as married migrants and indices of isolation of migrant groups are not statistically related to better health outcomes (although a discussion of the validity of the independent variables used is also warranted).

Data Artifacts

A counter-explanation for the Paradox is that misreporting on administrative data is at fault in inflating the positive health outcomes of Hispanics. In general, authors that illuminate problems

in data demonstrate that mortality rates can be artificially lowered for Hispanics by deflating the numerator (i.e. death counts) and/or inflating the denominator (i.e. population counts), thereby artificially producing favorable health outcomes for Hispanics. For death data on Hispanics, ethnicity is recorded by a funeral director, ideally using information from an informant but otherwise using observation. In contrast, data on ethnicities from Census surveys and birth certificates rely on self-identification. Rosenberg et al. (1999) conducted an early study of this phenomenon, using the National Longitudinal Mortality Study (NLMS) data which links the Census Bureau's Current Population Survey files (CPS) with death certificates for select years (in this case 1979-1985). This study effectively compares the rates of ethnic identification by self-reporting (CPS) and funeral director reporting (death certificates).⁸ These authors find that record-by-record agreement for Hispanic origin was 89.7% and that the ratio of Hispanic CPS deaths to death certificate deaths was 1.07, meaning that more CPS-identified Hispanics were identified as dead than the vital statistics would suggest. Arias et al. (2008) use updated NLMS data and find that the classification ratio for the same period under the Rosenberg et al. (1999) study was 1.04 (indicating better results with the use of a larger sample size), although the classification ratio for the 1990s increased to 1.05. These authors find that matching ethnicities across data sets was more consistent for foreign-born Hispanics, as well as in areas of high concentration of Hispanics.

The effect that the misclassification of ethnicities has on the Paradox is up for debate.

Arias et al. (2010) use an updated NLMS dataset and create "classification ratios" that measure net misclassification of ethnicities on death certificates, calculated as CPS Hispanic deaths by

⁸ The CPS' method of data collection is not always the same as that of the Decennial Census. The CPS conducts telephone interviews, whereas the Decennial Census only utilizes interviews when mail responses are not received.

corresponding counts from death certificates. After adjusting mortality rates by classification ratios by age, these authors still find that Hispanics enjoy a mortality advantage. Elo et al. (2004) additionally adjust death rates using Medicare-NUMIDENT (Social Security) records which identify Hispanics by place of birth, race, and a robust surname logarithm, continuing to find that Hispanics (particularly older Hispanics) have a mortality advantage over whites. What these authors do not do, however, is adjust downward the deaths to non-Hispanic whites and recalculate death rates to see if Hispanics still have higher health outcomes than do whites.

Smith and Brashaw (2006) suggest that one factor that contributed to the apparent mortality advantage was the large increase in persons identified as Hispanic on the Census due to the differing means of collecting data on ethnicity (i.e. self-reporting on the Census and third-person observation on death certificates). Thus, while the denominators of death rates (population count) increased, there was a corresponding lack of an increase in the numerators (death count) which led to a decrease in mortality rates as a product of data artifacts. Adjusting for the different coding procedures for 1990 and 2000 (and working under numerous assumptions) and transferring deaths to Hispanics and from NH whites, the authors find that the mortality advantage disappears.

While data artifacts have been argued to lower mortality rates, data artifacts can additionally raise mortality rates for Hispanics. When relying on the Census for population counts, the undercount of the Hispanic population would subsequently raise their respective mortality rates (see Chapter 2 and Mulry 2006).

In sum, it is clear that the data sources available are all subject to questionable issues of comparability. In this case, the difference between data based on self-identification (i.e. on the Census and CPS) may not be completely compatible with data collected by an observer (i.e. on

death certificates). Aside from these issues, changing questions that identify Hispanics as well as the changing concept of the very idea of “Hispanic” may shake the foundations of data on ethnicity (as in Chapter 1). Fortunately, Arias et al. (2008) alleviate some of these issues by producing rates of misclassification across CPS self-identifiers and death certificate third-person observations.

Migration Effects

One factor that may influence death rates is the hypothetically healthier genetic stock of migrants in comparison with non-migrants. That is, having the fortitude to move to a foreign country may be selective on people of good physical and mental condition. Studies show that foreign-born persons exhibit lower death rates (Hummer et al. 1999) and generally have better health (Stephen et al. 1994) than native-born persons. This finding has also been demonstrated in Australia (Donovan et al. 1992) and Canada (Chen, Ng, and Wilkins 1996). Evidence suggests that even internally-migrant populations have more positive health outcomes than their stationary counterparts (Wingate and Alexander 2005). Perhaps the most irrefutable finding comes from Hummer et al. (2007) who use data that allows for the fewest alternate explanations. These authors showed that mortality for infants born in the United States to Mexican immigrant women have lower first-hour, first-day, and first-week mortality rates than do NH white women (Hummer et al. 2007).

A second type of migration effect that may alter death rates is referred to as the “salmon bias,” in which unhealthy or elderly Hispanics return to countries of origin to die, which artificially lowers death rates as their deaths are not included in U.S. vital statistics. Palloni and Arias (2004) have found indirect evidence of the salmon bias, as Mexicans have a higher mortality advantage than other groups that do not have the same capacity of returning to their

countries of origin (e.g. Cubans). Palloni and Arias (2004) are unable, however, to account for the higher mortality advantage of Other Hispanics⁹ (i.e. Hispanics who are not Mexican, Cuban, or Puerto Rican) who may not have as high of a capacity to return to countries of origin to die. Abraido-Lanza et al. (1999) have confirmed this finding, but frame this finding as evidence against the salmon bias, since Other Hispanics share the mortality advantage without being as able to return to countries of origin. Elo and Turra (2008) use the Master Beneficiary Record and NUMIDENT data files to provide direct evidence of the salmon bias, as out-migrants (not exclusive to Hispanics) are shown to have higher mortality than those who remain in the United States.

HYPOTHESES

This chapter will address two of the explanations of the Paradox as mentioned above. In sum, this research aims to add to the current body of research by testing whether the Hispanic Mortality Paradox will continue to hold after accounting for both data artifacts and the salmon bias. In spirit with Arias et al. (2010), this research will upwardly adjust the numbers of deaths occurring to Hispanics. In spirit with Smith and Bradshaw (2006), this research will also downwardly adjust the numbers of deaths occurring to NH whites. This research will also take one step further by adjusting population counts by removing persons (emigrants) who would not be exposed to the chance of death in calculating mortality statistics. These aforementioned steps will test whether the Paradox exists under the lowest estimate of life expectancies for Hispanics and highest for Non-Hispanics. However, to treat all evidence of data-artifacts fairly, this research will also consider the role that undercount plays in deflating life expectancies.

⁹ In the Decennial Census, the Hispanic identifier item contains check boxes for Mexicans, Cubans, and Puerto Ricans. Persons who identify as Hispanic but do not fall under the three categories can check an "Other Hispanic" box and are so termed.

First, this research will create life expectancies under two scenarios: raw death data as they are, and death data adjusted upwards for Hispanics and downwards for NH whites. While the existing literature largely relies on age-specific mortality rates in analyzing the Paradox, this research will analyze life expectancies by age group. It will be hypothesized that *(H1) Hispanics will have higher life expectancies than non-Hispanic whites*. Second, this research will re-visit the estimates of emigration produced in Chapter 2, hypothesizing that *(H2) Hispanics show evidence of emigration at later ages*, as one may suspect from literature on the “salmon bias.” Third, this research will revisit life expectancies after adjusting death counts by death misclassification as well as adjusting population counts by emigration for elderly Hispanics. After the data have been adjusted, it will be hypothesized that *(H3) the Hispanic Mortality Paradox will still hold after data are adjusted for both misclassification error and the salmon bias*.

METHODS

The Life Table will be used to create life expectancies by age groups for the Hispanic and NH white populations. The Life Table is traditionally organized by rows that designate age-groups and columns that display figures (see Kintner 2004 for more details). The columns of the Life Table begin with the counts of the population by age group, followed by the deaths that occur within each age group for a representative year. A basic death rate, m_x , is then created by dividing the amount of deaths for an age group into the total population in the age group. “x” in the following life table figures represents an age group, and “n” the number of years that span the interval. The subsequent columns follow.

${}_nq_x$ is an adjusted death rate, which calculates the proportion of the persons alive at the beginning of an age interval (x) that will survive and enter into the following age interval. As it is assumed that everyone eventually dies, the q_x for the last age interval is set at 1. This chapter

will use a simple formula in calculating q_x for the other age intervals which is: $1 - e^{(-m^*n)}$, which is a method developed by Fergany (1971). There are a number of methods with which to estimate ${}_nq_x$ including the Greville (1943) and Keyfitz-Frauenthal (1975) methods. The method chosen here was chosen for its elegance and simplicity, and it does not produce substantively different results that change the basic findings of this chapter.

l_x is a hypothetical number of persons who enter into a given age interval. This figure is commonly set at 100,000 who enter into the population at age 0 and progressively die off as they age into later age intervals. Each l_x for the age groups represents the remaining persons that have survived into their respective age groups out of the initial 100,000; that is, assuming that death rates remain the same for the entirety of a person's life, the l_x represents the population counts that are expected to reach the age interval out of the original population. The formula for calculating this is as follows: $l_x = (1 - q_{x-n}) / l_{x-n}$.

${}_nd_x$ is the number of deaths that occur to the hypothetical populations within each age interval. It is simply calculated as $d_x = l_x - l_{x+n}$.

${}_nL_x$ is the total number of "person years" that are lived by the persons in the age interval from the time of entry to the time of exit of the age interval. For example, if five people live through a ten-year interval of age 20 – 30, the ${}_{10}L_{20}$ would be 50 (because 5 people lived 10 years each across the interval. One way to calculate this would be ${}_nL_x = n(l_x + .5{}_nd_{x-n})$.

T_x sums the "total years lived" that all persons in respective age cohorts are expected to live throughout their lifetime. This is summed backward from the ${}_nL_x$ of the oldest age interval, as these figures are the sums of the ${}_nL_x$ values of the age interval in question and all subsequent age intervals. In mathematical script, it can be displayed as $T_x^{(-i)} = T_{x+n}^{(-i)} + {}_nL_x^{(-i)}$.

This finally bring us to the life expectancies, written as e_x . Life expectancies represent the number of remaining years of life that a given person within an age cohort is expected to live, once he/she reached that age interval. These are calculated as $e_x = T_x / l_x$.

Based on the calculations of collective error in Chapter 2, undercount by age cohort will be assumed to apply at the highest rate (with a base coefficient of 11.7%) for the 0 – 5 age cohort, and linearly decrease to 0 by age 65. Distributing these base rates by age cohort will be calculated as follows:

$$UndercountBaseRate_a = 11.7 \times (1 - \frac{a}{5 \times 13}) \quad [3.1]$$

Where a is the starting age of the five-year cohort. $a/5$ is also divided by 13 because there are 13 age categories in which to distribute the undercount rates.

Since the total undercount for Hispanics is estimated to be 1,115,468 (using a total undercount rate of 4.99%, Mulry 2006), the base coefficient rates of emigration will be multiplied by a constant (depicted as b below) in order to produce the total emigration count of 1,115,468. These emigration rates will be calculated thusly.

$$Undercount_{total} = \sum(b \times UndercountBaseRate_a \times HispanicCount_a) \quad [3.2]$$

Substituting $UndercountBaseRate_a$ in equation 3.2 with equation 3.1, $Undercount_{total}$ with 1,115,468, and solving for b produces the following formula.

$$b = \frac{1,115,468}{\sum[HispanicCount_a \times 11.7(1 - \frac{a}{5 \times 13})]} \quad [3.3]$$

DATA

Data on the age-sex structure of the population for Hispanics and NH whites will be collected by the U.S. Census Bureau's American Factfinder. Deaths will be gathered from NCHS' Multiple

Cause of Death Data,¹⁰ and the year of death will be adjusted into Census years (i.e. April 1, 1990 to March 31, 1991) to align with the presumed deaths applicable to the population counted in the Census in 1990. All deaths of “unknown” age will be assumed to be deaths to persons of age over 84. Deaths to Hispanics that report residence outside of the U.S. will not be included.

As noted in Chapter 2, Hispanic ethnicity was not included on death certificates for Oklahoma, New Hampshire, and Louisiana for 1990. Due to this lack of data, deaths to Hispanics that have occurred to residents of said states would need to be imputed for. As performed in Chapter 2, age-specific mortality rates for Hispanics that reside in states that include Hispanic origin on death certificates will be applied to the Hispanics counted in the states that lack a Hispanic item of death certificates.

The classification ratios that measure Hispanic death certificate undercount (i.e. CPS death counts divided by death certificate counts) with which Hispanics will be allocated deaths are found in Arias et al. (2010). These authors list that classification ratios are .96 for ages 0-24; 1.06 for ages 25-44; 1.05 for 45-54; 1.03 for ages 55-64; 1.07 for ages 65-74; and 1.05 for ages 75 and above. In order to maximize the possible undercount of Hispanic deaths, the .96 classification ratio for those aged 0-24 will be changed to 1 so that Hispanics will not “lose” deaths for this age category. These ratios will be multiplied by Hispanic deaths, thus “re-allocating” deaths to Hispanics. It will be assumed that all of the deaths being allocated to Hispanics will come from NH whites. Thus, adjusted death counts for NH whites will be the number of deaths minus the number of deaths that are allocated to Hispanics in the Hispanic death data adjustment scenario.

¹⁰ The invaluable SAS and STATA codebooks used to read the vital statistics data were written by Jean Roth at the National Bureau of Economic Research.

The alternate emigration counts will be taken from Chapter 2, which produces rates of emigration based on Van Hook et al. (2006), Passel et al. (2006), and Schwabish (2009). This chapter will produce new residual estimates of emigration by using 10-year survivorship ratios that are calculated through a life table based on five-year age intervals. The alternate sample-based estimates of emigration used the rates of emigration produced by the three aforementioned studies and weigh these rates by age, country/region of origin, and length of stay in the United States (see the Methods section of Chapter 2). For brevity's sake, these estimates of emigration will simply be referred to as the "Van Hook," "Passel," and "Schwabish" estimates and collectively as the "sample-based" estimates.

For the residual estimate of emigration it will be assumed that emigrants migrated at a steady pace throughout the decade, thus producing a count of emigration for 1990 at 1/10 of the total emigration for the ten-year period. The sample-based estimates of emigration will apply the rates of emigration calculated for 1990 to derive emigration counts (see equation 2.8). When adjusting the population count in the life tables, the population will be subtracted by $\frac{1}{2}$ of the emigrants for the year, as it will be assumed that emigration is constant throughout the year and that half of the emigrants (on average) will be exposed to the chance of death for 1990.

[TABLE 3.1]

Table 3.1 displays the population counts of Hispanics for 1990 that have been adjusted upward for undercount. The total undercount rate estimated for Hispanics was 4.99%, and the resulting 1,115,468 that was undercounted was distributed based on the proportions of undercount applied to that each age cohort (see Methods above). Undercount for non-Hispanic whites will be applied uniformly across all age groups at .68% (Mulry 2006). These adjusted

population counts will be used as alternate data inputs in the analyses in the subsequent sections.

RESULTS

Life Expectancies of Hispanics and Non Hispanic Whites, Before and After Death Data

Adjustment

[TABLE 3.2]

Table 3.2 displays mortality rates and life expectancies for Hispanics in 1990 based on life table estimates (of which tables utilizing unmodified data can be seen in the Appendix Tables). As can be seen in the life expectancy column (e_x), life expectancy for Hispanics surpasses 79 years of age (79.38) for unadjusted data and drops to 78.66 under data where deaths are adjusted upwards. These life expectancies are quite high considering the lower general socioeconomic status of Latinos, since low socioeconomic statuses are associated with lower life expectancies (Feinstein 1993). High life expectancies are particularly evident for those who have reached 85 years of age, as these persons are expected to live an additional 9 years (!).

[TABLE 3.3]

Contrasting life expectancies of Hispanics to NH whites (Table 3.3), this analysis confirms that Hispanics have a mortality advantage over NH whites. Life expectancy for NH whites at birth is calculated as 78.54 using unadjusted death data, which is almost one year younger than the corresponding life expectancy for Hispanics (79.38). The high estimate for life expectancies for NH whites is not much higher (78.58) and still trails behind the low estimate for Hispanics. Thus, we find support of the first hypothesis in Hispanics' higher life expectancy in comparison to NH whites, even after accounting for death undercount for Hispanics and overcount for NH whites.

When comparing mortality rates, one finds that it is not until cohorts reach 60 years of age that mortality rates are lower for Hispanics than for NH Whites. In all younger cohorts (aside from the 5-9 cohort), mortality rates for Hispanics are higher than those of NH Whites. Despite this finding, crude mortality rates are still higher for NH Whites than for Hispanics. Viewing the data skeptically, one may assume that mortality rates at younger ages should theoretically be reflective of mortality rates at older ages. It may thus be assumed that either Hispanic mortality rates at younger ages are inflated, and/or that mortality rates at older ages are deflated. While the calculations performed here have adjusted for data count issues in the numerator (deaths), count issues in the denominator (population counts) remain. The undercount of younger Hispanics, and/or the emigration of older Hispanics may lead to the artificial production of these seemingly contrasting mortality rates.

A Qualification of the Salmon Bias

[TABLE 3.4]

Table 3.4 displays the emigration counts of Hispanics for 1990. The first section of Table 3.4 displays the emigration counts that are based on the raw data, while the second displays emigration counts that are adjusted by the undercount of the Hispanic population in the 1990 Census using figures calculated in Table 3.1. The main differences between the two estimates are found in the residual estimates of emigration, where emigrants are added to each age cohort descending from nearly 20,000 for the 0-5 age cohort to 300 for the 60-64 age cohort. The undercount-adjusted estimates continue to produce “negative emigration” counts for Hispanics aged 0 – 19 using the residual method, although the counts of emigration for Hispanics aged 20 – 29 turn from negative to positive. Regardless, adjusting the 1990 Hispanic population counts by presumed rates of undercount produce counts of emigration that are

much more plausible than otherwise. The sample based estimates of emigration also increase by 8% in the youngest age cohort, down to .6% for the 60-64 year cohort when the population is adjusted for undercount, although no dramatic changes occur to the pattern of emigration by age.

[FIGURE 3.1]

Perhaps the strongest support for the “salmon bias” proper comes from the residual estimates, which calculates a steady stream of emigrants that move out of the country in later ages. Figure 1 displays the emigration rates of Hispanics for 1990 out of the total Hispanic population (adjusted for undercount). The residual method estimates that Hispanics emigrate at the highest rates at the latest age groups, which is a finding that is most closely mirrored by Schwabish’s estimates. This pattern bears evidence of the “salmon bias” hypothesis, which posits that older Hispanics return home to die. This artificially elevates life expectancies since emigrants are not captured in U.S. death counts (i.e. the numerators of death rates), despite the fact that they have been counted in the Census (i.e. the denominators of death rates).

While still demonstrating sizeable emigration rates for Hispanics at the latest age groups, the Van Hook and Passel estimates qualify the salmon bias phenomenon. While technically any exit of the country leaves the emigrants unexposed to the chance of death in U.S. statistics, these estimates do not display the highest exit rates for the populations that one would assume are the most prone to be ill; namely the persons of the oldest age groups. Rather, the Van Hook and Passel estimates display the highest emigration rates for those of working age between 25 and 44. This pattern likely reflects the return of temporary workers (Massey, Durand, and Malone 2002) rather than the return of mortally ill persons.

In sum, the estimates of emigration do confirm the exit of Hispanics of later ages (H2), particularly when using the residual methodology or Schwabish rates of emigration. However, the two other estimates of emigration suggest that the highest emigration rates apply to Hispanics of working ages. Thus, it is clear that emigration will serve to depress mortality rates. However, the argument that emigration is due to infirmity would need to be qualified, as high emigration rates are also estimated to apply to persons of generally healthier ages.

Revisiting Life Expectancies: Does the Mortality Paradox Still Hold When Considering Death Misclassification and the Salmon Bias?

This section will introduce two possible modifications that can be made to the initial population counts in order to estimate life expectancies. The first modification is the subtraction of emigrants from population counts, as these persons are not exposed to the risk of death and should hence not be included in mortality rates.¹¹ The second modification made will be the adjustment of population counts based on estimates of undercount, which was estimated to be 4.99% for Hispanics and .68% for NH whites in 1990.

[TABLE 3.5]

Table 3.5 displays the estimates of life expectancies at birth based on the varying combinations of modifications made to the data. The first three rows display the life expectancies of NH whites, including the figures that are produced when allocating deaths from whites and adjusting for population undercount. As mentioned earlier, adjusting for the overcount of NH white deaths increases life expectancies at birth from 78.54 to 78.58. Further adjusting for population undercount increased life expectancies to 78.68.

¹¹ Negative emigration rates will not be included in these analyses and will be assumed to be 0. Additionally, emigration for the 0-4 component will not be included, as these persons were not alive for the 1990 Census.

For Hispanics, emigration estimates alone do not alter the life expectancies dramatically, resulting in the falling of life expectancies from 79.38 to 79.29 under the lowest scenario. Combining both death and emigration adjustments, however, the lowest estimate of the life expectancies of Hispanics falls to .01 year less than the higher life expectancy of NH white life expectancies (without accounting for population undercount), and the second lowest matches said life expectancy. The other two low estimates of Hispanic life expectancies continue to surpass the high NH life expectancy. These low estimates of life expectancies for Hispanics are uniformly lower than the highest life expectancies for NH whites that use both adjusted death and population data. In short, when accounting for both data artifacts and the salmon bias, this analysis demonstrates that Hispanics continue to have high life expectancies that match (and potentially surpass) those of NH whites. However, when also adjusting NH white population data, this analysis finds that the highest estimate of the life expectancies of NH whites surpass the lowest estimate of Hispanics (not confirming H3).

Treating both sides of the data fairly, one would also be advised to adjust the population counts of Hispanics for undercount. When adjusting the Hispanic population for undercount, the positive health outcomes of Hispanics increase. While undercount adjustment also increases estimates of emigration (over two-fold using residual estimates), it also increases life expectancies by increasing the denominator in death rates. Under these scenarios, Hispanic life expectancy at birth ranges from 78.79 – 78.87 even after accounting for death undercount and

emigration. Under all of these scenarios (which arguably use the most accurate data), the life expectancies of Hispanics are higher than the highest life expectancies of NH whites.¹²

CONCLUSION

This chapter revisited the Hispanic Mortality Paradox by analyzing the life expectancies of Hispanics compared to those of NH whites after accounting for two alternate explanations of the Paradox. The first is the “data artifacts” explanation, which suggests that the underreporting of Hispanic deaths on death certificates may lower mortality rates. Death underreporting for Hispanics has been confirmed in earlier studies, and rates of underreporting were applied to this study based on Arias et al. (2008). Adjusting deaths upward for Hispanics and downward for NH whites, Hispanics are shown to continue to have a mortality advantage over whites.

The second alternate explanation of the Paradox is the salmon bias hypothesis, which posits that Hispanics of declining health return to their countries of origin, decreasing mortality rates by not appearing in death registries in the United States. Using a residual-based methodology, this research showed support of the salmon bias by detecting high emigration rates of elderly Hispanics. However, two of the sample-based estimates of emigration demonstrate that the highest rates of exit occur for Hispanics of working age. While emigration has been demonstrated to lower life expectancies of Hispanics, it does not do so at a magnitude that is large enough to close the gap in life expectancies between Hispanics and NH Whites. However, after accounting for both death underreporting *and* emigration of Hispanics while also accounting for death overreporting and undercount for NH whites, the life expectancies of Hispanics fall below the highest life expectancies of NH whites when using half of the emigration

¹² Life expectancies for Hispanics can also be increased when accounting for yearly immigration (which this study does not do), as this study assumes that mortality for immigrants arriving in 1990 (the year for which these data apply) is 0.

estimates. Accounting for the undercount of Hispanics in the 1990 Census (which would arguably produce the most reliable figures), raises the life expectancies of Hispanics to levels that are higher than those of NH whites.

DISCUSSION

This research qualifies the salmon bias phenomenon by demonstrating that, according to two emigration estimates, the highest rates of emigration are found for persons of working age. Thus, while emigration continues to be associated with higher ages – presumably for persons who return to their country of origin to die – it is also associated with temporary work patterns. Thus, in addition to circular migration for persons who seasonally work in the U.S., this research bears evidence that the salmon bias needs be qualified by the population that simply arrives, works, and exits.

The existence of the Hispanic Mortality Paradox is partially demonstrated in this research, as life expectancies of Hispanics under what is presumably the most accurate data-scenario (death undercount, emigration, and undercount) surpass the highest life expectancies of the highest life expectancies of NH whites. However, using the most disadvantageous assumptions for the health of Hispanics, namely not accounting for the undercount of Hispanics for 1990, the higher health outcomes of Hispanics vis-à-vis NH whites becomes challenged.

This research underscores the importance of accounting for data problems in population research, as varying assumptions and adjustments to data can overturn theoretical conclusions. The problems with residual counts of emigration is a case in point, as residual methods revealed a mountainous level of “negative emigration” for persons aged 10 – 19 in 2000 (0-9 in 1990). A fundamental flaw in these calculations is that residual migration should never be negative for a population for which immigration does not apply, as has been the case

here. While not displayed in this Chapter (which will be covered in the Conclusion), residual emigration rates of children born between Censuses would be estimated at around 80,000 for 2000, suggesting that the youngest children are undercounted at much higher rates than are assumed. That means that it is not that more persons 10 – 19 magically appeared in 2000; rather, it is that these persons were not counted in 1990. Undercount by age need to be adjusted even further for the youngest populations, and it is likely that undercount for Hispanics in 1990 is greater than 4.99% as has been found in previous research (i.e. Mulry 2006).

The differing means of data collection also have an influence on the health results of Hispanics. Previous studies have identified the differentials between Hispanic self-reporting (as on the Census) in contrast to identification by a third-party member (as on death certificates). Accounting for this ethnic misclassification on death certificates would be important in producing more realistic figures of Hispanic population data (Arias et al. 2008 is of particular use in this aim).

In light of the problems that are evident in the decennial Census, demographic analysis will be of tremendous use in assessing the Census' quality and in further exploring rates of undercount. Demographic analysis relies on a simple formula that includes births, deaths, and migration, each derived from data sources independent of the decennial Census. This dissertation will continue in the steps that are necessary to conduct a demographic analysis of the Hispanic population to arrive at alternate counts of Hispanics in the Conclusion. At the close of Chapter 3, alternate estimates of emigration have been calculated and adjustments to death counts and baseline populations (Census 1990, when it will be used) have been provided. Migration can be gathered from alternate sources such as Passel and Suro (2005). What remains

to be tackled is driver of populations, namely births. As will be seen in the subsequent chapters, even this data set is in need of cleaning.

REFERENCES

- Abraido-Lanza, Ana, Bruce Dohrenwend, Daisy Ng-Mak, and Blake Turner. 1999. "The Latino Mortality Paradox: A Test of the 'Salmon Bias' and Healthy Migrant Hypothesis." *American Journal of Public Health* 89(10):1543-1548.
- Arias, Elizabeth 1998. "The Demography of Assimilation: The Case of Cubans in the United States." Ph.D. Dissertation, Department of Sociology, University of Wisconsin, Madison, WI.
- Arias, Elizabeth, Karl Eschbach, William Schauman, Eric Backlund, and Paul Sorlie. 2010. "The Hispanic Mortality Advantage and Ethnic Misclassification on US Death Certificates." *American Journal of Public Health*. 100(S1):171-177.
- Arias, Elizabeth., William Schauman, Karl Eschbach, Paul Sorlie, and Eric Backlund. 2008. "The Validity of Race and Hispanic Origin Reporting on Death Certificates in the United States." *Vital Health Stat* 2:1-24.
- Chen, Jiajian, Edward Ng, and Russell Wilkins. 1996. "The Health of Canada's Immigrants," *Health Reports* 7:33-50.
- Donovan, John, Edouard d'Espaignet, Carolyn Merton, and Marijke van Ommerson. 1992. "Immigrants in Australia: A Health Profile." *Australian Institute of Health and Welfare, Ethnic Health Series*, 1. Canberra: AGPS.
- Elo, Cassio, and Irma Turra. 2008. "The Impact of Salmon Bias on the Hispanic Mortality Advantage: New Evidence from Social Security Data." *Population Research Policy Review* 27:515-530.
- Elo, Irma, Cassio Turra, Bert Kestenbaum, and B. Ferguson. 2004. "Mortality Among Elderly Hispanics in the United States: Past Evidence and New Results." *Demography* 41(1):109-128.
- Feinstein, Jonathan. 1993. "The Relationship between Socioeconomic Status and Health: A Review of the Literature." *The Milbank Quarterly* 71(2):279-322.
- Fergany, Nader. 1971. "On the Human Survivorship Function and Life Table Construction." *Demography* 8(3):331-334.
- Gibbs, Jim, Gregory Harper, Mark Rubin, and Hyon Shin. 2003. "Evaluating Components of International Migration: Native-Born Emigrants." *Population Division Working Paper Series* No. 63. Washington, D.C.: U.S. Census Bureau.
- Greville, Thomas. 1943. "Short methods of Constructing Abridged Life Tables." *Record of the American Institute of Actuaries* 32(65):29-42.

Hummer, Robert, Daniel Powers, Starling Pullum, Ginger Gossman, and Parker Frisbie. 2007. "Paradox Found (Again): Infant Mortality Among the Mexican-Origin Population in the United States." *Demography* 44(3):441-457.

Hummer, Robert, Maureen Benjamins, and Richard Rogers. 2004. "Racial and Ethnic Disparities in Health and Mortality among the U.S. Elderly Population." Pp. 53-94 in *Understanding Racial and Ethnic Difference in Health in Late Life: A Research Agenda*, edited by R. Bulatao and N. Anderson. Washington, D.C.: National Academy Press.

Hummer, Robert, Richard Rogers, Charles Nam, and Felicia LeClere. 1999. "Race/Ethnicity, Nativity, and U.S. Adult Mortality." *Social Science Quarterly* 80:136-153.

Hummer, Robert, Richard Rogers, Sarit Amir, Douglas Forbes, and Parker Frisbie. 2000. "Adult Mortality Differentials Among Hispanic Subgroups and Non-Hispanic Whites." *Social Science Quarterly* 81:459-476.

Keyfitz, Nathan and James Frauenthal. 1975. "An Improved Life Table Method." *Biometrics* 31:889-899.

Kintner, Hallie. 2004. "The Life Table." Pp. 301-340 in *The Methods and Materials of Demography*, edited by Jacob Siegel and David Swanson. Boston, MA: Elsevier Academic Press.

Liao, Youlian, Richard Cooper, Guichan Cao, Ramon Durazo-Arvizu, Jay Kaufman, Amy Luke, and Daniel McGee. 1998. "Mortality Patterns Among Adult Hispanics: Findings from the NHIS 1986 to 1990." *American Journal of Public Health* 88:227-32.

Markides, Kyriakos and Jeannine Coreil. 1986. "The Health of Southwestern Hispanics: An Epidemiologic Paradox." *Public Health Reports*, 101:253-265.

Markides, Kyriakos, Laura Rudkin, Ronald Angel, and David Espino. 1997. "Health Status of Hispanic Elderly in the United States." Pp. 285-300 in *Racial and Ethnic Differences in the Health of Older Americans*, edited by L.G. Martin and B.J. Soldo. Washington, DC: National Academy Press.

Massey, Douglas. Jorge Durand, and Nolan Malone. 2002. *Beyond Smoke and Mirrors: Mexican Immigration in an Era of Economic Integration*. New York, NY: Russell Sage Foundation.

Mulry, Mary. 2006. *Summary of Accuracy and Coverage Evaluation for Census 2000*. Washington, DC: U.S. Census Bureau.

O'Hare, William. 2009. *Why are Young Children Missed So Often in the Census?* Kids Count Working Paper, Annie E. Casey Foundation, Baltimore, MD.

Palloni, Alberto and Elizabeth Arias. 2004. "Paradox Lost: Explaining the Hispanic Adult Mortality Advantage." *Demography* 41:385-415.

Passel, Jeffrey and Roberto Suro. 2005. *Rise, Peak, and Decline: Trends in U.S. Immigration 1992-2004*. Washington, DC: Pew Hispanic Center.

Rendall, Michael and Bema Torr. 2008. "Emigration and Schooling among Second-Generation Mexican-American Children." *IMR* 42:729-738.

Rosenberg, Harry, Jeffrey Maurer, Paul Sorlie, Norman Johnson, Marian MacDorman, Donna Hoyert, James Spitler, and Chester Scott. 1999. "Quality of Death Rates by Race and Hispanic Origin: A Summary of Current Research, 1999." *Department of Health and Human Services Publications* No PHSP 99-132B.

Smith, David, and Benjamin Bradshaw. 2006. "Rethinking the Hispanic Paradox: Death Rates and Life Expectancy for US Non-Hispanic White and Hispanic Populations." *American Journal of Public Health* 96:1686-1692.

Sorlie, Paul, Erick Backlund, Norman Johnson, and Eugene Rogot. 1993. "Mortality by Hispanic Status in the United States." *JAMA* 270:2646-2648.

Stephen, Elizabeth, Karen Foote, Gerry Hendershot, and Charlotte Shoenborn. 1994. "Health of the Foreign-Born Population: United States, 1989-90. Advance Data." *Vital and Health Statistics*, 241, Washington, DC: National Center for Health Statistics.

Van Hook, Jennifer, Weiwei Zhang, Frank Bean, and Jeffrey Passel. 2006. "Foreign-Born Emigration: A New Approach and Estimates Based on Matched CPS Files." *Demography* 43:361-382.

Wingate, Martha and Greg Alexander. 2005. "The Healthy Migrant Theory: Variations in Pregnancy Outcomes among US-Born Migrants." *Social Science & Medicine* 62:491-498.

TABLES

Table 3.1: Hispanic Count Adjusted for Undercount

Age	Total Count 1990	Base Undercount %	Adjusted Undercount %	Undercount	Adjusted Count 1990
Under 5 years	2,387,524	11.69	8.14	194,363	2,581,887
5 to 9 years	2,193,852	10.80	7.51	164,858	2,358,710
10 to 14 years	2,001,617	9.90	6.89	137,878	2,139,495
15 to 19 years	2,053,957	9.00	6.26	128,622	2,182,579
20 to 24 years	2,304,441	8.10	5.64	129,876	2,434,317
25 to 29 years	2,341,239	7.20	5.01	117,289	2,458,528
30 to 34 years	2,062,303	6.30	4.38	90,401	2,152,704
35 to 39 years	1,660,726	5.40	3.76	62,398	1,723,124
40 to 44 years	1,284,268	4.50	3.13	40,211	1,324,479
45 to 49 years	953,910	3.60	2.50	23,894	977,804
50 to 54 years	755,989	2.70	1.88	14,202	770,191
55 to 59 years	639,308	1.80	1.25	8,007	647,315
60 to 64 years	553,642	0.90	0.63	3,467	557,109
65 to 69 years	436,257	0	0	0	436,257
70 to 74 years	286,772	0	0	0	286,772
75 to 79 years	213,265	0	0	0	213,265
80 to 84 years	130,425	0	0	0	130,425
85 years and over	94,564	0	0	0	94,564
Total	22,354,059	7.169	4.99	1,115,468	23,469,527

Table 3.2: Hispanic Mortality Rates and Life Expectancies, with and without Death Adjustments: 1990

Age Cohort	Population	Death Adjustment	Deaths	Deaths adjusted	$n\mathbf{m}_x$	$n\mathbf{m}_x$, adjusted	e_x	e_x adjusted
Under 5 years	2,387,524	1	5,456	5,456	0.002285	0.002285	79.38	78.66
5 to 9 years	2,193,852	1	417	417	0.00019	0.000190	75.26	74.53
10 to 14 years	2,001,617	1	483	483	0.000241	0.000241	70.33	69.60
15 to 19 years	2,053,957	1	1,837	1,837	0.000894	0.000894	65.41	64.68
20 to 24 years	2,304,441	1	2,737	2,737	0.001188	0.001187	60.69	59.96
25 to 29 years	2,341,239	1.06	3,147	3,336	0.001344	0.001424	56.04	55.30
30 to 34 years	2,062,303	1.06	3,443	3,650	0.001669	0.001769	51.40	50.68
35 to 39 years	1,660,726	1.06	3,363	3,565	0.002025	0.002146	46.81	46.11
40 to 44 years	1,284,268	1.06	3,208	3,400	0.002498	0.002647	42.26	41.58
45 to 49 years	953,910	1.05	3,096	3,251	0.003246	0.003407	37.76	37.10
50 to 54 years	755,989	1.05	3,421	3,592	0.004525	0.004751	33.34	32.69
55 to 59 years	639,308	1.03	4,384	4,516	0.006857	0.007063	29.04	28.42
60 to 64 years	553,642	1.03	5,700	5,871	0.010295	0.010604	24.97	24.35
65 to 69 years	436,257	1.07	6,704	7,173	0.015367	0.016442	21.16	20.54
70 to 74 years	286,772	1.07	6,699	7,168	0.02336	0.024995	17.65	17.09
75 to 79 years	213,265	1.05	6,991	7,341	0.032781	0.034419	14.54	14.04
80 to 84 years	130,425	1.05	6,961	7,309	0.053372	0.056040	11.69	11.22
85 years and over	94,564	1.05	9,916	10,412	0.10486	0.110103	9.54	9.08

Table 3.3: Non Hispanic White Mortality Rates and Life Expectancies, with and without Death Adjustments: 1990

Age Cohort	Population	Deaths allocated to Hispanics	Deaths	Deaths adjusted	$n\mathbf{m}_x$	$n\mathbf{m}_x$, adjusted	e_x	e_x adjusted
Under 5 years	11,194,346	0	22,660	22,660	0.002024	0.002024	78.54	78.58
5 to 9 years	12,303,903	0	2,364	2,364	0.000192	0.000192	74.32	74.35
10 to 14 years	12,882,540	0	2,712	2,712	0.000211	0.000211	69.39	69.42
15 to 19 years	12,759,934	0	9,298	9,298	0.000729	0.000729	64.46	64.49
20 to 24 years	11,594,742	0	11,657	11,657	0.001005	0.001005	59.68	59.72
25 to 29 years	11,990,863	189	14,921	14,732	0.001244	0.001229	54.97	55.01
30 to 34 years	13,365,410	207	19,399	19,192	0.001451	0.001436	50.30	50.33
35 to 39 years	15,665,973	202	23,541	23,339	0.001503	0.00149	45.65	45.68
40 to 44 years	16,135,362	192	28,961	28,769	0.001795	0.001783	40.97	41.00
45 to 49 years	14,908,211	155	35,393	35,238	0.002374	0.002364	36.32	36.34
50 to 54 years	13,478,949	171	46,436	46,265	0.003445	0.003432	31.72	31.75
55 to 59 years	10,545,669	132	71,005	70,873	0.006733	0.006721	27.23	27.25
60 to 64 years	8,482,012	171	117,596	117,425	0.013864	0.013844	23.08	23.10
65 to 69 years	7,650,827	469	170,549	170,080	0.022292	0.02223	19.56	19.58
70 to 74 years	7,327,622	469	212,792	212,323	0.02904	0.028976	16.57	16.59
75 to 79 years	6,307,373	350	248,024	247,674	0.039323	0.039267	13.78	13.80
80 to 84 years	4,284,906	348	251,753	251,405	0.058753	0.058672	11.25	11.26
85 years and over	3,674,132	496	395,805	395,309	0.107727	0.107593	9.28	9.29

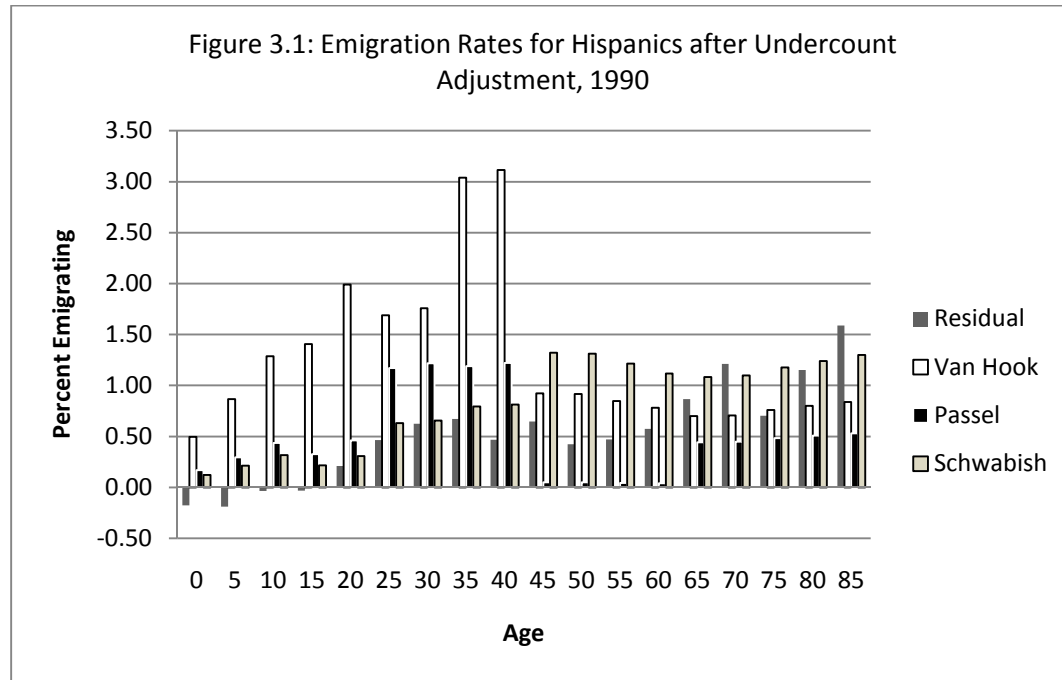
Table 3.4: Emigration counts for Hispanics before and after undercount Adjustment, 1990

	Emigration, 1990				Emigration, 1990 Undercount			
Age	Van				Van			
	Residual	Hook	Passel	Schwabish	Residual	Hook	Passel	Schwabish
Under 5 years	-23,985	11,804	4,009	2,892	-4,552	12,765	4,335	3,127
5 to 9 years	-20,958	19,009	6,456	4,657	-4,479	20,437	6,941	5,007
10 to 14 years	-14,499	25,759	8,749	6,311	-724	27,533	9,352	6,746
15 to 19 years	-13,547	28,908	6,682	4,464	-701	30,718	7,100	4,744
20 to 24 years	-7,861	45,858	10,600	7,081	5,105	48,443	11,197	7,480
25 to 29 years	-247	39,527	27,409	14,743	11,454	41,507	28,782	15,482
30 to 34 years	4,440	36,242	25,131	13,518	13,451	37,831	26,233	14,111
35 to 39 years	5,353	50,467	19,780	13,170	11,563	52,363	20,523	13,665
40 to 44 years	2,193	39,983	15,671	10,434	6,182	41,235	16,162	10,761
45 to 49 years	3,961	8,804	470	12,604	6,316	9,025	482	12,920
50 to 54 years	1,890	6,923	369	9,911	3,269	7,053	376	10,097
55 to 59 years	2,293	5,420	289	7,759	3,044	5,488	293	7,856
60 to 64 years	2,895	4,327	231	6,195	3,198	4,354	232	6,234
65 to 69 years	3,784	3,047	1,937	4,725	3,784	3,047	1,937	4,725
70 to 74 years	3,479	2,029	1,290	3,147	3,479	2,029	1,290	3,147
75 to 79 years	1,501	1,620	1,030	2,512	1,501	1,620	1,030	2,512
80 to 84 years	1,501	1,044	663	1,618	1,501	1,044	663	1,618
85 years and over	1,501	792	503	1,228	1,501	792	503	1,228
Total	-46,307	331,563	131,269	126,969	64,890	347,284	137,431	131,458

Table 3.5: Life Expectancies of Non-Hispanic Whites and Hispanics Using Varying Data Adjustments

Samples	Adjustments	1990 Emigration Estimate	Life Expectancy Age 0
NH White	None	n/a	78.54
	Deaths	n/a	78.58
	Deaths, Undercount	n/a	78.68
Hispanic	None	n/a	79.38
	Deaths	n/a	78.66
Residual Emigration Component, Hispanics	Emigration	34,790	79.31
	Deaths, Emigration	31,449	78.59
	Deaths, Emigration, Undercount	71,610	78.87
Van Hook et al. (2006) Emigration Component, Hispanics	Emigration	331,563	79.30
	Deaths, Emigration	331,563	78.58
	Deaths, Emigration, Undercount	347,284	78.80
Passel et al. (2006) Emigration Component, Hispanics	Emigration	131,269	79.34
	Deaths, Emigration	131,269	78.62
	Deaths, Emigration, Undercount	137,431	78.84
Schwabish (2009) Emigration Component, Hispanics	Emigration	126,969	79.29
	Deaths, Emigration	126,969	78.57
	Deaths, Emigration, Undercount	131,458	78.79

FIGURE



APPENDIX TABLES

Table 3.6: Life Table for Hispanics, 1990, Unadjusted Data

Age Cohort	Population	Deaths	$n m_x$	$n q_x$	l_x	$n d_x$	$n L_x$	T_x	e_x
Under 5 years	2,387,524	5,456	0.002285	0.011361	100,000	1,136	497,154	7,937,701	79.38
5 to 9 years	2,193,852	417	0.00019	0.00095	98,864	94	494,085	7,440,547	75.26
10 to 14 years	2,001,617	483	0.000241	0.001206	98,770	119	493,552	6,946,462	70.33
15 to 19 years	2,053,957	1,837	0.000894	0.004462	98,651	440	492,153	6,452,910	65.41
20 to 24 years	2,304,441	2,737	0.001188	0.005921	98,211	581	489,598	5,960,757	60.69
25 to 29 years	2,341,239	3,147	0.001344	0.006698	97,629	654	486,509	5,471,159	56.04
30 to 34 years	2,062,303	3,443	0.001669	0.008313	96,975	806	482,858	4,984,649	51.40
35 to 39 years	1,660,726	3,363	0.002025	0.010074	96,169	969	478,420	4,501,791	46.81
40 to 44 years	1,284,268	3,208	0.002498	0.012412	95,200	1,182	473,041	4,023,372	42.26
45 to 49 years	953,910	3,096	0.003246	0.016097	94,019	1,513	466,300	3,550,330	37.76
50 to 54 years	755,989	3,421	0.004525	0.022372	92,505	2,070	457,333	3,084,030	33.34
55 to 59 years	639,308	4,384	0.006857	0.033706	90,436	3,048	444,515	2,626,697	29.04
60 to 64 years	553,642	5,700	0.010295	0.050175	87,388	4,385	425,882	2,182,182	24.97
65 to 69 years	436,257	6,704	0.015367	0.073958	83,003	6,139	399,471	1,756,300	21.16
70 to 74 years	286,772	6,699	0.02336	0.110237	76,864	8,473	362,725	1,356,829	17.65
75 to 79 years	213,265	6,991	0.032781	0.151177	68,391	10,339	315,401	994,104	14.54
80 to 84 years	130,425	6,961	0.053372	0.234218	58,052	13,597	254,757	678,702	11.69
85 years and over	94,564	9,916	0.10486	1	44,455	44,455	423,945	423,945	9.54

Table 3.7: Life Table for Non Hispanic Whites, 1990, Unadjusted Data

Age Cohort	Population	Deaths	$n m_x$	$n q_x$	l_x	$n d_x$	$n L_x$	T_x	e_x
Under 5 years	11,194,346	22,660	0.002024	0.01007	100,000	1,007	497,478	7,854,396	78.54
5 to 9 years	12,303,903	2,364	0.000192	0.00096	98,993	95	494,727	7,356,918	74.32
10 to 14 years	12,882,540	2,712	0.000211	0.001052	98,898	104	494,230	6,862,191	69.39
15 to 19 years	12,759,934	9,298	0.000729	0.003637	98,794	359	493,071	6,367,961	64.46
20 to 24 years	11,594,742	11,657	0.001005	0.005014	98,435	494	490,938	5,874,891	59.68
25 to 29 years	11,990,863	14,921	0.001244	0.006203	97,941	607	488,185	5,383,953	54.97
30 to 34 years	13,365,410	19,399	0.001451	0.007231	97,334	704	484,906	4,895,768	50.30
35 to 39 years	15,665,973	23,541	0.001503	0.007485	96,630	723	481,338	4,410,862	45.65
40 to 44 years	16,135,362	28,961	0.001795	0.008934	95,906	857	477,387	3,929,524	40.97
45 to 49 years	14,908,211	35,393	0.002374	0.0118	95,050	1,122	472,438	3,452,137	36.32
50 to 54 years	13,478,949	46,436	0.003445	0.017078	93,928	1,604	465,618	2,979,698	31.72
55 to 59 years	10,545,669	71,005	0.006733	0.033105	92,324	3,056	453,936	2,514,080	27.23
60 to 64 years	8,482,012	117,596	0.013864	0.066973	89,268	5,978	431,219	2,060,145	23.08
65 to 69 years	7,650,827	170,549	0.022292	0.105471	83,289	8,785	394,076	1,628,926	19.56
70 to 74 years	7,327,622	212,792	0.02904	0.135149	74,504	10,069	346,740	1,234,850	16.57
75 to 79 years	6,307,373	248,024	0.039323	0.178493	64,435	11,501	292,481	888,110	13.78
80 to 84 years	4,284,906	251,753	0.058753	0.25455	52,934	13,474	229,337	595,629	11.25
85 years and over	3,674,132	395,805	0.107727	1	39,460	39,460	366,291	366,291	9.28

CHAPTER 4

PREDICTING INTERETHNIC CHILDBIRTH BY THE CHARACTERISTICS OF MOTHERS:

A PEEK INTO THE MELTING POT

The United States is home to a smorgasbord of persons of varying social backgrounds. Persons from all walks of life, from all over the world have found the U.S. their home, creating enclaves of ethnic neighborhoods and/or mixing in with persons of different ethnic backgrounds. Surely, intraethnic families are the norm. However, interethnic families have a sizeable presence in the U.S., breaking down the boundaries that separate traditionally-defined ethnicities and creating ethnic blends. This research seeks to explain some of the factors that contribute to interethnic family formation. What are the dynamics that underlie the formation of interethnic families? What kinds of people are more likely to have interethnic children?

There has been a wealth of literature that explores both Hispanic family formation and interethnic relationships. A neglected area of research combines these two literatures in explaining the demographic factors that influence to interethnic childbirth. This chapter is dedicated to addressing this gap in the literature by identifying the demographic characteristics of mothers who give birth to interethnic children. Using the same variables that have been useful for predicting Hispanic family formation and interethnic relationships, this research takes an exploratory investigative perspective in seeking to find the commonalities and differences that interethnic parentage has with the aforementioned literatures. This research will work under ethnic definitions used by the Census Bureau, where “ethnicity” includes the categories “Hispanic” and “Non Hispanic.”

This research begins by summarizing some of the key findings in the existing literature on Hispanic family formation and interethnic relationships. Using birth certificate data, the important variables in the previous literatures will be used to predict interethnic childbirth. The Discussion section will give meaning to the statistical results, largely demonstrating that race and ethnicity strongly predict the ethnicity of one's reproductive partner. However, it will also be demonstrated that social context also has an influence on the ethnicity of one's partner, as the networks that one is embedded in serve to "break down" ethnic divisions. The Conclusion section will contrast the findings of this research on interethnic childbirth with the literature on interethnic relationships in general.

HISPANIC FAMILY FORMATION

Family formation for Hispanics can provide a knowledge base in understanding interethnic childbirth. This section will focus on the interplay between relationship type (i.e. cohabitation versus marriage), childbearing, and the demographic factors that are related to Hispanic family formation.

Cohabitation for Hispanics may be partially influenced by the popularity of the consensual union in Latin America, which is a historical form of co-residential union that is similar to marriage (Castro 2002). Consensual unions can be seen as alternate forms of "marriage" for persons of fewer economic resources (Oropesa et al. 2000; Oropesa and Landale 2004). In the United States, the consensual union can take the form of cohabiting families which is a familial arrangement that is more common for Hispanics than for whites (Bumpass and Lu 2000) and does not appear to affect the stability of Hispanic families (Manning, Smock, and Majumdar 2004). Cohabitation for Hispanics may be a factor of generational effects, as rates of

cohabitation for Hispanics tend to increase generationally (Brown, Van Hook, and Glick 2008; Landale, Oropesa, and Bradatan 2006).

Just as the consensual union is influenced by a person's economic resources in Latin America, cohabitation is similarly more common for those who have fewer economic resources and education in the United States (Bumpass and Lu 2000; Casper and Bianchi 2002; Manning and Brown 2006). Adding to the stresses of having fewer economic resources, cohabiters also tend to be younger and have less family member support (Bumpass and Lu 2000; Eggebeen 2005).

Cohabitation leads to higher fertility rates, and thus to family formation for Hispanics (Manning et al. 2004; Musick 2002). Correlates of cohabitation, including education and socioeconomic status, also influence fertility rates. Education in particular plays a strong role in childbirth, suppressing fertility rates for Hispanics (as well as for other ethnicities, Bumpass and Lu 2000; Glick et al. 2006; Musick 2002; Stockard, et al. 2009; Wildsmith and Raley 2006). Socioeconomic status and employment also have effects on childbirth (Wildsmith and Raley 2006).

Nonmarital childbirth is associated with the same variables that are related to family formation for Hispanics. For example, education and the education of one's parents have effects on nonmarital childbearing (Glick et al. 2006). Being raised in a family with a higher socioeconomic background also suppresses nonmarital fertility (Wildsmith and Raley 2006). Generation and culture also influence nonmarital fertility, as the cultural context in which one grows up influences the perceptions of gender roles, contraceptive use, and educational opportunities that frame one's expected life-course (Musick 2002). Birth cohorts that grow up

with higher rates of less traditional family structures (e.g. children born out of wedlock), for example, tend to have higher nonmarital fertility rates themselves (Stockard et al. 2009).

INTERETHNIC RELATIONSHIPS

As with the literature on Hispanic family formation, the literature on interethnic relationships is also useful in exploring the factors that contribute to interethnic childbirth. Interethnic relationships are more common for cohabiting couples than for married couples. Joyner and Kao (2005) go so far as to argue that, over the life-course of a person, persons are more likely to be in interethnic relationships before marriage, although they eventually settle with persons of the same ethnicity when they marry. Partially due to social pressures and family support (Hohmann-Marriott and Amato 2008), cohabitation may be seen as an alternative to marriage for mixed-race couples (Blackwell and Lichter 2000; Lichter and Qian 2004; Qian and Lichter 2007).

Interethnic marriages, in contrast to interethnic cohabiting relationships, tend to be more common with increasing education (Batson, Qian, and Lichter 2006; Rosenfeld 2005; Qian and Lichter 2007). One's background also plays a role in interethnic marriage, as having parents with higher levels of education is related to interethnic marriage (King and Bratter 2007).

Race also interplays with socioeconomic status in marriage. Heterogamous marriages (but not cohabitations) suggest spousal trades: high education in one spouse is associated with higher color status in another (Blackwell and Lichter 2000). Interethnic families tend to have higher education and family incomes than intraethnic Hispanic families, although interethnic families still trail behind non Hispanic white families in education and income (Lee and Edmonston 2005).

Race and acculturation add additional dimensions in out-marriage for Hispanics, as Hispanic White women are more likely to have interracial relationships than are Hispanic non-

Whites (King and Bratter 2007; Qian and Cobas 2004). Native-born Hispanics are more likely to intermarry (as well as cohabit) with whites than are foreign-born Hispanics, suggesting that acculturation also influences interethnic relationships (Qian and Lichter 2007; Qian and Cobas 2004).

Marriage market effects also play a role in interethnic marriage, as interethnic marriage is related to residing in diverse areas (Lichter et al. 2007). The larger the representation of one's ethnicity within a state, the more likely it will be that a person will marry within his/her group (Kalmijin and Tubergen 2010).

INTERETHNIC CHILDREN

This research asks a very specific question that has little been discussed: predictors of having interethnic children. The above literatures are all relevant to the topic at hand, yet producing predictions of interethnic childbirth specifically on the basis of the previous literatures will not be comparing apples to apples.

The closest study known by the author that touches upon interethnic childbirth by marital status is mentioned in passing by Landale et al. (2006), who provide rates of "exogamous" births by marital status by Hispanic type. Contrary to what one would expect, given that cohabitation is more popular for interethnic relationships, interethnic children are far less frequently born to interethnic unmarried partners than for married partners. However, these low rates of interethnic children in cohabiting partnerships may partially be a function of missing father ethnicities, which is on the order of tens-of-times more frequent for unmarried mothers.

Some common threads exist across the Hispanic family formation and interethnic relationship literatures. In sum, Hispanic family formation is related to cohabitation (and marital

status in general), generation, economic resources, and education. The interethnic relationship literature also includes effects that result from the aforementioned variables, as well as by race and the marriage market. The degree to which these variables interact, however, is in question. The situation gets convoluted when viewing education's relationship with Hispanic family formation and interethnic relationships, as an example. Education is negatively related to Hispanic family formation, yet positively related to interethnic relationships. But cohabitation is more common for less educated persons, as well as for interethnic relationships. Education's relationship with interethnic childbirth is thus unclear, and may be contingent upon factors such as marital status.

This research will use the same variables mentioned above in predicting interethnic childbirth, namely age, race, education, nativity, number of live births,¹³ and the percentage of persons that are Hispanic in the Public Use Microdata Area. This research will test the differences that these influences have on interethnic childbirth by the sex of the Hispanic parent as well as by marital status, resulting in four models. Dividing the model into four subsets is performed because it is hypothesized that the effects of the variables would be different under different conditions (i.e. by marital status and the ethnicity of the mother).¹⁴

¹³ Cohabiting relationships tend to feature more complex previous relationships and previous children (Bumpass and Lu 2000; Carlson and Furstenberg 2006; Osborne et al. 2007; Qian, Lichter, and Mellot 2005), making this included among predictor variables.

¹⁴ This research will take an exploratory approach without hypothesizing all of the (28) relationships that the independent variables have on each of the four models. However, some expected findings can illuminate the logic behind separating the data into the four samples between values of marital status and ethnicity of the mother. Variation by marital status is hypothesized to be related to education, as higher educational attainment is hypothesized to be positively related to interethnic childbirth when the pair is married, but negatively related when not married. Variation by the sex of the Hispanic parent is hypothesized to be related to the percent of persons that are Hispanic in the area, which is hypothesized to be positively related to interethnic childbirth when the mother is not Hispanic, and negatively related when the mother is Hispanic. Additionally, differences are expected across groups, as the number of live

DATA AND METHODS

All data are drawn from the National Center for Health Statistics (NCHS) micro data for mothers who state residence in the United States. Children who would be between the ages of 0 and 10 at the time of the 2000 Census were sampled. All demographic variables refer to the mother due to the large quantity of missing data for fathers. The mother's demographic characteristics of age, race, and foreign-born are drawn directly from the data.

The "state number of births" sums the number of births that have occurred in a state for a given Census year, in units of 10,000 and rounded to the nearest integer. Percent Hispanic in a state is calculated by state by year and grouped into five-percent units, which are also rounded to the nearest integer.¹⁵ Mother's education is re-coded into no high school, high school (the omitted category), some college, and bachelor's and above. Mother's race is classified into white, black, and other race. Imputed values on the mother's race are coded to missing. Number of live births is truncated to 0 to 10 and above, and is treated as a continuous variable.

Multivariate logistic regression will be used to determine the effects that the variables have on the odds that mothers bear interethnic children. The data will be divided into four models by marital status (married, not married) and ethnicity of the mother (Hispanic, non Hispanic).

[TABLE 4.1]

births is hypothesized to be positively related to interethnic childbirth for non Hispanic mothers who are not married, negatively related for Hispanic mothers who are not married, and negatively related to married families in general. Due to these convoluted relationships, no specific hypotheses will be made and the analyses will be separated into four models.

¹⁵ State population and percent Hispanic in state (in intervals of 5) were rounded to integers due to the enormity of the dataset (totaling 43,456,600) that would pose a strain on computations. Future research and technologies can include more detail for these figures, although the findings are not likely to change noticeably.

RESULTS

Table 4.1 displays simple rates at which mothers have interethnic children within the four categories. Hispanic mothers are much more likely to bear interethnic children than non Hispanic mothers. Across marital status, married Hispanic women are more likely to have interethnic children than unmarried Hispanic women. The relationship is the opposite for non Hispanic women, however, as unmarried non Hispanic women are more than twice as likely to have interethnic children than married non Hispanic women.¹⁶

[TABLE 4.2]

Table 4.2 displays the results of the multivariate logistic regression models which predict interethnic childbirth, with different models based on the marital status and ethnicity of the mother. The first thing to note is that marital status does not change the direction of the effects for nearly all of the variables. The only directional change exists for Non Hispanic mothers who are neither black nor white (comprising 5.8% of non Hispanic births, not shown) across Models 3 and 4, for whom interethnic childbirth was more likely to occur for unmarried mothers and less likely to occur for married mothers. Two statistical differences exist between marital status, as the population size (for Hispanic mothers, Model 2) and “some college” (for non Hispanic mothers, Model 4) lose significance for mothers who are not married.¹⁷ What this means is that the only true hypothesis of this study, namely that the predictors of interethnic childbirth will have different relationships by marital status and the ethnicity of the mother, is not supported. Instead, marital status has little effect on the demographic patterns of interethnic childbirth.

¹⁶ In regard to the interethnic relationship literature, the finding that interethnic relationships are more common in cohabiting relationships (Blackwell and Lichter 2000; Lichter and Qian 2004; Qian and Lichter 2007) may be relevant for non Hispanic women, as opposed to Hispanic women.

¹⁷ Due to the large sample size, only variables that were significant at the .001 level were considered to be “significant.”

The effects of a majority of the independent variables on interethnic childbirth are in opposite directions for Hispanic versus Non-Hispanic women. Hispanic women are more likely to have interethnic children if they are older, non white, highly educated, and have had fewer previous children (relative to other Hispanic women). On the other hand, non Hispanic women are more likely to have interethnic children if they are younger, less-educated, white, and have had more previous children (relative to other Non Hispanic women). Hispanic women that have interethnic children are thus more likely to have higher social status and perhaps fit an image of a more “professional” woman. On the other hand, Non Hispanic women who have children with Hispanic men may be more common for young white women of “lower” social status.¹⁸ Supporting this idea of “professionalism,” supplemental analyses indicate that Hispanic women are more likely to have interethnic children when married, whereas Non Hispanic women are less likely to do so (see Appendix Table).

Reflecting marital-market dynamics, the percentage of persons in a state that are Hispanic increases the probability of having children with a Hispanic person. The state number of births, a proxy for the state population size, decreases the probability of having a child with a Hispanic person for mothers of both ethnicities.

Only two variables have the same relationships across all groups and consistently predict interethnic births. Foreign-born women (particularly Hispanic women) are consistently less likely to have interethnic children. Year also has the same, albeit very small, directional

¹⁸ Although this research does not directly address interethnic relationships (but, rather, interethnic childbirth), these results may be useful in reconciling seemingly contradictory findings in previous literature on education’s relation to interethnic relationships (negative, a la Hohmann-Marriott and Amato 2008, or positive, a la Qian and Lichter 2007). These contradictory findings are reconciled when controlling for ethnicity, as education has opposite effects for white versus Hispanic women in regard to having interethnic children. Contrary to earlier research on interethnic relationships (Batson et al. 2006; Rosenfeld 2005; Qian and Lichter 2007), this research finds that education’s effect on interethnic childbirth does not change direction by marital status.

effect across samples, demonstrating that interethnic partnerships are demonstrated to be more common over time.

Conceptually, the statistical models provide some clues regarding the demographic predictors of parental partnerships. Mothers who are Hispanic are 54.28 times more likely to partner with a Hispanic father (as seen in the Appendix Table), indicating that the United States is far from the freely-blending melting pot that pairs persons without regard to ethnicity. Race similarly has strong effects on the model, indicating that white women are more likely to pair with a Hispanic father. However, the other demographic variables that predict pairing with a Hispanic father have similar effects for Hispanic mothers as well as for Non Hispanic mothers. Put another way, because education (for example) has a negative effect for having interethnic children for Hispanic mothers yet a positive for Non Hispanic mothers, education therefore has a negative effect with pairing with Hispanic fathers in general for mothers of both ethnicities. What this means is that, to some degree, one does not need to think in terms of “Hispanic mothers having interethnic children” nor of “Non Hispanic mothers having interethnic children.” Instead, the data are saying that certain demographic qualities influence persons to partner with (Non)Hispanic fathers, regardless of that person’s own ethnicity.

DISCUSSION: WHAT THE STATISTICS TELL US ABOUT INTERETHNIC PARENTAGE

This research highlighted several demographic factors that contribute to interethnic parentage. The results can be summarized by making general observations and “sketches” of the populations that are more likely to bear interethnic children. In general, Hispanic women having interethnic children tended to have more “professional” attributes in being older, highly educated, and having had fewer previous children. These women also tend to be non white and married when having interethnic children. Non Hispanic mothers, on the other hand, tended to

reflect “less professional” attributes, more likely to be younger, less-educated, and having more previous children. Non Hispanic mothers also tend to be white and non-married when having interethnic children. However, these relationships are conditional on the “average” characteristics of (Non)Hispanic women, preventing from making cross-comparisons for women of different ethnicities without accounting for averages within groups. Future research can explore this possibility using log-linear models.

It is clear that interethnic parentage is the exception, as a mother’s Hispanic ethnicity improves the odds of pairing with a Hispanic father by a factor of 54. However, aside from a woman’s ethnicity, it is also demonstrated that one’s social environment, as in neighborhood ethnic composition and educational networks, influence the selection of reproductive partners and partially neutralize the effect that ethnicities have on mating patterns. In terms of previous literature, these results indicate that the “marital market” that one is exposed to influences one’s reproductive partners. But perhaps more generally, this suggests that “colorblindness” simply results from sharing one’s life with different people. Going to college and living in a city with a different ethnic make-up neutralizes the boundaries that divide people in regard to one’s ethnicity. For example, instead of viewing one’s self as an “Asian” person, one may instead identify as a college-graduate (who happens to be “Asian”). Instead of viewing one’s self as “white,” one may instead identify as an “Angelino/a.” In these ways, demographic characteristics and social context are factors that contribute to the “melting pot” effect.

The data also suggest that there are generational and time-effects that neutralize ethnic mating boundaries. Generationally, persons born in the United States are also more likely to have interethnic children. Additionally, interethnic parentage is becoming more common with

time (controlling for the other variables), meaning that a melting pot is slowly mixing persons together.

CONCLUSION

The purpose of this research was to identify predictors for interethnic childbirth by using variables that have been used to predict Hispanic family formation and interethnic relationships. Despite the conceptual inextricability between interethnic relationships and interethnic childbirth, the models presented here have revealed a number of differences between the two literatures.

While studies of mixed-race relationships in general suggest that interethnic cohabitation is more common in cohabitations rather than in marriages (Blackwell and Lichter 2000; Lichter and Qian 2004; Qian and Lichter 2007), interethnic childbirth indicates that interethnic children born out of wedlock is only more common for Non Hispanic women, whereas it is less common for Hispanic women.

Previous research has also indicated that interethnic marriages tend to be more common as education increases, although this is not the case for cohabiting relationships (Batson et al. 2006; Rosenfeld 2005; Qian and Lichter 2007). For interethnic childbirth, however, this study finds that the relationship between education and interethnic childbirth depends on the ethnicity of the mother: education seems to have a positive relationship on interethnic childbirth for Hispanic women, although it has a negative relationship for Non Hispanic women. This may, however, be a reflection of the “spousal trades” phenomenon, in which the high education in one spouse is associated with higher color status in another (Blackwell and Lichter 2000). Further research would be necessary to determine the roles that the educational attainment of the father (unavailable in this data set) and the races of the mothers vis-à-vis the

fathers have on interethnic childbirth. Unlike Blackwell and Lichter's (2000) finding for interethnic relationships, the role of education on interethnic parentage does not vary by marital status.

While this research finds that white Hispanic mothers are less likely to have interethnic children, this may be due to a data artifact resulting from the fact that NCHS racial categories lack an "other" category as found in the Census.¹⁹ This finding contrasts with previous studies of interracial relationships that find that such relationships are more common for white Hispanics (King and Bratter 2007; Qian and Cobas 2004).

In common with the literature on interethnic relationships, native-born Hispanics are more likely to have interethnic children (similar to Qian and Cobas 2004; Qian and Lichter 2007). Also similar are marital market effects, which are a function of interethnic residential exposure (Lichter et al. 2007) and an ethnic group's population size (Kalmijin and Tubergen 2010).

For the purposes of the Demographic Analysis of Hispanics, the interethnic population and the way in which interethnic persons identify will have a large influence on the count of the Hispanic population. An understanding of the factors that produce interethnic children will be useful in making informed decisions regarding the imputation of missing ethnicities, as 14% of birth certificates lack data on the father's ethnicity. On this note, Chapter 4 will further the research on DA by using these models (and other methods) to impute missing data on birth certificates.

¹⁹ Previous research, however, indicates that white Hispanics tend to be more likely to pair with non Hispanics (i.e. Qian and Cobas 2004). This differing result is due to the racial categories available on birth certificates, which do not include an "other" option as opposed to the Census. Because of this, all Hispanics who are neither Asian nor black are likely coded as "white." This produces contrasting results across data sets because Hispanics that identify as "some other race" are more likely to marry fellow Hispanics, while Hispanic whites are less likely to do so (Tafoya 2004). The lack of the "other" option in birth certificates in this case leads to higher proportions of "white" Hispanics (who would otherwise be classified as "other") to pair with fellow Hispanics.

In summary, this research adds complexity to our understanding of interethnic relationships in general. While some variables have similar relationships for interethnic relationships and interethnic parentage, there are also many differences that are found. Future research can use other data sets with additional variables, including more detail on father characteristics, in furthering our knowledge of interethnic childbirth.

While the United States is far from a melting pot that features ethnically inter-breeding persons en masse, there are portions of the populations who do produce interethnic children. As years pass and more persons melt together within this pot, ethnic relations between “groups” will surely change. This study highlighted some of the factors that contribute to this melting pot by identifying demographic characteristics of persons and communities that fuel this intermixing. Future studies of interethnic breeding and the resultant ways that the concept of ethnicity (let alone the categories used to distinguish ethnicities) changes will be important for persons wishing to sketch a portrait of the ethnic landscape of the United States in the coming years.

REFERENCES

- Batson, Christie D., Zhenchao Qian, and Daniel T. Lichter. 2006. "Interracial and Intraracial Patterns of Mate Selection Among America's Diverse Black Populations." *Journal of Marriage and Family* 68:658–72.
- Blackwell, Debra L. and Daniel T. Lichter. 2000. "Mate Selection Among Married and Cohabiting Couples." *Journal of Family Issues* 21:275–302.
- Brown, Susan L., Jennifer Van Hook, and Jennifer E. Glick. 2008. "Generational Differences in Cohabitation and Marriage in the US." *Population Research and Policy Review* 27:531–550.
- Bumpass, Larry and Hsien-Hen Lu. 2000. "Trends in Cohabitation and Implications for Children's Family Contexts in the United States." *Population Studies* 54:29–41.
- Carlson, Marcia J., and Frank Furstenberg. 2006. "The Prevalence and Correlates of Multipartnered Fertility Among Urban US Parents." *Journal of Marriage and Family* 68:718–732.
- Casper, Lynne and Bianchi, Suzanne M. 2002. *Continuity and Change in the American Family*. Thousand Oaks, CA: Sage Publications.
- Castro, Martin T. 2002. "Consensual Unions in Latin America: Persistence of a Dual Nuptiality System." *Journal of Comparative Family Studies* 33:35–55.
- Eggebeen, David. 2005. "Cohabitation and Exchanges of Support." *Social Forces* 83:1097–1110.
- Glick, Jennifer, Stacey, D. Ruf, Michael J. White, and Frances Goldscheider. 2006. "Educational Engagement and Early Family Formation: Differences by Ethnicity and Generation." *Social Forces* 84:1391–1415.
- Hohmann-Marriott, Bryndl E., and Paul Amato. 2008. "Relationship Quality in Interethnic Marriages and Cohabitations." *Social Forces* 87:825–855.
- Joyner, Kara and Grace Kao. 2005. "Interracial Relationships and the Transition to Adulthood." *American Sociological Review* 70:563–581.
- Kalmijn, Matthijs and Frank Van Tubergen. 2010. "A Comparative Perspective on Intermarriage: Explaining Differences among National-Origin Groups in the United States." *Demography* 47:459–479.
- King, Rosalind Berkowitz and Jenifer L. Bratter. 2007. "A Path Toward Interracial Marriage: Women's First Partners and Husbands across Racial Lines." *The Sociological Quarterly* 48:343–369.
- Landale, Nancy, R. Salvador Oropesa, and Cristina Bradatan. 2006. "Hispanic Families in the United States: Family Structure and Process in an Era of Family Change." Pp. 138–178 in

Hispanics and the Future of America, edited by Marta Tienda and Faith Mitchell. Washington, DC: The National Academies Press.

Lee, Sharon and Barry Edmonston. 2005. "New Marriages, New Families: U.S. Racial and Hispanic Inter-marriage." *Population Bulletin* 60(2).

Lichter, Daniel, J. Brian Brown, Zhenchao Qian, and Julie H. Carmalt. 2007. "Marital Assimilation Among Hispanics: Evidence of Declining Cultural and Economic Incorporation?" *Social Science Quarterly* 88:745-765.

Lichter, Daniel T. and Zhenchao Qian. 2004. *Marriage and Family in a Multiracial Society*. New York and Washington, DC: Russell Sage Foundation and Population Reference Bureau.

Manning, Wendy, Pamela Smock, and Debarun Majumdar. 2004. "The Relative Stability of Cohabiting and Marital Unions for Children." *Population Research and Policy Review* 23:135–159.

Manning, Wendy D., and Susan L. Brown. 2006. "Children's Economic Well-Being in Married and Cohabiting Parent Families." *Journal of Marriage and Family* 68:345–362.

Musick, Kelly. 2002. "Planned and Unplanned Childbearing among Unmarried Women." *Journal of Marriage and Family* 64:915-929.

Oropesa R.S, Bridget Gorman, Linda Waite, Christine Bachrach, Michelle Hindin, Elizabeth Thompson, and Arland Thornton. 2000. "Ethnicity, Immigration, and Beliefs about Marriage as a 'Tie that Binds.'" Pp. 188-211 in *The Ties That Bind: Perspectives on Marriage and Cohabitation*, edited by L.J. Waite. New York, NY: Aldine de Gruyter.

Oropesa R.S. and Nancy Landale. 2004. "The Future of Marriage and Hispanics." *Journal of Marriage and Families* 66:901–20.

Osborne, Cynthia, Wendy Manning, and Pamela Smock. 2007. "Married and Cohabiting Parents' Relationship Stability: A Focus on Race and Ethnicity." *Journal of Marriage and Family* 69:1345–1366

Qian, Zhenchao and Daniel T. Lichter. 2007. "Social Boundaries and Marital Assimilation: Interpreting Trends in Racial and Ethnic Inter-marriage." *American Sociological Review* 72:68

Qian, Zhenchao, Daniel T. Lichter, Leanna M. Mellott. 2005. "Out-of-Wedlock Childbearing, Marital Prospects and Mate Selection" *Social Forces* 84: 473-491.

Qian, Zhenchao and Jose Cobas. 2004. "Latinos' Mate Selection: National Origin, Racial, and Nativity Differences." *Social Science Research* 33:225–247.

Rosenfeld, Michael. 2005. "A Critique of Exchange Theory in Mate Selection." *American Journal of Sociology* 110:1284–1325.

Royston, Patrick. 2005. "Multiple Imputation of Missing Values: Update of ICE." *The Stata Journal* 5:527-536.

Stockard, Jean, Jo Gray, Robert O'Brien, and Joe Stone. 2009. "Race Differences in Cohort Effects on Non-Marital Fertility in the United States." *Social Forces* 87:1449-1479.

U.S. Census Bureau Population Division. 2010. *The Development and Sensitivity Analysis of the 2010 Demographic Analysis Estimates*. Census Bureau, Washington, D.C.

Wildsmith, Elizabeth and Kelly Raley. 2006. "Race-Ethnic Differences in Nonmarital Fertility: A Focus on Mexican American Women." *Journal of Marriage and Family* 68:491-508.

TABLES

Table 4.1: Percent of Births that are
Interethnic, by Ethnicity and Marital
Status of the Mother, 1990 – 2000

	Married	Not Married
Hispanic	15.30	11.82
Non Hispanic	2.56	5.43

Table 4.2: Interethnic Childbirth by Characteristics of the Mother: Logistic Regression Odds Ratios

	Model 1	Model 2	Model 3	Model 4
	Mother Hispanic Married	Mother Hispanic Not Married	Mother Non Hispanic Married	Mother Non Hispanic Not Married
Race (omitted: white)				
Black	1.66	2.73	0.50	0.21
neither black nor white	3.78	4.45	0.96	1.11
Education (omitted: high school)				
no high school	0.32	0.53	1.26	1.19
some college	1.64	1.41	0.91	0.98^
bachelor's and above	2.30	1.89	0.64	0.76
Other Characteristics of Mother				
foreign born	0.25	0.18	0.55	0.76
number of live births	0.82	0.88	1.08	1.09
Age	1.06	1.03	0.95	0.96
Social Context				
number of births in state	1.01	1.00^	0.99	0.99
percent Hispanic in state	0.86	0.86	1.29	1.27
Year	1.00	1.00	1.00	1.00
Pseudo R ²	0.21	0.17	0.09	0.12

Number of births in state in units of 10,000.

Percent Hispanic in state in units of 5%

^ not significant at the .001 level

APPENDIX TABLE

Table 4.3: Alternate Models Predicting the Ethnicities of Fathers. Logistic Regression Odds Ratios

	Model 5	Model 6	Model 7
Mother's Ethnicity	Hispanic	Non Hispanic	Both
Father's Ethnicity (DV)	Non Hispanic	Hispanic	Hispanic
New Variables			
Mother Hispanic	-	-	54.28
Married	1.02	0.53	0.68
Race (omitted: white)			
Black	2.07	0.30	0.34
neither black nor white	4.10	1.01^	0.36
Education (omitted: high school)			
no high school	0.40	1.21	1.80
some college	1.61	0.92	0.76
bachelor's and above	2.38	0.64	0.53
Other Characteristics of Mother			
foreign born	0.22	0.59	2.74
number of live births	0.83	1.09	1.12
Age	1.05	0.96	0.96
Social Context			
number of births in state	1.00	0.99	0.99
percent Hispanic in state	0.86	1.29	1.24
Year	1.01	1.00	1.00^
Pseudo R ²	0.200	0.106	0.647

Number of births in state in units of 10,000.

Percent Hispanic in state in units of 5%

^ not significant at the .001 level

CHAPTER 5
IMPUTATION PROCEDURES FOR BIRTH RECORDS:
A CONTINUATION OF CHAPTER 4

The bane of a statistician's existence is missing data. Although birth certificate data is perhaps among the most complete datasets that exist, it is not without its problems. For the purposes of this study, the main flaw in birth certificate data is that it has missing data for fathers' ethnicities for 14% of cases. As missing father ethnicity is treated as Non Hispanic in the National Center for Health Statistics' (NCHS) tabulations, the resultant error that is created can be quite high.

This chapter is the most purely methodological chapter in this dissertation, leaving the realm of Sociology altogether. The Sociological explorations that lie at the foundation of this chapter are found in Chapter 4, which deals with regression models that predict interethnic childbirth. This chapter is dedicated to follow up on Chapter 4, using the findings of Chapter 4 to impute missing data on parental ethnicities. This research approaches imputation from the perspective of a disseminator of descriptive statistics of a population. Specifically, this research will explore imputation options that will best apply to a governmental body that is called upon to provide counts of a population.

The main purpose of this chapter is to test the accuracy of various imputation methods that can be used to impute missing Hispanic ethnicities on birth certificate data. Four imputation procedures are tested, including two that are based on regression methods, a previous observation methodology, and the method that is currently used by the U.S. Census Bureau Population Division that is based on proportional distributions of observed cases. Error patterns of the imputation procedures will be identified for each of the four procedures, and suggestions

for future research will be made in the first Conclusion. This research will demonstrate that imputation procedures that are based on regression models, particularly one that relies on posterior predictive distributions of values, perform remarkably well. Additionally, a simpler model used by the Census Bureau, which essentially imputes by using conditional averages, also performs well. However, imputations based on means produce greater errors by demographic detail (i.e. by age), as the Census method creates higher estimates of Hispanic children of older ages. A second conclusion will use the imputation procedures tested in this analysis in order to produce counts of Hispanic native-born children who reside in the United States.

COMMON IMPUTATION PROCEDURES

Missing data leads to a number of problems when conducting analyses. Barnard and Meng (1999) identify three fundamental problems that accompany missing data: loss of power or efficiency for statistical tests, complications that arise from irregularities in the data patterns, and bias that results from the selectivity of cases that feature missing data. For Demographic Analysis (DA), missing data affects the counts of a target population, thereby compromising the very objective of DA and the validity of the data used.

The methods that are used to deal with missing data are numerous, although not all of these methods effectively deal with the three problems outlined above. The simplest way to deal with missing data is to simply perform a list-wise deletion of cases with missing data, which is an acceptable strategy if the sample size is large and values are missing at random. However, when a list-wise deletion of cases leads to a study founded on untenable assumptions and faulty data, some form of data imputation for missing cases would be advisable. One must be careful in selecting the appropriate methodology for imputing data as the methodologies can

dramatically alter the findings of a study, at worst making the results of any subsequent analyses worse-off than simply performing a list-wise deletion of cases with missing data.

A popular and hassle-free way to impute data is to replace missing values with population means (and/or medians). This can be done in a number of ways, including by using conditional means (within groups) or row/column means, depending on how the data are structured. While an easy method to utilize, using means may squash variation when inferential statistics are called for and may work under the false assumption that missing cases are not statistically different from complete cases. However, when the goal is simply to produce descriptive statistics (i.e. counts), imputing with means is an attractive procedure since the end goal is not to perform inferential statistics. Additionally, when carefully conditioned, the use of means can account for certain sources of variation. The Demographic Analysis program at the U.S. Bureau of the Census, for example, imputes missing values on missing parental ethnicities on birth certificates using a modified “mean” approach that is conditional on the state, decade, and ethnicity of the partner (when available, U.S. Census Bureau 2010).

When missing data is not missing at random and follows a general pattern, imputing on the basis of other observed values can be performed with success. Engels and Diehr (2003) utilize a longitudinal dataset in which missing values are commonly found following an initial missing value. These authors take values that are commonly missing, but actually observed, and impute the values as missing to test the success rate of varying imputation measures. Under this scenario, imputing on the average of the following and final observed values is demonstrated to have the best performance. When these data are unavailable, the previous observed value or previous case mean/median produces the next most satisfactory results. Finally, when no longitudinal data are available, these authors find that a class of imputation measures termed

“baseline” imputations – running the gamut from regression predicted values, hot-deck, or class means/medians – are suggested to be used. Before taking these findings to heart, however, one must consider that these findings are gathered from a specific type of (longitudinal health) data with an equally specific type of missing data pattern.

Another popular type of imputation method is known as the “hot deck” procedure, which can assume the form of a “nearest-neighbor” imputation methodology. While a hot deck can take many forms, at its heart the hot deck imputation procedure relies on the presence of complete, clean, and complete cases of data that are similar to the cases that need imputing (Fellegi and Holt 1976). These complete cases are treated as “donor” cases to the cases in need of imputation wherein missing values are drawn from the corresponding donors that match on a number of observed characteristics. This procedure has been used extensively by the Census Bureau for missing ethnicities on the Decennial Census, although utilizing such procedures may have led to the over-estimation of Hispanic persons for the imputed cases (Cresce, Schmidley, and Ramirez 2002).

A final important imputation procedure to introduce relies on covariates in imputing values on the basis of the values of the variable’s covariates in a statistical model. Using Bayesian probability statistics, Rubin (1987) proposes fitting a model that predicts values on the variable in need of imputing. On the basis of the values of the covariates for the cases in need of imputation, a distribution of possible values can be drawn from. Random draws from the posterior predictive distribution can then be used to impute for missing values. Repeating this procedure multiple times in “multiple imputation,” random variation in imputation draws will be evened out over the imputation waves, producing small yet reliable standard errors that appropriately fit the data (Schenker et al. 2006). This type of multiple imputation is shown to be

quite flexible and reliable insofar as the models used are robust (see Barnard and Meng 1999 for some examples of its usage).

The National Center for Health Statistics uses a number of simple methods for imputing missing data on birth certificates. For missing mother's age, for example, the imputed age is drawn from the age of the previous record with a mother of the same race and total amount of previous births. Missing marital status for women get imputed a "married" status, which is the most common marital status for mothers. When producing public reports on measures such as fertility rates, the child is assigned the race and ethnicity of the mother. The race of the child is assigned the race of the father if the race of the mother is missing, and the race of the mother of the previous record if both the mother's and father's races are missing. If a mother's ethnicity is missing, the child is simply assumed to be Non Hispanic (National Center for Health Statistics 2000).

The aforementioned imputation procedures are by no means exhaustive, although they represent a good sample of the procedures used in population and health research. The National Center for Education Statistics (2002, Appendix B) provides examples of other imputation procedures including historical imputation, deductive imputations, random imputation, flexible matching imputation, ratio imputation, EM algorithm imputation, distance function matching, and composite methods. While many of these imputation methods are used in population and health research, the majority either will not work for the purposes of this research given the structure of the data, and/or the imputation measures previously described offer statistical improvements. However, institutions with access to the complete, 100% NCHS data may be able to implement some of these imputation procedures (e.g. the historical

imputation method – as well as more sophisticated “nearest neighbor hot deck” methodologies as currently used by the Census Bureau).

This study aims to offer alternatives for the NCHS in assigning the ethnicity of a child’s parents, although the implications of this research extend beyond simply ethnicity. The goal is to test the successes of the different imputation procedures for missing ethnicities on birth certificate data and to provide suggestions for future publications of summary statistics. As mentioned previously, the goal is not to impute data *per se*, but rather to provide total expected counts of children by parental ethnic arrangements. As will be explained below, the methods that will be tested include a “previous observation” method, a “modified means” method, a “probabilities sum” method, and a “Bayesian model-based imputation” method. Using these imputation procedures, this research also aims to predict counts of Hispanic children by predicting the number of children that would be identified as Hispanic on the Census on the basis of the ethnicities of one’s parents.

DATA

All birth data are taken from the National Center for Health Statistics micro data for vital statistics birth data. As performed in Chapter 3, the demographic characteristics of the mother were used as the independent variables. Complete birth records with no missing data for children who would be between the ages of 0 and 10 on April 1, 2000 were sampled. Mothers who stated residence outside of the United States or in states for which Hispanic ethnicity was not collected were not included in this analysis. Independent variables age, race, and foreign-born were drawn directly from the data. Year of birth is coded into Census years (beginning on April 1). The state number of births sums the births by Public Use Microdata Area (PUMA) and the result is divided by 10,000 and rounded to the nearest integer. The percent births to

Hispanic women is also grouped by PUMA by year, grouped in units of five, and rounded to the nearest integer. Education is coded to no high school, high school (which is omitted), some college, and bachelor's and above. Mother's race is (normally) coded to white (omitted), black, and other. Number of previous live births is truncated to 10 at its upper limit.

Only complete birth records (without missing data on the variables used) were included, which totaled 36,045,922. Missing data was imputed into this dataset using a hot-deck method, which matched complete cases with incomplete cases by marital status, number of births in the PUMA, year, and the age of the mother (all of these variables are significant on predicting missing values). The father's ethnicity and mother's race, education, and nativity status were then imputed as missing by using a hot deck "donor" case with missing values. Using this procedure, missing values were distributed based on the patterns and correlates of missing cases found in the data. The procedure produced missing values for 8.2% of fathers' ethnicities, .5% of mothers' ethnicities, 2.1% of mothers' education, and .2% of mothers' nativity status.

Estimating the number of Hispanic children that would be identified on the Census will require assumptions regarding the proportions at which children with varying parental ethnic arrangements identify as "Hispanic." It will be assumed that all children with both Hispanic mothers and fathers will be identified as "Hispanic." It will similarly be assumed that all children with both Non Hispanic mothers and fathers will be identified as Non Hispanic. For children of mixed ethnicity, the rates of imputation will be drawn from the U.S. Census Bureau (2010). In this estimate, 61.4% of children will be imputed as Hispanic when the mother is Hispanic (but the father is not), while 69.8% of children will be imputed as Hispanic when the father is Hispanic (but the mother is not).

METHODS

Four methods of imputation were used to fill in the missing data for the fathers' ethnicity. After the data are imputed, percentages and counts of Hispanic children by the ethnicities of the parents will be calculated, and these figures will be compared with the corresponding figures calculated by using the observed values. Net percentages and counts of "incorrect" imputations were calculated to assess the effectiveness of each imputation procedure.

The *previous observation* method will work under the assumption that the observations directly preceding observations with missing cases aggregately represent the population as a whole – or, better yet, that such observations represent the population with missing values. As the name implies, cases with missing ethnicities will be imputed by using the ethnicity of the father of the previous observation for mothers of the same ethnicity. Cases are sorted by the ethnicity of the mother, year, state, and case number.

For the Demographic Analysis program of the Census Bureau, cases with missing parental ethnicities are imputed on the basis of the proportions of parental ethnic arrangements found for the state. This method will be termed the *Census modified means* (or simply the "Census") method. For example, fathers of unknown ethnicities who are paired with Hispanic mothers are assigned ethnicities based on the proportions of Hispanic fathers who are paired with Hispanic mothers for the state, over the decade of the birth record. Where the mother's ethnicity is missing, the same process is reversed and applied to the mother. Where both parents' ethnicities are missing, overall proportions of parental ethnic arrangements for the state and decade are used as substitutes. As this method effectively imputes ethnicities on the basis of the "mean rate" of occurrence, it basically imputes values through the use of means. These means, however, are conditioned by state, decade, and the ethnicity of the mother. Like

the previous observation method, this method rests on the assumption that the proportions of parental ethnic arrangements for cases with missing data are the same as the proportions of parental ethnic arrangements for complete cases.

[TABLE 5.1]

For the regression methods of imputation, logistic regression was used to predict the father's ethnicity. The models used for the imputations are displayed on Table 5.1, which are based on the Appendix Table in Chapter 4. Table 5.1 also displays the relationships that the independent variables have on the dependent variable when using the original dataset. For Non Hispanic mothers, race was coded to white/non-white because "other race" was insignificant in the original models. The regression methods of imputation use two models: one for Hispanic mothers, and the other for Non Hispanic mothers.

As the goal of this research is simply to provide aggregate descriptive statistics, assigning probabilities of being Hispanic for individual cases and summing these probabilities can estimate the total counts of children that have specific ethnic arrangements. Determining such probabilities can be derived from a logistic regression model that predicts odds (which can be converted into probabilities) of having a Hispanic father, based on the values of the covariates in the model outlined in Table 5.1. By summing the predicted probabilities by the ethnicity of the parental partner (that is, the sum of the $\hat{y}$ values), one can estimate the number of children that were born by the ethnicities of their parents. This method will be called the *probability* method of imputation. Due to the limitations of the regression method when using incomplete data, the *probability* method is unable to impute for missing father ethnicities for 771,726 cases (representing 26% of missing cases for father ethnicities). For these cases, missing values were imputed by using the matched imputation value in the method described below.

Lastly, imputations can be performed by using a regression model to provide a distribution of possible scores that one would expect to find for each observation. This distribution would hence be conditioned by the covariates found for the respective observation. Random draws from the posterior predictive distribution would then be performed and imputed into the dataset. Each random draw will condition subsequent draws, ensuring that the end set of imputations will be based on a distribution of draws that adequately reflects the standard errors that one would expect to find in the data set. Normally, this procedure would be repeated multiple times (hence this method's title of "multiple imputation") so as not to bias the results of inferential statistical procedures on the basis of one set of draws. However, as the goal of this research is simply to estimate aggregate population totals, this research will only perform a single-imputation round using such a method. STATA's Multiple Imputation by Chained Equations (ICE) procedure was used to perform this imputation procedure (see Royston 2005). All variables in the model have missing values imputed by modeling them on the other variables in the main model (unless otherwise specified), beginning with the variable that has the fewest number of missing cases. The models used for STATA's ICE procedure are also displayed on Table 5.1. This method will be called the *ICE* method in this text.

RESULTS OF IMPUTATION PROCEDURES

[TABLE 5.2]

Table 5.2 displays the results of the imputation procedures for all of the fathers' ethnicities that were imputed. The first two columns display the true values against which the imputation measures would be evaluated. The first two rows compare the percentages of children by mother and father ethnicities (in row percentages) and display similar results for all imputation

procedures, only exceeding a 1% difference in cell percentages for the Census procedure for Hispanic mothers.

The “net percent error” rows (rows 2 and 3) display the direction and strength of the imputation bias in terms of percentage deviations for each cell. That is, for each cell, the net percent error rows calculate the difference between the counts in the imputation columns and the observed columns, which are then divided by the observed counts and multiplied by 100. Rows 2 and 3 show that the regression models (ICE and probability methods) perform stupendously, never having deviations from the observed values that exceed .08%. The previous observation method also performs well, although the errors across all categories are higher than are found for the regression methods. The largest deviation from the observed values is found for the Non Hispanic mother, Hispanic father category, as using a previous observation method underestimates this parental arrangement by 4.5%. The Census method performs the most poorly, underestimating interethnic children by 7.4% (when the mother is Hispanic) and 4.1% (when the father is Hispanic). Intraethnic children are predicted at more accurate rates and are within 1.3% from the observed values.

The relative poor performance for the previous observation and Census methods of imputation are due to the patterns of missing data. These methods work under the assumption that data are missing at random, as these methods produce imputations based on all the other cases using a blanketed approach. However, these procedures underestimate the counts of interethnic children, suggesting that missing data are more common for interethnic couples.

[TABLE 5.3]

Table 5.3 uses the counts of children by parental ethnicities that were produced by each imputation rule and produces theoretical counts of Hispanic children that would result from the

respective estimate. The theoretical counts of Hispanic children are calculated by multiplying the counts for each parental ethnic arrangement by the assumed proportion with which children are identified as “Hispanic” on the Census.

Again, the regression procedures produced surprisingly accurate results that are estimated at .005% (ICE) or .004% (probability) higher than the observed counts. These results even outperform the results in Table 5.2, partially due to the “balancing” of overestimates with underestimates. The balancing of over- and under-estimates is much more dramatic for the other two methods of imputation. For the previous value method, error of counts fall to -.48%. Perhaps the “poorest” performer in Table 5.2, the Census method ends up performing very well in predicting the counts of Hispanic children, underestimating the count by .02%.

[TABLE 5.4]

The first set of columns on Table 5.4 break down the count errors of Hispanic children by age based on the imputation procedure used. Viewing the Census method column, one sees further evidence that the “balancing” effect aided in this method’s accuracy in predicting total Hispanic counts. Specifically, the Census method underestimates the count of Hispanic children of younger ages, and overestimates such counts for children of older ages. This result is likely due to the error produced by using a decade-based average for interethnic childbirth, which does not consider the changes of the rates of interethnic childbirth. Since interethnic childbirth is more common over time, a fixed decade-based average would increasingly underestimate a moving average that increases every year, as well as conversely overestimating the smaller rates for earlier years. For the middle age (age 5), the Census method nearly predicts the number of Hispanic children perfectly. Similarly, the previous observation method increasingly underestimates the number of Hispanic children for more recent years, although this method

even underestimates Hispanic children at the earliest time point. The previous observation method misses Hispanic father ethnicities at increasingly higher rates, underestimating Hispanic children for the most recent time point at rates over three times higher than the underestimate at the earliest time point.

The next set of columns on Table 5.4 display the absolute percent error of Hispanic children by age. Using this criterion of accuracy, the ICE procedure outperforms all of the other imputation procedures, having less than half of the total absolute error than the next most accurate procedure. The Census procedure is the next most accurate, netting a total error of a touch over 1%. Unsurprisingly the previous observation method produces the most total error. Surprisingly, the otherwise most “successful” procedure under the other measures produced relatively high rates of error by age. Like the Census method, the probability method benefitted the most from the balancing of overestimates and underestimates by age. However, the volatile (and occasionally dramatic) changes in the under- and over-estimates by age make one question whether the probability method simply “got lucky” in its final tally of children.

CONCLUSION, PART 1: WHICH IMPUTATION PROCEDURE TO USE

This research suggests that the performance of the various imputation measures ultimately depends on the researcher’s usage of the data. Each procedure has varying strengths and weaknesses, depending on the context for which the imputations are used. However, some conclusions can be made, as follows.

1. Across all measures, the ICE procedure – which rests on regression models and makes imputations from random draws of the posterior predictive probability based on the values of the covariates – performs very well. The ICE procedure also comes with the added benefit of imputing values directly into the dataset, which comes in handy when

inferential statistics are called for. The difficulty with using this procedure rests on the massive computing power needed to perform this task (which is the reason why the socio-geographic variables were truncated to into integers).

2. In terms of producing total counts of children by parental ethnic arrangements and in total estimates of Hispanic children, the sum of predictive probabilities outperforms the ICE procedure. However, a closer look at the erratic and dramatic age-distributions of the errors calls into question the utility of this method. Additionally, this method's inability to produce estimates for cases with missing data makes this method only truly reasonable when the errors follow a monotonic pattern.
3. The Census modified means procedure produces very good estimates of totals of Hispanic children, although it tends to underestimate the counts of interethnic children. Additionally, the simplicity of the method makes it very appealing and readily replicable. The holes in this method become evident by age, as its underestimate of interethnic children produce underestimates of Hispanic children in more recent years. However, the linear error by age, and the spot-on accuracy of the estimates in the middle age category suggest that this method can produce quite favorable estimates if slight age-adjusted modifications can be made.
4. Imputations based on previous observations are not recommended. Revisions to this suggestion may be made for data with randomly missing values, and future research can determine the potential utility of this imputation procedure under data that are missing at random (or highly autocorrelated data).

In sum, this research concludes with the recommendation that the ICE procedure be used when substantial computing processing power is readily available. As evidenced in this

study, even when data are not missing at random, descriptive statistics can still be computed at high accuracy using ICE. Replicating the missing data patterns found in birth certificate data, this research demonstrates that the ICE procedure continues to perform well, never exceeding .08% error using any of the previous criteria (with the exception of the total absolute percent error for eleven age groups). Another benefit that ICE has over the other procedures is the fact that ICE directly imputes values into the micro-data. While the other procedures only produce counts (and thus do not necessarily qualify as being “imputation” procedures *per se*), ICE generates values for missing data that are informed by the covariates in the model. Thus, the utility of ICE goes beyond producing estimate counts, but can also be used for other purposes such as cleaning datasets.

This research also highlights the favorable performance of the Census procedure (which uses a “modified means” imputation procedure), although such a procedure underestimates the number of interethnic children. Displaying much potential promise when examining error rates by age, future research can continue to refine this method and make it very robust. On the other hand, it can be argued that the Census procedure simply “got lucky” and, despite the errors by age and parental ethnic arrangement, managed to balance underestimates with overestimates to closely approximate observed values.

A cautionary note may be made for this research, as the goal is simply to produce descriptive statistics (e.g. counts) based on large micro-data sets. The performance of these procedures may have different results when the uses of the data are different. On a positive note, imputations based on the posterior predictive probability of cases (based on covariates) have been demonstrated to perform rather well when using multiple imputation procedures (as

opposed to single-imputation as done here, see Donders et al. 2006; Schaefer 1999; Schenker et al. 2006).

CONCLUSION, PART 2: ESTIMATED COUNTS OF HISPANIC CHILDREN BORN IN THE UNITED STATES

Having explored the possible options for imputing missing parental ethnicities on birth certificates and tested their attendant error structures, one can apply these methodologies to the entire ten-year dataset in order to produce a complete count of Hispanic children. This second Conclusion does just this, using the probability, ICE, and Census methods to account for missing data on parental ethnicities. Due to the poor performance of the previous observation method, this Conclusion will not replicate this procedure. As in the previous analyses, the probability procedure was unable to predict probabilities of having a Hispanic father for all cases due to cases with larger amounts of missing data. In these cases, results from the ICE procedure were used in lieu of the probability procedure.

Due to the potential conflation of the race and ethnicity concepts, ethnicities were coded to Non Hispanic if the parent's ethnicity is missing but a race is stated. This is done under the assumption that these cases "racialize" the Hispanic term, identifying persons with a race and thereby making the (non)Hispanic identity redundant (see Chapter 1). Even after such procedures were taken, the mother's ethnicity was missing for 76,643 cases. For these cases, a random sample of 10,000,000 observations with complete data (with no missing data) was included to assist in imputing for missing values. The ICE procedure was used in the same models that are displayed on Table 5.1 (without the race independent variables since they were also missing). Otherwise, imputations were conducted using two models (as done previously): one for Hispanic mothers, and another for Non Hispanic mothers.

Following imputation, estimated counts of Hispanic children are also made by assuming that all children with two Hispanic parents identify as Hispanic, whereas those with only one Hispanic parent identify as Hispanic 61.4% or 69.8% of the time when the mother or father is Hispanic, respectively.

Before taking a complete count of native-born Hispanic children, however, it would first be necessary to impute missing data of a completely different variety. Namely, Hispanic ethnicity was not included on all birth certificates from Louisiana until 1990, Oklahoma until 1991, and New Hampshire until 1993. These “holes” in missing data would best be imputed separately than through the other methods described above.

To deal with the lack of a Hispanic ethnicity item on birth certificates for the aforementioned states, predicted values from a regression model were used to estimate the number of children who would be identified as Hispanic by the given state, by year. To perform this task, a dataset was created that included year, state, total births, and total counts of Hispanic children (which were calculated by using the ICE procedure for imputed values and assigning proportions identifying as Hispanic based on the ethnicities of the parents).²⁰ Regression models were conducted separately for each state that needed imputed counts.

[TABLE 5.5]

Table 5.5 displays the regression models that can be used to estimate the counts of Hispanic children for states that lacked a Hispanic item. For New Hampshire and Oklahoma, year

²⁰ Traditionally, predicting the numbers of children born is performed through Child-Woman-Ratios, which basically assume that the number of births is directly a function of women of childbearing age (commonly 15-44). Additionally, total fertility rates, which also factor in the number of males (who are able to produce “Hispanic” children in the event that they reproduce with a Non-Hispanic woman) in the population to predict the number of children can be an option to predict the numbers of births. In other words, Hispanic females (and males) can be used as independent variables to impute for missing data. This research chooses to incorporate neither, as differences in coverage and immigration will make it difficult to predict yearly Hispanic counts.

(in the form of Age of child in 2000) and the number of births for the state are able to account for over 90% of the variation in births by year. Louisiana has less success in predicting births, and the more parsimonious model uses only year as the independent variable to predict counts of Hispanic children. The simpler model was used to estimate the counts of Hispanic children for Louisiana.

[TABLE 5.6]

Table 5.6 displays the predicted counts of native-born Hispanic children to mothers who claim residence in one of the states of the U.S. As done previously, three imputation measures were used, and an average of the three imputation measures are also included. All three of the imputation procedures produced similar counts within 1% of each other, although the Census procedure consistently imputed at rates higher than the others. As also seen in Table 5.4, the Census procedure produced higher counts of Hispanic children of older ages. With the counts of Hispanic children displayed on Table 5.6, this research is now in a position to produce a demographic analysis estimate of Hispanic children aged 0 – 10 for 2000.

REFERENCES

- Barnard, John and Xiao-Li Meng. 1999. "Applications of Multiple Imputation in Medical Studies: from AIDS to NHANES." *Statistical Methods in Medical Research* 8:17-36.
- Cresce, Arthur, A. D. Schmidley, and Roberto Ramirez. 2004. "Identification of Hispanic Ethnicity in Census 2000: Analysis of Data Quality for the Question on Hispanic Origin." *Population Division Working Paper Series* No. 75. Washington, D.C.: U.S. Census Bureau.
- Donders, A. Rogier T., Geert J.M.G. van der Heijden, Theo Stijnen, and Karel G.M. Moons. 2006. "Review: A Gentle Introduction to Imputation of Missing Values." *Journal of Clinical Epidemiology*, 59:1087-1091.
- Engels, Jean Mundahl, and Paula Diehr. 2003. "Imputation of Missing Longitudinal Data: A Comparison of Methods." *Journal of Clinical Epidemiology*, 56:968-976.
- Felleng, I. P. and D. Holt. 1976. "A Systematic Approach to Automatic Edit and Imputation." *Journal of the American Statistical Association*, 71(353):17-35.
- National Center for Education Statistics. 2002. "NCES Statistical Standards." Retrieved July 23, 2011 (<http://nces.ed.gov/statprog/2002/appendixb3.asp>).
- National Center for Health Statistics. 2000. *Documentation for the Detail Natality Tape File for 2000 Data*. Hyattsville, MD: National Center for Health Statistics.
- Royston, Patrick. 2005. "Multiple Imputation of Missing Values: Update of ICE." *The Stata Journal* 5:527-536.
- Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: J. Wiley & Sons.
- Schafer, Joseph L. 1999. "Multiple imputation: a primer." *Statistical Methods in Medical Research* 8:3-15.
- Schenker, Nathaniel, Trivellore E. Raghunathan, Pei-Lu Chiu, Diane M. Makuj, Guangyu Zhang, and Alan J. Cohen. 2006. "Multiple Imputation of Missing Income Data in the National Health Interview Survey." *Journal of the American Statistical Association* 101(475):924-933.
- U.S. Census Bureau. 2010. "The Development and Sensitivity Analysis of the 2010 Demographic Analysis Estimates." Retrieved July 24, 2011 (http://www.census.gov/newsroom/releases/pdf/20101206_da_revpaper.pdf).

TABLES

Table 5.1: Models Used for Imputation of Father Ethnicities: Odds Ratios

Mother's Ethnicity	Hispanic	Non Hispanic
Father's Ethnicity (DV)	Hispanic	Hispanic
Race (omitted: white)		
Black	0.24	
neither black nor white	0.48	
Nonwhite		0.46
Education (omitted: high school)		
no high school	2.48	1.24
some college	0.62	0.92
bachelor's and above	0.42	0.64
Other Characteristics of Mother		
Married	0.99	0.56
foreign born	4.52	0.84
number of live births	1.21	1.08
Age	0.95	0.96
Social Context		
number of births in state	1.00	0.99
percent Hispanic in state	1.16	1.30
Year	0.99	1.00
Pseudo R ²	0.200	0.101

Number of births in state in units of 10,000.

Percent Hispanic in state in units of 5%

All variables significant at .001 level

Table 5.2: Results of Imputation Procedures on Missing Father Ethnicities

Sample		Father's Ethnicity									
Mother Ethnicity	Measure	Observed Percentage		ICE Imputation		Probability		Previous Observation		Census Method	
		Hispanic	Non Hispanic	Hispanic	Non Hispanic	Hispanic	Non Hispanic	Hispanic	Non Hispanic	Hispanic	Non Hispanic
Hispanic	Percent	85.82	14.18	85.83	14.17	85.83	14.17	85.74	14.26	86.88	13.12
Non Hispanic		3.08	96.92	3.08	96.92	3.08	96.92	2.94	97.06	2.95	97.05
Hispanic	Net Percent Error			0.012	-0.074	0.013	-0.079	-0.097	0.588	1.228	-7.433
Non Hispanic				0.010	0.000	-0.001	0.000	-4.482	0.142	-4.115	0.131
Hispanic	Count Discrepancy			650	-650	687	-687	-5,136	5,136	64,965	-64,965
Non Hispanic				93	-94	-9	8	-41,257	41,257	-37,881	37,880

Table 5.3: Counts of Hispanic Children Aged 0 - 10 by Imputation Method

	% Assigned as Hispanic	Observed Count	ICE	Probability	Previous	Census
Both Hispanic	100	5,290,488	5,291,138	5,291,175	5,285,352	5,355,453
Mother Hispanic, Father Non Hispanic	61.4	874,047	873,397	873,360	879,183	809,082
Mother Non Hispanic, Father Hispanic	69.8	920,485	920,578	920,476	879,228	882,604
Theoretical Count, Hispanic Children		6,469,651	6,469,967	6,469,910	6,438,872	6,468,287
Percent Net Error			0.005	0.004	-0.476	-0.021
Net Difference from Observed Count			316	259	-30,780	-1,364

Table 5.4: Predicted Counts of Total Hispanic Children by Imputation Rule and Age

Age in 2000	Total, Observed	Net Count Error				Absolute Percent Error			
		ICE	Probability	Previous	Census	ICE	Probability	Previous	Census
0	675,153	-366	-3,022	-4,339	-1,503	0.054	0.448	0.643	0.223
1	643,143	9	-490	-3,852	-998	0.001	0.076	0.599	0.155
2	623,115	249	1,428	-3,569	-880	0.040	0.229	0.573	0.141
3	615,450	460	3,323	-3,249	-305	0.075	0.540	0.528	0.050
4	599,350	406	1,140	-2,980	52	0.068	0.190	0.497	0.009
5	583,437	-249	-1,846	-2,971	-8	0.043	0.316	0.509	0.001
6	585,951	-130	-142	-2,403	54	0.022	0.024	0.410	0.009
7	573,159	-274	95	-2,181	352	0.048	0.017	0.381	0.061
8	558,193	220	3,295	-2,190	481	0.039	0.590	0.392	0.086
9	532,080	27	-2,353	-1,762	564	0.005	0.442	0.331	0.106
10	480,620	-35	-1,168	-1,284	828	0.007	0.243	0.267	0.172
TOTAL		317	260	-30,780	-1,363	0.402	3.116	5.130	1.014

Table 5.5: Regression Models Predicting Hispanic Births for States
Missing a Hispanic Item

	Louisiana	Louisiana	New Hampshire	Oklahoma
Age in 2000	-33.59*	-51.05†	-55.38**	-13.81***
Number of Births		0.03	0.17*	6.41***
Adjusted R ²	0.3494	0.3436	0.9039	0.9826

†significant at .10 level

*significant at .05 level

**significant at .01 level

***significant at .001 level

Table 5.6: Predicted Counts of Native-Born Hispanic Children, Complete Data Set

AGE2000	Predicted Counts				Net Percent Difference from Average		
	ICE	Probability	Census	Average	Probability	ICE	Census
0	820,975	820,860	821,780	821,205	-0.04	-0.03	0.07
1	782,529	782,492	783,770	782,930	-0.06	-0.05	0.11
2	754,183	754,293	755,853	754,776	-0.06	-0.08	0.14
3	743,367	743,211	744,805	743,794	-0.08	-0.06	0.14
4	723,292	723,231	725,456	723,993	-0.11	-0.10	0.20
5	704,822	704,675	707,886	705,794	-0.16	-0.14	0.30
6	697,977	697,972	701,447	699,132	-0.17	-0.17	0.33
7	684,770	684,911	688,430	686,037	-0.16	-0.18	0.35
8	671,049	671,296	674,791	672,379	-0.16	-0.20	0.36
9	643,733	643,943	648,929	645,535	-0.25	-0.28	0.53
10	589,755	589,809	595,292	591,618	-0.31	-0.31	0.62

CONCLUSION

A DEMOGRAPHIC ANALYSIS OF THE HISPANIC POPULATION

At long last, this research is finally able to complete its ultimate goal. The ultimate goal is simply to estimate the size of the Hispanic population by using the method of Demographic Analysis (DA) to the extent that the data allow. All that is needed for the execution of DA is clean data on births, deaths, and migration. As this research has shown, the trick lies in the “clean” part of “clean data.”

The chapters of the dissertation were designed to build toward the goal of producing clean data. Throughout the process, the chapters have touched upon larger dynamics that relate to the Sociology of the Hispanic population. Traveling through the Chapters and recalling some of the lessons learned in dealing with Hispanic data, there are specific assumptions and decisions that can be made regarding the data that will be used for Hispanic DA. The following is a recap of these lessons learned from the previous research and the attendant empirically-based assumptions/decisions made in handling the data on Hispanics.

Chapter 1 is the asterisk of this entire project. While the dissertation is devoted to measuring the size of the Hispanic population, Chapter 1 takes a step back and asks “who identifies as ‘Hispanic’ anyway?” The “Hispanic” identity confronts contrasting notions of its very meaning. On the one hand, it is framed as a linguistic ethnicity. On the other hand, it is commonly viewed as a race. It is a real identity to some, while it is also a construct that has been crafted from above in highly political contexts. These contrasting views of the ethnicity are reflected in the way in which people respond to the item that asks for Hispanic identification.

Nearly half of Hispanics identify as “some other race,” largely reflecting these persons’ use of “Hispanic” as a *racial* (i.e. not merely ethnic) identity. As these persons do not feel as if they can be identified within the given racial categories (including white, black, and some form of indigenous or Asian race), they simply mark “other.” Another large proportion, arguably for lack of better option, identify as “white.”

The fluidity of race and ethnicity similarly leads to the fluidity of the identification with the Hispanic ethnicity. This is most evident for the persons who do not identify as Hispanic despite their technical classification as Hispanic due to their birth in or ancestry of an American Spanish-speaking country. Theoretical approaches that imply a “hierarchy” of status and power are useful in explaining some of the exits from the Hispanic category. A “colonized mentality” (Freire 1993; Memmi 1991) is imparted to dominated groups of people, wherein the labels that become assigned to them from above (in this case the “Hispanic” ethnicity) become accepted. Those who eschew their culture and attempt to be integrated into the dominant society through such means as marriage, social identification, or socioeconomic advancement can leave their minority label and approach the majority label, as is also reflected in assimilation theory (see Gordon 1964; Neidert and Farley 1985; Park 1950). This process can be seen through patterns of racial identification, as minorities are divided into a hierarchy of races that include “collective black,” “honorary white,” and “white” categories (Bonilla-Silva 2004; Bonilla-Silva and Glover 2006).

As theorized above, race and indicators of acculturation and assimilation are highly related. Several variables dealing with socioeconomic status (e.g. income and education) are related to “whiteness” and acculturation (i.e. language spoken at home and nativity). Additionally, there is evidence that the tri-racial hierarchy has a real social significance, as

Hispanics who are more “white” are more likely to display high levels of “integration,” as variables such as the years of stay in the U.S. and marriage with a Non Hispanic white spouse (as well as the aforementioned variables) are related to degrees of “whiteness.” There is evidence of gradations of whiteness for Hispanics, as the identification as “Hispanic, non white,” “Hispanic, white,” and “Non Hispanic, white (despite having been born in or having ancestry of a Hispanic country)” are progressively-related to the acculturation and status variables.

Arguing that the “hierarchical” theoretical approaches are not completely representative of all of Hispanic identification patterns, Chapter 1 also offers an alternate explanation that the Hispanic ethnicity is contextual and ambiguous. Chapter 1 postulates that the Hispanic ethnicity is embedded within contrasting political and personal definitions which can push and pull individuals in varying distances from the Hispanic ethnicity. Persons who hold firm identities outside of the Hispanic ethnicity, despite their technical classification as Hispanic through birth or ethnicity, are argued to be “ambiguous ethnics” who straddle a fine line between (not) being Hispanic. For these persons, the social context plays a strong role to push the individual from, or pull the individual into the Hispanic category. For these persons, racial identification have huge effects with identification with the Hispanic ethnicity, as “black” and “Asian” races serve as pushes away from the Hispanic ethnicity by competing their racial identity with their ethnic identity (which is often racialized). Additionally, the ethnicity and race of one’s spouse and the size of the Hispanic population in which one lives have strong effects on the identification with the Hispanic ethnicity, indicating that the persons one is surrounded by have a large influence on one’s own identity.

Chapter 1 serves to contextualize the rest of this research by calling into question the entire notion of the Hispanic ethnicity. The Hispanic population will continue to be fluid, and

thereby introduce error in estimation of the population as the meaning of what it means to be Hispanic changes. In this way, Chapter 1 is the asterisk of the rest of the research. This caveat aside, the rest of the chapters will build upon each other to culminate in the production of a DA estimate of the Hispanic population size.

Chapter 2 highlights additional difficulties involved in the measurement of the Hispanic population that arise from the response rate of the Decennial Census. Comparing the Hispanics counted in the 1990 Census to the presumably same population counted in 2000 illustrates the questionability of the quality of the 1990 Census in its capturing of the Hispanic population. Chapter 2 demonstrated that surviving the Hispanics counted in the 1990 Census and subtracting the result by the same cohorts of persons in the 2000 to estimate emigration produces, frankly, impossible estimates of emigration for Hispanics under age 30.

Due to high level of uncertainty in the validity of estimates of emigration constructed using a residual method, alternate estimates of emigration were necessary. Fortunately, recent scholarship has produced three elegant sample-based estimates of emigration rates for Hispanics in Van Hook et al. (2006), Passel et al. (2006), and Schwabish (2009). Using a set of assumptions regarding the structure of the data and emigration rates for Hispanics vis-à-vis the entire population in general, these three studies were used to create estimates of emigration counts by age for Hispanics in place of the unreliable counts produced by the residual method. Each of these estimates produced higher counts of emigration compared to the residual estimates, and each had unique patterns of emigration by age.

As each of the emigration estimates were intended to measure the same construct, the alternate emigration rates were used to provide windows into the patterns of undercount for the 1990 Census. By comparing the average of the sample-based estimates to the residual-

based estimate of emigration, it was suggested that undercount was principally found for persons under age 65 in 1990, with rates of undercount decreasing linearly by age – or, perhaps more appropriately, by year of entry.

The rates of undercount are telling in regard to the nature of data-collection and the trust of the public in governmental institutions. The rates of undercount for the Hispanic population are related to the passage of the Immigration Reform and Control Act (IRCA) of 1986. IRCA effectively criminalized many Hispanic workers in the U.S., and Chapter 2 argues that the resultant trust of the government and the Census had suffered. For example, Hispanics who entered after the passage of IRCA were found to be heavily undercounted, while those who entered during a period within which immigrant Hispanics were able to attain “legal” documentation were reliably counted. As mentioned in the conclusion of Chapter 2, this finding implies that the government and its institutions may have difficulties soliciting engagement and participation from groups that may feel threatened by either the government or the population, perhaps affecting the counts of blacks, Hispanics, South Asians, or Middle Easterners. Data quality in general, from government institutions or otherwise, are also subject to compromise when individuals lack trust in the data collectors.

For the purposes of Demographic Analysis, Chapter 2 provides alternate estimates of emigration by age for Hispanics, which will be particularly useful in light of the difficulties with which emigration can be estimated by using residual methods. Additionally, Chapter 2 implies that undercount for 1990 decreases linearly by age, from around 11.7% for the youngest cohorts, and declining to 0 by the age of 65. Both of these results will be carried into Chapter 3 (as well as this Conclusion).

Chapter 3 explores the Hispanic Mortality Paradox in the face of its counter-explanations. One of the counter-explanations of the Paradox is that mortality rates of Hispanics are artificially low due to errors that are found in the data. Death certificate undercount has received much of the blame for the low mortality rates of Hispanics, and fortunately Arias et al. (2008) has provided rates of undercount by age for Hispanics. The second counter-explanation of the Paradox deals with the salmon bias, wherein foreign-born persons of declining health migrate to their countries of origin to die. Chapter 2 has provided a methodology to estimate emigration by age which can allow for the adjustment of death rates by the presumed population that leaves the country.

The estimates of emigration rates by age that are calculated in Chapter 3 qualify the salmon bias as a function of poor health. Indeed, residual and Schwabish estimates of emigration demonstrate that Hispanics of higher age are the most likely to emigrate, even after accounting for the undercount of younger Hispanics. However, Van Hook and Passel estimates of emigration produce emigration rates that are highest for Hispanics of working age. This means that, while the salmon bias is demonstrated to contribute to the inflation of population counts in mortality rates, it is instead possible that the strongest explanation for emigration is due to temporary migration wherein Hispanics enter, work, and leave the U.S.

After accounting for both death undercount and the emigration of Hispanics, as well as for death overcount and population undercount of Non Hispanic whites, one finds that the health advantage of Hispanics over Non Hispanic whites becomes challenged. However, when taking the population undercount of Hispanics in consideration, Hispanics are again shown to have a health advantage over Non Hispanic whites. These findings bear evidence that the assumptions one makes, and the quality of the data that are available, have big impacts on the

theoretical patterns that are found in the population. In this case, the Hispanic Mortality Paradox can be challenged on the basis of the quality of the data that are available. The most probable conclusion of Chapter 3, however, is that Hispanics do have a mortality over Non Hispanic whites, as the population undercount of Hispanics (at 4.99%, Mulry 2006) has a large effect in swaying life expectancies in favor of Hispanics.

For the purposes of DA, Chapter 3 provides death adjustments to the data that will be incorporated in this Demographic Analysis. Additionally, Chapter 3 demonstrates that distributing undercount to the population under 65 for Hispanics produces emigration rates that are much more reasonable than otherwise, leading this conclusion's DA to repeat this procedure. Because of the high undercount of Hispanics on the 1990 Census, this Conclusion will apply undercount rates to the 1990 baseline population when used.

Chapter 4 takes an exploratory approach in explaining the factors that predict interethnic childbirth. Chapter 4 takes the literatures that explain Hispanic family formation and interethnic relationships and teases out the relationships for the variables that explain interethnic childbirth specifically. While the hypothesis that the predictors of interethnic childbirth would vary greatly by the ethnicity of the mother was validated, the hypothesis that the variables would also vary by the marital status of the parents did not bear fruit.

The most notable finding of Chapter 4 is the contrasting social profiles of the women who have interethnic children when comparing Hispanic with Non Hispanic women. In general, it is suggested that Hispanic women who have interethnic children tend to be more "professional" than on average for Hispanic women, as these women tend to have had fewer previous children, be older, and be more educated. Non Hispanic mothers who have interethnic children, on the other hand, display the opposite features (compared to other Non Hispanic

mothers) including being younger, less-educated, and having had more previous children. Race, age, nativity, year, and the marital market are also shown to play roles in the choice of one's reproductive partner.

The presence of interethnic children will change the landscape of the Hispanic population in the coming years. Interethnic relationships and reproduction are becoming more popular with time, and complications regarding the count of a "Hispanic" population will also arise. Namely, the extent to which interethnic children identify as Hispanic will play a large role in the count of the Hispanic population. Fortunately, estimates of the proportions of interethnic children that are identified as Hispanic on the Census are provided by the U.S. Census Bureau (2010), and these proportions will be applied to the number of interethnic births that are counted in Chapter 5.

Chapter 5 is the continuation of Chapter 4, whose purpose for DA was to provide statistical models with which the imputation of missing data on parental ethnicities would be possible. Chapter 5 is a methodological chapter that tests the success rates of four methods that are used to provide counts of children by the ethnicities of one's parents, while also "imputing" missing data on parental ethnicities. Two of the methods used to impute data use regression techniques. One of the methods produces aggregate counts children by parental ethnic-arrangements by summing the predicted probabilities of missing parents' ethnicities that are produced from the regression models. A second method uses STATA's Multiple Imputation by Chained Equations (ICE), which directly imputes values in missing data through the use of a posterior predictive distribution of variable values that are produced by a series of regression models. A third method is used by the Census Bureau which assigns Hispanic ethnicities on missing data with the proportions of inter/intra ethnic pairings that are found in complete data

for cases that share the same mother's ethnicity, by state. The father's ethnicity, or general proportions of parental ethnicities are used if the mother's ethnicity is unavailable. A fourth method uses values from previous observations to impute missing data.

Chapter 5 demonstrates that all of the methods perform reasonably well in predicting the numbers of children born to parents of specific ethnic arrangements as found in the data. However, the previous observation and Census methods are found to have the highest rates of error when predicting the counts of children by parental ethnicities. These two methods fortuitously produce reasonable estimates of total Hispanic children, although these outcomes are largely due to the balancing of overestimates and underestimates of counts by demographic detail. Additionally, one of the regression methods (relying on predicted probabilities of parental ethnicities) produces highly erratic and extreme over- and under-estimates of Hispanic children by age, making this method potentially the most unstable. On all measures, STATA's ICE procedure performs very well.

Taking one step further, Chapter 5 predicts the number of children that would be identified as Hispanic on the Census based on the assumption that 61.4% of interethnic children are identified as Hispanic when their mother is Hispanic, while 69.8% do so when their father is Hispanic (based on U.S. Census Bureau 2010).

Net new immigration (1990-2000) has not been explored in the chapters of the dissertation. Alternate estimates of immigration by age exist in Passel and Suro (2005). Additionally, the 2000 Census "year of entry" variable can be used to count net immigrants over 1990 – 2000.²¹ This research will not use Passel and Suro (2005) estimates of migration due to its

²¹ As "Census years" begin on April 1, the "net migrants" that have arrived in 1990 according to the 2000 Census will not be entirely drawn from those who reported entry in 1990. Rather, it will be assumed that

production of lower counts of migrants than is recorded in the 2000 Census. Additionally, the immigration component that is estimated by Passel and Suro (2005) does not account for emigration (whereas the Census “net entrants” does so by definition), which would lower net immigration counts even further. The 2000 Census also has the largest sample – or “universe,” depending on one’s definition of a “sample” – making it presumably the most accurate count of net migrants from the 1990 – 2000 period. Deardoff and Blumerman (2001) estimate undercounts of the foreign-born in 2000 at 3.3% (low estimate), 6.0%, and 6.7% (high estimate). To provide three useful estimates of net migrants, this DA will use unadjusted counts as a “low” estimate, an average of the undercounted estimates as a “preferred” estimate, an average of all estimates as an “average” estimate, and a count increased by 6.7% as a “high” estimate. These percentages will be added at uniform rates by age. For brevity’s sake net new immigration will also be referred to as “net immigration.”

Death count modifications will also be conducted to provide more accurate estimates of the components of the Hispanic population. As performed by Chapters 2 and 3, death counts will be adjusted upward to account for death misclassification based on Arias et al. (2008). Additionally, deaths for states that lack a Hispanic ethnicity item on their death certificates will be imputed based on the mortality rates for states with complete data. Such age-specific mortality rates will be applied to counts (or estimates) of the Hispanic population.

As the publically-available Multiple Cause of Death dataset does not list date of birth for a decedent, the user is unable to perfectly determine the age that the decedent would have had (had he/she survived) by the beginning of the following Census year (April 1 – a decedent’s

3/4 of those who reported year of entry in 1990 immigrated after April 1. Thus, those reporting year of entry in 1990 on the 2000 Census will be weighted at $\frac{3}{4}$ of their original “person-weights” in the IPUMS dataset.

theoretical age by Census year 2000 is of particular concern). To estimate the age of a decedent for April 1, 2000, probabilities of having been born in, or having had a birthday within the Census year of one's death would be assigned to the decedent.

For persons of less than one month, only persons born in April would be at risk of not being born within the Census year. For these persons, all persons who passed away with less than one day of age were assumed to be born within the Census year. For persons over one day old, the following will be assumed:

$$Proportion_{born\ within\ year} = DaysOfAge \times \left(\frac{1}{30}\right) \quad [6.1]$$

When only ranges of days are available, *DaysOfAge* will assume the value of the middle value of the range.

For persons with one-to-eleven months of age, it will be assumed that if the months that a decedent had completed by the time of his/her death are greater than the complete months that have passed for the Census year, then the decedent has not yet had a birthday for the Census year. If the number of months that have passed through the year equal the months of age completed by the decedent, it will be assumed that half of them have been born within the Census year. Finally, if the decedent has completed fewer months than has the Census year, it will be assumed that the decedent was born within the Census year.

Following one year of age, age detail in the dataset is limited to the decedent's years of age in integers. For these persons, the following will be assumed:

$$Proportion_{birthday} = \frac{1}{24} + MonthsPassed \times \left(\frac{1}{12}\right) \quad [6.2]$$

where

Proportion_{birthday} represents the proportions of decedents that have had a birthday within the Census year prior to his/her death, and

MonthsPassed represents the numbers of months that have been completed for the Census year.

The majority of decedents will be assigned probabilities of having two ages by April 1, 2000. Total deaths for each age cohort will be estimated by summing the probabilities that decedents are members of a particular age cohort.

ONWARD: THE DEMOGRAPHIC ANALYSIS OF THE HISPANIC POPULATION, AGES 0 – 9

The only “true” Demographic Analysis that can be conducted given the structure of the data applies to the population under 10 in 2000, as these are the persons who have birth certificates that identify “Hispanic” origin. Table 6.3 in the Appendix displays the three alternate estimates of Hispanic children (Chapter 5), deaths adjusted for undercount (Chapter 3), and net migration (IPUMS, described in the previous section). Based on Van Hook et al. (2006, see also Gibbs et al. 2003), emigration for native-born Hispanics is assumed to be 0.²² Population estimates are calculated in four scenarios: low, preferred, average, and high. Most estimates of the DA inputs, in this case the birth and net migration inputs, also include an “average” estimate. All scenarios utilize the adjusted death component, and the other components used for the scenarios are as follows: the low scenario sums the ICE estimates of births and unadjusted immigrant estimates; the “preferred” scenario sums ICE births and average of the undercounted immigrant estimates; the “average” scenario sums the average estimates; and the high scenario the Census births and the high estimate of immigrants.

²² Although “native born” emigration is assumed to be 0, all native-born children whose mother reports residence in a foreign country are excluded in this analysis. De facto, all births to women who claim foreign-born residence are assumed to be emigrants (which is likely to be a high estimate).

[TABLE 6.1]

As the low scenario does not adjust for the undercount of the net immigrants, its measures of undercount are essentially the undercount of the native-born Hispanic children in the U.S. As seen in Table 6.1, the undercount of Hispanic children of age 0 rests around 11.3%, which decreases relatively linearly by age to reach 4.6% for Hispanic children of age 9. Adjusting net migration for undercount does not change the percent undercount by much for younger children, although more dramatic changes are found across the later ages (particularly under the high scenario). These counts indicate that the 2000 Census highly undercounted Hispanic children, particularly in younger ages.

The high proportions of undercounted children give greater perspective on the negative residual emigration estimates found in Chapters 2 and 3 (not to mention other previous studies including Warren and Peck 1980). Negative emigration rates using residual methodologies are largely a product of the undercount for young Hispanics (although young persons are undercounted in general, O'Hare 2009). For persons under 30, it is difficult to estimate emigration rates using residual methodologies simply because they were less likely to be counted in earlier time points when they were younger. The "collective error" scores in Chapter 2 additionally confirm that undercount decrease by age. Future research can explore reasons for the undercount of younger Hispanics. Possible reasons may be to large households that have more inhabitants than there are fields on Census forms. Another possible reason may be suspicion of the Census and the fear that the information provided may be discovered and used against respondents. Informants who provide counts for non-responding households may readily remember the number of adults living in a household, but lose count of the numbers of children. And, quite simply, heads of household may not take the Census seriously and tire after

partially filling the Census form out, accounting for themselves and perhaps other adults and sending it out before it is completed. In any of these scenarios, the youngest persons may be the most likely to be left off of the Census form.

RELATIVE UNDERCOUNT OF THE HISPANIC POPULATION, AGES 10 AND UP

The estimates of the Hispanic population over 9 cannot be fairly described as being derived from Demographic Analysis, as no births counts were used to estimate a base population. Instead, the 1990 Census count, adjusted for presumed undercount of the population (Chapter 3) was used as a baseline count of the population. For lack of a better term, the estimates of this population for which births were unavailable will be termed the “pseudo DA” estimates. The “undercount” estimates derived here would more properly be described as “relative undercount,” as the results will directly compare the undercount rates of 1990 vis-à-vis the rates for 2000. Negative scores indicate the relative undercount of the 1990 Census (after it has been adjusted for 4.99% undercount; see Chapter 3 and Mulry 2006), and positive scores the relative undercount for the 2000 Census.

Aside from the baseline population counts, the demographic components that produce the pseudo DA estimates are adjusted deaths (Chapter 3), immigration (described above), and emigration (Chapter 2). All pseudo DA estimates start with the adjusted baseline counts and adjusted deaths, and the “low” to “high” estimates are functions of the immigration and emigration inputs chosen (the figures are listed in Table 6.4 in the Appendix). The low scenario uses unadjusted net immigration and the Van Hook emigration estimates. The preferred scenario works under the assumption that the most reliable migration components are the average of the undercount-adjusted net new immigration and average of the emigration measures. Like the preferred scenario, the average scenario uses the average emigration

measure, although it uses the average of all net immigration scores (instead of the average of the undercount-adjusted net migrant scores). The high scenario uses the residual emigration estimate and 6.7% undercount assumption net immigration estimates.

[TABLE 6.2]

Table 6.2 displays the pseudo DA counts and relative rates of undercount. The “relative undercount” figures are the pseudo DA estimates subtracted by the 2000 Census counts, divided by the 2000 counts (multiplied by 100). The relative undercount figures can be interpreted as the “percent relative error” between the 1990 and 2000 Censuses, where negative scores indicate relative undercount of the 1990 Census and positive scores the relative undercount of the 2000 Census as a percentage of the 2000 Census count. The relative undercount percentage figures are predominantly negative over all the estimates (except for the high estimate). The fact that the majority of the figures remain negative indicates that one (or more) of the following are true: the undercount of Hispanics in 1990 is even greater than previously estimated; emigration estimates are too high; death counts are too high; or there is overcount in the 2000 Census. While post-Censal studies have shown that Non Hispanic whites were overcounted in 2000, they have also shown that Hispanics have been undercounted in 2000 (Mulry 2006). Thus, it is with evidence that one can rule out the overcount of Hispanics in 2000. Additionally, the wealth of literature that suggests that death counts are too low would rule out the high death count explanation (e.g. Arias et al. 2008; Rosenberg et al. 1999).

[FIGURE 6.1]

Given the lack of a set of “true” emigration counts (let alone population counts) with which to verify the accuracy of the emigration component, one would be unable to make firm decisions regarding the accuracy of the emigration components. However, one observation that

can be made is that the low DA estimate, which uses the Van Hook estimate of emigration, shows high 1990 rates of relative undercount across all ages and exorbitantly high said figures for the middle age groups. This suggests that the emigration rates by age (and magnitude) for the Van Hook estimate may be the most inaccurate of the emigration estimates used. A second observation is that Hispanics at the very last age group show very high relative undercount for the 2000 Census, suggesting that the salmon bias and/or death undercounts in Chapter 3 have been highly understated for older Hispanics. Future research into this finding may lead to the conclusion that Hispanics do not retain a mortality advantage over Non Hispanic whites, as emigration for elderly Hispanics (as well as death misclassification) may be much higher than previously anticipated.

In terms of the mean absolute relative percent error, the high DA estimates produce counts that are the most similar to the Census 2000 count which, for the most part, estimates relative undercount for the 2000 Census. However, the high DA estimate utilizes the residual estimate of emigration (which is known to be very low). The preferred and average DA estimates, on the other hand, continue to display relative undercounts for the 1990 Census. However, all three of these estimates produce close estimates, producing total relative undercount between 1.5% for 1990, to 1.3% for 2000. Additionally, these three estimates produce relative undercount scores that generally stay close to 0 for most of the age cohorts.

Clearly, all of the observations made regarding the relative undercount of Hispanics across the 1990 and 2000 Censuses are all speculation. This research would not be able to statistically verify which data component is leading to the continued relative undercount of the 1990 Census. Further research needs to be done in order to evaluate the population components. However, the most “conclusive” observations that can be made are the following:

1. Death and emigration undercount for the oldest Hispanic cohorts are likely underestimated in this research; 2. Undercount for the 1990 Census is likely higher than 4.99%; and 3. Averaging alternative estimates of the same component seems to produce reliable results.

CONCLUSION OF THE CONCLUSION

Five chapters later, this dissertation has produced estimates of the Hispanic population under age 10 for 2000 by using the method of Demographic Analysis. It has been shown that the undercount of children is particularly high, surpassing 11% for children under 1 year of age. Undercount for children falls to 5% or thereabouts for children of age 9 in a somewhat linear fashion.

For the rest of the population, pseudo DA estimates only produced higher undercount rates for the 1990 Census, and only estimates undercount for the 2000 Census when using the residual estimates of emigration (which are known to be too low). Deaths and/or emigration counts for the oldest Hispanics are shown to be in need of revision upward, as the relative undercount for the 2000 Census for these Hispanics is very high. Otherwise, the allocation of the 4.99% undercount to Hispanics under 65, the upward adjustment of Hispanic deaths, and the averaging of emigration estimates may be treated as acceptable modifications to data components, if the criterion for acceptability is to attempt to match the Hispanic population count in 2000. However, there continues to be evidence that the undercount for Hispanics in the 1990 Census is higher than 4.99%.

Clearly, the Hispanic population continues to grow in the United States. The extent of this growth, however, becomes clouded by the questionability of the statistics gathered by the Decennial Census. In order to have a firm grasp on the progression, discrimination, and political representation of Hispanics, an accurate count and accurate statistics of this population are

pivotal. As demonstrated in this dissertation, the “voices” of many young Hispanics are not being heard in the data.

The door is wide open for continued research on DA for Hispanics. The logical next step for continuing in this research project is the execution of a “sensitivity analysis” which produces a wider range of estimates by taking into account methodological uncertainty. In particular, confidence intervals and sampling error are considered in producing bands of estimates for the demographic components (U.S. Census Bureau 2010). In this case, confidence intervals can be introduced into both of the migration estimates (immigration and emigration), as well as in the models that were used to impute missing Hispanic origins on death certificates. Another logical step in this research is repeating the general research projects for 2010. With the progressive publication of the 2010 Census and NCHS health statistics, updated applications of DA for Hispanics can be performed and expanded to the population aged 0 – 19 in 2010.

As accurate data collection continues to increase in difficulty, alternate estimates of the population may become more important in evaluating the quality of the Decennial Census. The quality of the 2000 Census is far greater than the quality of the 1990 Census, although the undercount of Hispanic children using DA demonstrates that the 2000 Census has not adequately counted this demographic. The high undercount of native-born Hispanic children in this research also demonstrates that “undocumented” Hispanics are not the only populations that are not accurately counted (in contrast to what some of the more conservative politicians may decry). Count problems have also been evident in the difficulty of counting black male adults (Robinson, Adlakha, and West 2002). As Chapter 2 suggests, the capture of minority populations in the Census may be particularly difficult for populations that may feel marginalized or threatened from governmental institutions. Unless advances to Census data-

collection continue to be made, it may be advisable to completely re-imagine the Census in order to attain reliable population counts (see Swanson and Walashek 2011). In the meantime, we have Demographic Analysis.

REFERENCES

- Arias, Elizabeth., William Schauman, Karl Eschbach, Paul Sorlie, and Eric Backlund. 2008. "The Validity of Race and Hispanic Origin Reporting on Death Certificates in the United States." *Vital Health Stat* 2:1-24.
- Bonilla-Silva, Eduardo. 2004. "From Bi-Racial to Tri-Racial: Towards a New System of Racial Stratification in the USA." *Ethnic and Racial Studies* 27:931-950.
- Bonilla-Silva, Eduardo and Karen Glover. 2006. "'We Are All Americans': The Latin Americanization of Race Relations in the United States." Pp. 149-186 in *The Changing Terrain of Race and Ethnicity*, edited by Maria Krysan and Amanda Lewis. New York, NY: Russell Sage Foundation.
- Deardorff, Kevin E. and Lisa M. Blumerman. 2001. "Evaluating Components of International Migration: Estimates of the Foreign-Born Population by Migrant Status in 2000." *Population Division Working Paper Series* No. 58. Washington, D.C.: U.S. Census Bureau.
- Freire, Paulo. 1993. *Pedagogy of the Oppressed*. New York: Continuum Books.
- Gibbs, Jim, Gregory Harper, Mark Rubin, and Hyon Shin. 2003. "Evaluating Components of International Migration: Native-Born Emigrants." *Population Division Working Paper Series* No. 63. Washington, D.C.: U.S. Census Bureau.
- Gordon, Milton. 1964. *Assimilation in American Life: The Role of Race, Religion and National Origins*. New York: Oxford University Press.
- Memmi, Albert. 1991. *The Colonizer and the Colonized*. Boston, MA: Beacon Press.
- Mulry, Mary. 2006. *Summary of Accuracy and Coverage Evaluation for Census 2000*. Washington, DC: U.S. Census Bureau.
- Neidert, Lisa and Reynolds Farley. 1985. "Assimilation in the United States: An Analysis of Ethnic and Generation Differences in Status and Achievement." *American Sociological Review* 50:840-850.
- O'Hare, William. 2009. *Why are Young Children Missed So Often in the Census?* Kids Count Working Paper, Annie E. Casey Foundation, Baltimore, MD.
- Park, Robert. 1950. *Race and Culture*. Glencoe, IL: The Free Press.
- Passel, Jeffrey, Jennifer Van Hook, Frank Bean, and Weiwei Zhang. 2006. "Emigration: Measurement Issues and New Estimates." Presented at Congress and Tomorrow's Foreign-Born Population: Evaluating Immigration Projections, September 26-27, Washington, DC.

Passel, Jeffrey and Roberto Suro. 2005. *Rise, Peak, and Decline: Trends in U.S. Immigration 1992-2004*. Washington, DC: Pew Hispanic Center.

Robinson, J., Arjun Adlakha, and Kirsten West. 2002. "Coverage of Population in Census 2000: Results from Demographic Analysis." Presented at the Annual Meeting of the Population Association of America, May 8-11, Atlanta, GA.

Rosenberg, Harry, Jeffrey Maurer, Paul Sorlie, Norman Johnson, Marian MacDorman, Donna Hoyert, James Spitler, and Chester Scott. 1999. "Quality of Death Rates by Race and Hispanic Origin: A Summary of Current Research, 1999." *Department of Health and Human Services Publications* No PHSP 99-132B.

Schwabish, Jonathan. 2009. "Identifying Rates of Emigration in the United States using Administrative Earnings Records." Working Paper, Congressional Budget Office, Washington, DC.

Swanson, David and Paula Walashek. 2011. "CEMAF as a Census Method." *Springer Briefs in Population Studies, Volume 1*. Dordrecht, The Netherlands: Springer, B.V. Press.

U.S. Census Bureau. 2010. "The Development and Sensitivity Analysis of the 2010 Demographic Analysis Estimates." Retrieved July 24, 2011
(http://www.census.gov/newsroom/releases/pdf/20101206_da_revpaper.pdf).

Van Hook, Jennifer, Weiwei Zhang, Frank Bean, and Jeffrey Passel. 2006. "Foreign-Born Emigration: A New Approach and Estimates Based on Matched CPS Files." *Demography* 43:361-382.

Warren, Robert, and Jennifer Marks Peck. 1980. "Foreign-Born Emigration from the United States: 1960 to 1970." *Demography* 17(1):71-84.

TABLES

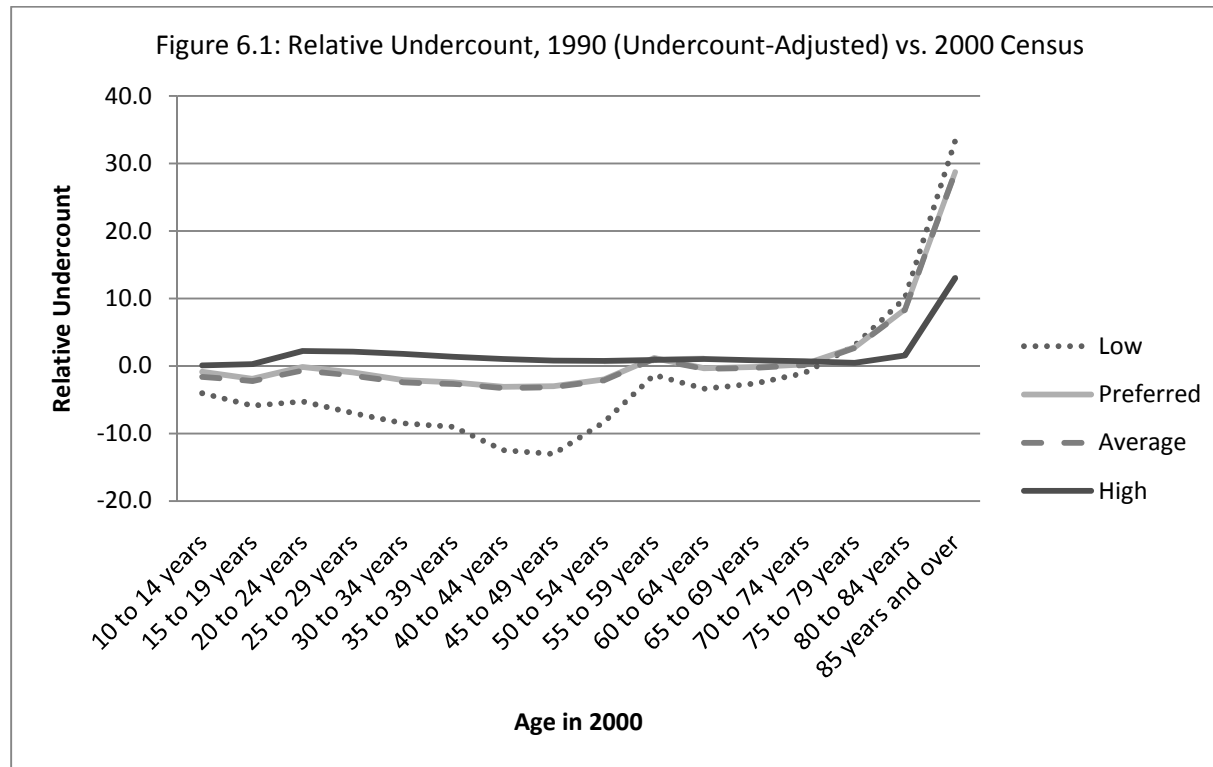
Table 6.1: Demographic Analysis Estimate of Hispanic Children

Age	DA Estimate				Census 2000 Count	Percent Undercount			
	Low	Preferred	Average	High		Low	Preferred	Average	High
0	860,500	861,355	861,024	862,736	772,918	11.3	11.4	11.4	11.6
1	826,964	827,929	827,697	829,818	752,873	9.8	10.0	9.9	10.2
2	808,759	809,926	809,754	812,380	739,528	9.4	9.5	9.5	9.9
3	799,386	800,583	800,226	802,826	723,206	10.5	10.7	10.6	11.0
4	789,117	790,510	790,297	793,610	729,450	8.2	8.4	8.3	8.8
5	786,673	788,383	788,234	792,597	724,777	8.5	8.8	8.8	9.4
6	782,515	784,287	784,280	788,948	716,774	9.2	9.4	9.4	10.1
7	780,558	782,552	782,512	787,553	721,636	8.2	8.4	8.4	9.1
8	776,062	778,248	778,144	783,459	736,165	5.4	5.7	5.7	6.4
9	757,641	760,000	760,255	766,782	724,328	4.6	4.9	5.0	5.9

Table 6.2: Pseudo DA and Relative Undercount for Hispanics, 1990 (adjusted) vs. 2000

Age	Pseudo DA Count				Census 2000 Count	Relative Undercount, 1990:2000 (After 1990 Undercount Allocation)			
	Low	Preferred	Average	High		Low	Preferred	Average	High
10 to 14 years	3,035,711	3,136,694	3,112,431	3,165,901	3,163,412	-4.0	-0.8	-1.6	0.1
15 to 19 years	2,984,700	3,111,292	3,100,934	3,179,949	3,171,646	-5.9	-1.9	-2.2	0.3
20 to 24 years	3,229,384	3,403,808	3,386,757	3,485,946	3,409,427	-5.3	-0.2	-0.7	2.2
25 to 29 years	3,149,398	3,353,822	3,337,563	3,457,872	3,385,334	-7.0	-0.9	-1.4	2.1
30 to 34 years	2,859,515	3,059,468	3,049,153	3,180,669	3,124,901	-8.5	-2.1	-2.4	1.8
35 to 39 years	2,571,135	2,756,638	2,749,699	2,863,750	2,825,158	-9.0	-2.4	-2.7	1.4
40 to 44 years	2,017,104	2,233,002	2,228,626	2,328,072	2,304,152	-12.5	-3.1	-3.3	1.0
45 to 49 years	1,543,892	1,722,383	1,719,583	1,789,292	1,775,168	-13.0	-3.0	-3.1	0.8
50 to 54 years	1,247,214	1,333,649	1,331,766	1,371,500	1,360,935	-8.4	-2.0	-2.1	0.8
55 to 59 years	947,157	971,434	970,231	968,732	960,033	-1.3	1.2	1.1	0.9
60 to 64 years	725,149	747,887	747,017	758,212	750,407	-3.4	-0.3	-0.5	1.0
65 to 69 years	583,756	598,405	597,768	604,519	599,353	-2.6	-0.2	-0.3	0.9
70 to 74 years	472,275	478,386	477,923	480,545	477,266	-1.0	0.2	0.1	0.7
75 to 79 years	336,568	335,714	335,445	328,228	326,726	3.0	2.8	2.7	0.5
80 to 84 years	197,620	194,566	194,394	182,306	179,538	10.1	8.4	8.3	1.5
85 years and over	201,030	194,047	193,929	170,346	150,708	33.4	28.8	28.7	13.0
Total	26,101,608	27,631,195	27,533,220	28,315,837	27,964,164	-6.7	-1.2	-1.5	1.3
Mean Absolute Percent Relative Undercount						8.0	3.6	3.8	1.8

FIGURE



APPENDIX TABLES

Table 6.3: Data Components for Demographic Analysis, 2000

	Component	Births				Deaths	Net Immigration			
	Source	ICE	Predicted Sum	Census		Adjusted	Unadjusted	Average Undercount	Average All	6.7 Undercount
	Estimate	Low (Preferred)	Middle	High	Average		Low	Preferred	Average	High
Age in 2000	0	820,975	820,860	821,780	821,205	3,960	43,485	44,341	43,780	44,916
	1	782,529	782,492	783,770	782,930	4,580	49,015	49,980	49,347	50,628
	2	754,183	754,293	755,853	754,776	4,676	59,253	60,419	59,654	61,203
	3	743,367	743,211	744,805	743,794	4,791	60,811	62,008	61,223	62,812
	4	723,292	723,231	725,456	723,993	4,946	70,771	72,164	71,250	73,100
	5	704,822	704,675	707,886	705,794	5,047	86,898	88,609	87,487	89,758
	6	697,977	697,972	701,447	699,132	5,489	90,027	91,799	90,637	92,990
	7	684,770	684,911	688,430	686,037	5,523	101,311	103,306	101,998	104,646
	8	671,049	671,296	674,791	672,379	6,016	111,029	113,214	111,781	114,683
	9	643,733	643,943	648,929	645,535	5,951	119,859	122,218	120,671	123,804

Table 6.4: Data Components for Pseudo-Demographic Analysis, 2000

	Component	1990 Baseline Counts	Deaths	Net Immigration				Emigration		
	Source	Adjusted	Adjusted	Unadjusted	Average Undercounted	Average All	6.7 Undercount	Residual	Average All	Van Hook
	Estimate	High		Low	Preferred	Average	High	Low	Average	High
Age in 2000	10 to 14 years	2,581,887	6,396	553,336	582,847	575,469	590,410	0	38,529	93,116
	15 to 19 years	2,358,710	7,598	776,792	818,221	807,864	828,837	0	58,041	143,204
	20 to 24 years	2,139,495	18,087	1,278,854	1,347,060	1,330,008	1,364,537	0	64,660	170,879
	25 to 29 years	2,182,579	25,868	1,219,457	1,284,495	1,268,235	1,301,161	0	87,383	226,770
	30 to 34 years	2,434,317	30,225	773,575	814,832	804,518	825,405	48,828	159,457	318,152
	35 to 39 years	2,458,528	38,138	520,376	548,129	541,191	555,241	111,882	211,883	369,632
	40 to 44 years	2,152,704	42,658	328,178	345,681	341,305	350,166	132,140	222,725	421,120
	45 to 49 years	1,723,124	43,894	210,043	221,245	218,445	224,116	114,054	178,092	345,381
	50 to 54 years	1,324,479	43,486	141,244	148,777	146,894	150,707	60,201	96,121	175,023
	55 to 59 years	977,804	44,287	90,295	95,111	93,907	96,345	61,130	57,194	76,655
	60 to 64 years	770,191	50,831	65,305	68,788	67,917	69,680	30,829	40,261	59,516
	65 to 69 years	647,315	66,038	47,758	50,305	49,668	50,958	27,716	33,177	45,279
	70 to 74 years	557,109	85,410	34,674	36,523	36,061	36,997	28,151	29,837	34,098
	75 to 79 years	436,257	95,249	20,118	21,191	20,923	21,466	34,246	26,486	24,558
	80 to 84 years	286,772	86,238	12,918	13,607	13,435	13,784	32,011	19,575	15,832
	85 years and over	438,254	222,200	8,842	9,314	9,196	9,434	55,142	31,320	23,866