THE EFFECTS OF LANGUAGE AND GEOGRAPHY-DEFINED GROUPS ON HEALTH INSURANCE CHOICE

Fredric E. Blavin

A DISSERTATION

In

Healthcare Management and Economics

For the Graduate Group in Managerial Science and Applied Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2011

Supervisor of Dissertation

*Signature*_________________________

Mark V. Pauly, Bendheim Professor of Health Care Management

Graduate Group Chairperson

*Signature*_________________________

Eric T. Bradlow, K.P. Chao Professor of Marketing, Statistics, and Education

Dissertation Committee

Scott Harrington, Ph.D., Alan B. Miller Professor of Health Care Management

Guy David, Ph.D., Assistant Professor of Health Care Management

Jeremy Tobacman, Ph.D., Assistant Professor of Business and Public Policy

The Effects of Language and Geography-Defined Groups on Health Insurance Choice

DEDICATION

I dedicate this dissertation to my wife, Susan, and parents, Harvey and Judy, for their unconditional love, support, and patience.

ACKNOWLEDGEMENTS

I would like to thank Mark Pauly, Scott Harrington, Guy David, and Jeremy Tobacman for their mentorship and insightful comments. I would also like to thank Steve Bednar, Melinda Beeuwkes Buntin, Linda Blumberg, Joanne Levy, Sean McElligott, John Holahan, Genevieve Kenney, Stacey McMorrow, Michael Morrissey, Andrew Mulcahy, Amol Navathe, Andrea Puig, Kenneth Reynolds, and David Rothschild for their feedback and support.

ABSTRACT

THE EFFECTS OF LANGUAGE AND GEOGRAPHY-DEFINED GROUPS ON HEALTH INSURANCE CHOICE

Fredric E. Blavin

Mark V. Pauly

The objective of this study is to measure how language and geography-defined groups influence participation in public health insurance programs. The theoretical model in this paper shows how better information on insurance states, gleaned through language group contacts in one's local area, can help individuals decide whether or not to take up a public benefit or remain uninsured. This study focuses on Medicaid-eligible adults and Medicaid/CHIP-eligible children who speak a non-English language at home, and uses pooled cross-sections of the 2008-2009 American Community Survey (ACS). Adapting an empirical method developed by Bertrand, Luttmer, and Mullainathan (2000), I define the main variable of interest as the interaction between *contact availability*, the density of an individual's language group in an individual's local area, and group *quality*, the information and preferences related to Medicaid that an individual's language group may possess, as measured by the language group's Medicaid take-up rate. The empirical framework also uses language group and Public Use Microdata Area (PUMA) fixed effects to control for observable and unobservable differences across language groups and local areas. The main results and sensitivity analyses strongly suggest that language and geography groups have a statistically significant impact on an individual's probability of taking-up Medicaid/CHIP: For a policy change that increases Medicaid use by 1 percentage point, the network for these language groups will increase the probability of taking-up Medicaid by 10 percentage points for adults and 7 percentage points for children. As eligibility expands under the Affordable Care Act and more people in a given language group enroll in Medicaid/CHIP, the multiplier effect could lead to higher overall program participation than might otherwise might be anticipated in a scenario without non-market interactions. These results can also help policymakers target outreach funds towards uninsured non-English speakers who are eligible for public benefits.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Chapter 1: Introduction

1.A Research Objective

Empirical evidence strongly suggests that targeted efforts to increase Medicaid or Children's Health Insurance Program (CHIP) participation reduce ambulatory care sensitive hospital admissions among children (Aizer 2003). In addition, compared to the uninsured, children enrolled in Medicaid are more likely to have a usual source of care outside of the emergency room, less likely to have unmet or delayed health needs, are more satisfied with the care they receive, and are more likely to utilize preventive health and dental care (Dubay and Kenney 2001). However, despite these potential benefits and low cost-sharing levels from the enrollees perspective, millions of low-income uninsured children and adults are eligible for Medicaid/CHIP coverage. According to the 2009 American Community Survey (ACS), approximately 19% of uninsured adults (39.9 million) and 71% of uninsured children (7.1 million) are income eligible for Medicaid/CHIP coverage through either mandatory "categorically needy" or optional "categorically related" pathways.[1]

While the majority of the literature attributes low take-up rates to lack of information (e.g., not knowing about program eligibility), low perceived benefits

[1] These are approximations and do not include all of the Medicaid eligibility pathways. For example, they do not incorporate citizenship criteria or health/disability status. These issues will be further addressed in the paper.

associated with participation, and administrative and policy design complexities (Remler and Glied 2003), there is a growing interest in the role of “social networks” in potentially reducing the costs of participation (Bertrand, Luttmer, Mullainathan 2000; Aizer and Currie 2004). Using economic theory, a new data source, and well-established empirical methods, this is the first study to measure the effects of “networks”, defined by language group behavior and geographic location, on individual health insurance take-up decisions.

Researchers in sociology and economics tend to use the terms “social networks”, “social interactions”, “peer effects”, and “neighborhood effects”, interchangeably. In its simplest form, social interactions are defined as direct non-market interactions between individuals that can potentially influence individual choices and economic outcomes such as use of physician services (Moffitt 2001; Pauly and Satterthwaite 1981). In this study, social interactions, such as conversations between friends related to Medicaid benefits, are unobservable to the researcher, whereas networks, defined by the agents for whom individuals rely upon for social interactions, are defined with available data sources. For the purpose of this study, the concept of networks is defined very broadly and provides a noisy signal about social interactions, but the application is specific to language-geography groups.

Focusing on the low-income, Medicaid-eligible population, this study uses non-English language spoken at home and geographic location to proxy for the social links between individuals and explores if there is a causal effect of language group behavior on an individual's probability of taking-up seemingly free Medicaid/CHIP benefits relative to being uninsured. Borrowing from Bertrand et al. (2000), the main variable of interest in this paper is defined as the interaction between *contact availability*, the density of an individual's language group in an individual's local area, and group *quality*, the information related to Medicaid/CHIP that an individual's language group may possess, as measured by the language group's Medicaid take-up rate. A simple example demonstrates this approach. For an individual that is part of a high Medicaid/CHIP take-up language group (e.g., above the mean), living among a high concentration of his/her language group can increase the person's probability of taking-up Medicaid. For example, suppose a Cantonese speaker migrates to the U.S. and lives in an area that his heavily concentrated with other Cantonese speakers. Because Cantonese speakers in the U.S. as a whole have a high Medicaid take-up rate, these potential contacts in the local area can provide information related to the benefits of enrollment relative to being uninsured. In contrast, for those, such as Koreans, that are part of a low take-up group (e.g., below the mean), living among a high concentration of the language group can decrease the person's probability of taking-up Medicaid relative to living among a low concentration of the language group. These potential contacts might believe that costs of enrollment outweigh the benefits (e.g., it is more convenient to remain uninsured and utilize necessary care from safety net providers), and could discourage the individual from enrolling. It is also possible that living among a high concentration of the language group increases the probability of take-up, regardless if the person is from a low or high take-up group. However, the *differential effect* on the probability of take-up will be larger among those that are part of a high take-up group, as these groups might possess more practical knowledge (e.g., information related to eligibility and necessary documentation) that could help the individual enroll in Medicaid. In other words, the main question for this study is as follows: What is the differential effect (between low and high take-up language groups) of living in areas of high concentration of a common language group on an individual's probability of taking-up Medicaid?

While language groups could influence the decision to obtain private health insurance, this paper does not focus on this outcome for two main reasons. First, networks could only indirectly influence rates of employer-sponsored health insurance (ESI) through labor market decisions, and several other studies have already measured network effects in the labor group (e.g., Ioannides and Loury 2004). Earlier versions of this paper tested the same methodology with ESI instead of Medicaid, but did not find any statistically significant results. Second, language group behavior could potentially influence decisions in the individual non-group market by reducing search costs and spreading information. However, the proportion of non-group enrollees that speak a non-English language is relatively small, and ACS does not contain sufficient information (e.g, ESI offer information) to determine eligibility for the individual non-group market. As such, the main independent variable of interest would capture unobservable characteristics that influence an individual's likelihood of searching for coverage in the individual non-group market.

This question is relevant for several reasons. First, Medicaid eligibility and language spoken at home are important; the number of adult and children who speak a language other than English at home has substantially increased over the past few decades (U.S. Census Bureau and American Community Survey, 1979-2008). Currently, approximately 12% of adults are eligible for Medicaid, 26% of whom speak a non-English language at home, and 50% of children are eligible for Medicaid/CHIP, 32% of whom have a mother who speaks a non-English language at home.[2] Demographics are shifting, and those who speak a non-English language at home have different behavioral patterns and experience different outcomes than English-speakers; for instance, children without English-speaking parents are less likely to take-up Medicaid/CHIP relative to children with English-speaking parents (Kenney et al. 2010).

Second, sociological research suggests that people who speak a non-English language at home interact mainly with others who speak that language, and are more closely linked than individuals who merely share the same ethnic background (Alba

[2] Author's tabulation of the 2009 ACS. Medicaid eligibility is determined by state income thresholds. See chapter 4 for more details.

1990; Lazear 1995). For example, several studies explored the impact of racial and ethnic group behavior on individual outcomes, but given high levels of variation within racial and ethnic groups as opposed to across groups (e.g., as a group, Asians might behavior similarly to Whites, but there is considerable variation among Asians who speak Japanese, Hmong, Cambodian, Hindi, etc…), it is very difficult to come up with accurate theoretical predictions and empirical estimates. This is consistent with the fact that on the ACS, Cantonese adults and children have relatively high take-up rates, whereas Korean adults and children have some of the lowest take-up rates. Similarly, immigrant populations may lack knowledge about the U.S. health care system and could be more likely to rely on those who speak a common non-English language for information related to Medicaid benefits. Characteristics of health care systems, such as levels of out-of-pocket spending and the efficacy of government financing, vary across countries of birth and can create similar levels of information or attitudes towards government-sponsored health insurance programs such as Medicaid/CHIP within a given language group. This study primarily focuses on language groups, but also explores the strength of country of birth networks.

Third, there is considerable variation in Medicaid/CHIP participation both across and within states (Kenney et al. 2010) and geography can play an important role in determining the strength of a language group's network effect. This is one of the first studies to utilize the new health insurance coverage questions (added in 2008) on the

American Community Survey (ACS). The ACS is a particularly rich survey because it also includes identifiers for public use microdata area (PUMA) of residence, which are geographic units within states that contain at least 100,000 people, and more aggregated residence measures such as super-PUMAs (areas of 400,000 people) and metropolitan statistical areas (MSAs). No other major surveys that produce health insurance estimates, such as the Current Population Survey (CPS), National Health Interview Survey (NHIS), or Medical Expenditure Panel Survey (MEPS), contain such detailed geographic identifiers. The ACS also contains identifiers for language spoken at home, citizenship status, country of birth, and other socio-demographic characteristics that can help identify and determine the strength of network effects. Because of these unique variables, along with the fact that each ACS cross-section contains more than 3 million individuals, I can directly control for language groups and local areas characteristics.

Fourth, there are timely policy implications associated with this study. Effective January 1, 2014, the Affordable Care Act (ACA) expands Medicaid eligibility so that states must cover adult citizens up to 138 percent of federal poverty level (FPL), primarily affecting non-parents who are currently ineligible. This portion of the ACA could have a large impact on the majority of states, as only 11 states have eligibility thresholds for parents exceeding 133 percent FPL and more than half of states do not provide Medicaid coverage for childless adults (Artiga, 2009). Even though the number of uninsured is projected to decrease by 32 million, 23 million residents are predicted to remain uninsured by 2019, including those who are eligible for Medicaid but do not take-up the benefits (CBO, 2010). Language groups could play an important role in influencing take-up, as over 2 million newly eligible adults speaks a non-English language at home.[3] Additionally, a top Obama administration priority is to ensure that uninsured children are enrolled in Medicaid or the CHIP program (Sebelius 2010); as part of the CHIP Reauthorization Act of 2009, $100 million dollars were allocated to outreach and enrollment activities. The results from this study could assist states in developing practical policy tools, such as direct advertising campaigns, aimed at these groups could affect individual behavior directly and indirectly through a network multiplier-type effect.

Finally, this paper is motivated by the theoretical and empirical challenges associated with measuring the causal effects of group behavior on individual economic outcomes. The framework aims to convince the reader that language-geography defined groups have a causal impact on an individual's probability of taking-up Medicaid. This framework unambiguously predicts that an exogenous increase in Medicaid take-up (e.g., improved outreach efforts through the ACA) will increase the individual's probability of take-up, but more so for those that live in high contact availability areas. Holding all else constant, however, an exogenous increase in contact

[3] Author's tabulation of the 2009 ACS.

availability of an individual's language group will have an ambiguous effect on an individual's probability of take-up.

1.B Roadmap

Chapter 2 gives a review of the existing literature, describes the Medicaid take-up process, and elucidates the key contributions of this study. Few studies explore the role of networks in influencing health insurance choice and none focus on Medicaid/CHIP take-up as an outcome or language-geography as the measure of networks. The first part of this chapter (2.A) reviews the literature related to the take-up of government benefits with a focus on Medicaid and CHIP. As a whole, these studies show that take-up rates vary considerably by income levels, expansion or eligibility type, geography, and other individual or family characteristics. As mentioned in 1.A, the reasons for not taking-up Medicaid/CHIP also vary. This study expands on this literature by providing new take-up estimates using the ACS (overall and by language group) and exploring how language and geography can impact an individual's probability of taking-up benefits. The second part of this chapter (2.B) reviews the theoretical and empirical research related to networks and economic outcomes. The theoretical literature shows that networks influence individual economic outcomes through the spread of information, social learning, imitation, and stigma reduction via the spread of social norms. The empirical literature explores a wide range of outcomes and empirical methods, with the former ranging from crime to earnings to obesity, and the latter ranging from experiments to instrumental variables to fixed effects models. Given the wide variety of studies, this section will primarily focus on the papers with similar outcome variables (e.g., health insurance choice or the take-up of government benefits), network definitions such as language and race/ethnicity, and empirical methods.

I develop a formal expected utility maximization model in Chapter 3 to better understand the mechanisms by which language and geography influence health insurance outcomes. In this model, individuals face the choice of taking-up Medicaid/CHIP benefits or being uninsured. The model incorporates the expected private utility and a multiplicative network utility associated with each choice that illustrates how networks provides information on a language group's common tastes, knowledge related to health care options, and valuation of Medicaid benefits relative to being uninsured. The model predicts that an increase in a language group's Medicaid take-up rate is associated with an increase in the individual's probability of take-up, whereas an increase in a language group's uninsurance rate is associated with an increase in the individual's probability of being uninsured. An increase in contact availability has an ambiguous effect on an individual's probability of taking-up Medicaid and depends on the difference in magnitude of the total utility associated with Medicaid take-up relative to being uninsured; living in a high CA area could increase the probability of take-up among high take-up groups, whereas the opposite may be true among low take-up groups.

The major empirical challenge associated with this study is to properly identify the causal effect of one's language-geography network on health insurance outcomes. While it is easy for researchers to find correlations between individual outcomes and mean language group or neighborhood outcomes, it is much more challenging to demonstrate that networks have a causal effect on individual behavior. As a simple illustration, I ran two "naïve" OLS models, where the dependent variable is a 0/1 indicator for taking-up Medicaid, and the main network variables were either the mean Medicaid rate within language groups (language group effects) or the mean Medicaid rate within PUMAs (neighborhood effects). Even after controlling for individual and household-level characteristics, the coefficient on these network variables range from .70 to .95 depending on the model. These coefficients are statistically significant at the 1 percent level, but do not provide a causal estimate of network effects; they are merely correlations that may be attributable to unobservable individual, neighborhood, or language group characteristics. These correlations can be characterized as the "reflection problem", where individual behavior determines group behavior, and not vice versa (Manski 1993).

However, additional omitted variable biases could remain. For example, it is possible that differential geographic sorting among individuals or outreach efforts that are correlated with the main variable of interest could be also explain the main results from this paper. Chapter 4 further describes these omitted variable biases and identification issues (4.A) and presents the data sources (4.B) and empirical framework (4.C) used to address these challenges. Chapter 4.B describes the data and the development of the core and expanded samples. There are two core samples: Medicaid eligible adults (19-64) who speak a language other than English at home and Medicaid/CHIP eligible children who live in a non-English household (defined by the mother's language). Both samples exclude individuals with private or other public health insurance. The expanded sample analyses include those with private health insurance and are more theoretically sound because some individuals face multiple health insurance choices. However, the results in the core sample are easier to interpret and are consistent with the results from the expanded sample. Chapter 4.C describes the empirical models, inspired by Bertand et al. (2000) who created a unique measure of language-geography networks in the context of welfare use and used language group and local area fixed effects to control for unobservable language group and local area characteristics, respectively. This study primarily uses linear probability models[4] where the dependent variable is a 0/1 indicator for Medicaid take-up (core sample) or being insured (expanded sample). I also explore multinomial logit models with the expanded sample where Medicaid, any private health insurance, or being

[4] I also use logit and probit models as specification checks.

uninsured are the three choice outcomes for the dependent variable. This chapter also describes multiple sensitivity and sub-sample analyses associated with the main model, including the use of local area characteristics instead of fixed effects and interactions of these characteristics with the network variable.

The results from this dissertation reveal that language-geography defined networks have a strong impact on the probability that an individual takes-up Medicaid/CHIP benefits. Chapter 5 presents the core sample, expanded sample, and sensitivity analyses for adults and children. The regression coefficient on the main network variable is positive and statistically significant at the 1% level in all of the main models and remains robust across the vast majority of the sensitivity analyses. Interpretation of these coefficients, especially in the multinomial logit model, is not straightforward because the key independent variable is an interaction term between two continuous variables. The most intuitive way to interpret the network coefficient is to view it as a policy multiplier effect: The core model results imply that for a hypothetical policy change that increases Medicaid use by 1 percentage point, the network for these language groups will increase the probability of having Medicaid by 10 percentage points for adults and 7 percentage points for kids.

Chapter 6 summarizes and highlights the main results, policy implications, key contributions of the dissertation, study limitations, and areas for future research. From a policy perspective, changes such as Medicaid expansions or marketing campaigns can have a direct effect on Medicaid/CHIP take-up and an indirect multiplier-effect through language-geography networks. This result implies that CMS can achieve "more-bang-for-their-buck" in areas that have a high concentration of language groups that are more likely to value Medicaid relative to being uninsured as a whole. However, in order to maximize Medicaid/CHIP take-up, if desired, CMS would need to devote additional marketing and outreach resources towards language groups that are currently uninformed about government health insurance and/or have low perceived benefits or high face costs of enrollment. The results also imply that it will be more difficult or costly to convince uninsured "hermit-types"[5] to enroll in Medicaid or CHIP.

[5] I would like to thank Mark Pauly for coming up with this name!

Chapter 2: Literature Review

This chapter reviews the most relevant existing literature related to the take-up of Medicaid/CHIP benefits and the role of networks influencing individual economic outcomes. The goal of this chapter is to identify the main contributions of this paper, given the gaps in the existing literature. Readers should refer to Klees, Wolfe, and Curtis (2010) for additional background information related to Medicaid, such as eligibility rules, scope of services, amount and duration of services, and payment issue.

2.A The Take-Up of Government Benefits

It is widely known that some of the uninsured, adults and children alike, are eligible for "free" public coverage. Most recently, Kenney et al. (2010) estimated that 7.3 million children were uninsured on the 2008 ACS, of whom 4.7 million or 65% were eligible for Medicaid or CHIP but not enrolled.[6] The authors use the Urban Institute Health Policy Center's ACS Medicaid/CHIP Eligibility Simulation Model to determine Medicaid/CHIP eligibility. This dissertation uses a different methodology to define eligibility (see Chapter 4), but obtains consistent eligibility and take-up estimates. Kenney et al. (2010) also found that participation rates substantially varied across

[6] Similar to this study, Kenney et al. (2010) defines participation as the ratio of eligible children enrolled in Medicaid/CHIP to those children plus uninsured children who are eligible for Medicaid/CHIP.

states, ranging from 55% to 95%, and individual/household characteristics. Most relevant to this study, the authors estimated participation rates of 83% among children with at least one English-speaking parent in the home, compared to 77% among children without any English-Speaking parents in the home.

Other studies produce take-up estimates ranging from 50-70% and find take-up decreases as Medicaid/CHIP income eligibility thresholds increase. By 1996, the Medicaid take-up rate among eligible children was approximately 70% (Gruber 2003; Selden et al. 1998). This translates into 4.7 million uninsured children that were eligible for Medicaid benefits during this time period (Selden et al. 1998). Another study found that the percent of children eligible for Medicaid increased by 15 percentage points between 1984 and 1992, but the fraction covered increased by only 7.4 percentage points (Currie and Gruber 1996a). Cutler and Gruber (1996) and Currie and Gruber (1996b) also found that among newly eligible children and women of childbearing age, only 23% and 34%, respectively, took up public coverage. Using the CPS and SIPP, another study found that the OBRA 1990 expansion led to an 8-percentage point rise in Medicaid coverage for children just inside the eligibility limits, and a similar rise in coverall health insurance (Card and Shore-Sheppard 2004). The authors concluded that the effect of Medicaid expansions was limited by low take-up rates among newly eligible children rather than by the crowding out of private health insurance. Similarly, LoSasso and Buchmueller (2004) estimate SCHIP take-up rates ranging from 8 to 14 percent among the newly eligible populations, and hypothesize that the newly eligible population may not be aware of their benefits especially if they had not previously participated. However, many of these children were already covered by other sources of health insurance.

Individuals do not take-up government programs because of high transaction costs due to administrative barriers and/or low perceived benefits. Two major literature reviews (Currie 2006; Remler and Glied 2003) conclude that take-up could be hindered by administrative barriers, lack of information, and “stigma” associated with government programs. However, both studies conclude that administrative barriers matter the most, whereas stigma does not have significant effect on take-up. The literature reviews also found that larger program benefits have a positive effect on participation. For example, Ettner (1997) finds that elderly people with chronic functional limitations are four times more likely to take-up Medicaid than those without limitations. Similarly, many physicians do not treat publicly insured because of relatively low reimbursement rates (Currie 2006), which can alter a patient’s valuation of Medicaid benefits relative to being uninsured. The bullets below summarize some of the key findings on why eligible populations do not take-up government benefits:

- *Administrative barriers:* Up to a quarter of Medicaid applicants cannot produce the necessary documentation (e.g., birth certificate, citizenship papers, proof of residency, and proof of income) within the required time or fail to attend all of the required interviews necessary to receive Medicaid benefits (GAO 1994). Evidence from the behavioral economics literature suggests that some of these small hassles and procrastination might explain why some individuals do not take-up program benefits (Bertrand et al. 2004), whereas changing the program design to utilize the existing tax system might make enrollment easier for eligible populations with income levels above the tax filing threshold (Congdon et al. 2011).

The results from regression analyses show that measures of inconvenience, such as perceived application length, hinder take-up, while policies such as presumptive eligibility, which lower inconvenience costs, have a significant positive impact on take-up. Design mechanisms-eliminating asset tests, offering continuous eligibility and coverage, simplifying the application and renewal processes, and extending benefits to parents--have large statistically significant positive effects on CHIP take-up rates, while mandatory waiting periods reduce take-up (Bansak and Raphael 2006). Wolfe and Scrivner (2005) obtain consistent results and also find evidence suggesting that specific outreach activities can have a positive effect on SCHIP take-up. However, the validity of these results is questionable due the small number of policy changes relative to the long time-frames for each study and the fact policy changes tend to be correlated with state budgetary considerations. In a much more methodologically sound study, Aizer (2003) examines Medicaid enrollment in California from 1996 to 2000 and the timing and placement of community-based application assistants that were part of a 2008 outreach campaign. She finds that application assistance programs had a large impact on Medicaid enrollments, particularly among Hispanic (4.6 percent) and Asian (6 percent) children relative to other children in the same community.

- *Information:* Medicaid eligibility rules are complex and individuals can qualify through a number of pathways, some of which are required by federal law and others are optional for states (Artiga 2009; Hearne 2005).[7] While most parents might have heard of Medicaid or CHIP, they do not necessary know of details related to benefit levels and eligibility. Proxies for information, such as educational attainment, provide weak results. Some evidence shows that those who are confused about Medicaid eligibility rules are 1.8 times less likely to take-up Medicaid (Stuber et al. 2000). Learning over time might also occur as lagged eligibility has a greater effect on take-up than current eligibility (Yelowitz 2000). This is consistent with the relatively low take-up estimates among the newly eligible populations who might not be aware of their benefits.
- *Stigma:* Stigma can be defined as the psychological feeling of shame or a social sense of disrespect associated with program participation (Remler and Glied

[7] See chapter 4 for a more in-depth analysis of Medicaid/CHIP eligibility rules.

2003), or to put it in simpler economic terms, the disutility arising from participation in a welfare-related program per se (Moffit 1983). Moffit (1983) models non-participation in government programs as a utility-maximizing decision, where the main cost is the stigma associated with participation. Despite having a compelling theoretical underpinning, empirical measures of stigma are difficult to interpret and the results are generally weak (Remler and Glied 2003). Stigma associated with Medicaid/CHIP participation would also be difficult to separate from stigma associated with having sufficiently low income to be eligible in the first place.

2.B Theory of Networks and Economic Behavior

Network theory can be broadly divided into two categories: The theory behind network formation and the theory behind how networks impact economic outcomes. In the most simplisitic model of network formation, individuals interact with others in their network if the benefits of doing so outweigh the costs. More formally, Jackson (2005) and Jackson and Wolinsky (1996) show the net utility $u_i(g)$ that person *i* receives from a network *g* is

$$u_i(g) = \sum_{j \neq i} (\delta_{ij})^{p_{ij}(g)} - \sum_{j \neq i: ij \in g} c_{ij}$$

Where $p_{ij}(g)$ is the number of links in the shortest path between individuals *i* and *j*, $c_{ij}>0$ is the cost to maintain a direct relationship with person *j*, and δ_{ij} is a factor between 0 and 1 that indicates the benefit from a direct relationship with person *j* and is raised to higher powers for more distant relationships. For example, consider a network where person 1 is linked to 2, 2 is linked to 3, and 3 is linked to 4; person 1 gets a benefit of δ_{12} from the direct connection with person 2, a benefit of $(\delta_{13})^2$ from the indirect connection with person 3, and a benefit of $(\delta_{14})^3$ from the indirect connection with person 4. Since $\delta_{ij} < 1$, there is a lower benefit from an indirect connection than a direct one. However, individuals only pay costs for maintaining direct relationships whereas indirect relationships are costless. The model also shows which networks are efficient and which networks are likely to form when individuals choose their own links as modeled through pairwise stability. A network is pairwise stable if no player wants to sever a link and no two players both want to add a link in the network.

It is important to note that the power relationship between δ_{ij} *and* $p_{ij}(g)$ can only be empirically tested when complete data on the network structure are available. In this study's empirical analysis, actual direct and indirect connections within each individual's network are unobserved and therefor I assume network formation is exogenous. In other words, this dissertation assumes that individuals are born into language groups and area of residence is exogenous[8], with this latter assumption tested through various sensitivity models.

---

[8] Or endogenous, but in a way that is uncorrelated with health insurance choice.

The more pertinent models relate to how networks influence individual economic outcomes. Sociologists have had a major influence on this area of network research (Granovetter 1973; 2005) and have developed similar prediction as economics: networks affect the flow and quality of information and networks influence behavior through the spread of social norms. While some economists have analyzed how networks influence individual outcomes through social norms, such as peer pressure, role models, stigma, or social approval (For example, see Akerlof 1980; Lindbeck et al. 1999; Besley and Coate 1992; Moffit 1983), this section, and dissertation as a whole, focuses economic models that explain how networks influence individual behavior through the spread of information.

Bala and Goyal (1998) develop a model of Bayesian learning where agents use their own past experience as well as the experience of their neighbors to guide their decision making. Through various assumptions, the authors show that in a connected society, “local learning ensures that all agents obtain the same payoffs in the long run” and eventually converge to choosing the same action. Ellison and Fudenberg (1993) develop a model of social learning where agents take into account the experiences of their neighbors in deciding which of two technologies to use. Ellison and Fudenberg (1995) also examine how word-of-mouth communication accumulates information of individual agents and may lead all players to adopt the action that is on average superior, depending on what people are saying. These models are specific to technological adoption but they illustrate how social learning can lead to different outcomes among heterogeneous groups.

Banerjee (1992) and Bikhchandani et al. (1992) describe how networks influence individual behavior through imitation of behavior. Both of these models make predictions of individual behavior based on information from groups or previous decision makers. Banerjee (1992) analyzes a sequential decision model in which each agent looks at the decisions made by previous agents before making their own decisions. The model produces an inefficient equilibrium where people do what others are doing rather than using their private information. Bikhchandani et al. (1992) find that localized conformity of behavior can be explained by informational cascades, which occur when it is for an individual to follow the behavior of the preceding individual without regard to his own information.

Networks can also be viewed in the context of search costs. Pauly and Sattherwaite (1981) show that the reputation of a physician is formed through information shared between consumers. They find that a higher number of physicians lowers the ratio of friends per provider, and therefore increases search costs because consumers communicate with others to learn the reputation of the providers. Empirically, they look at primary care physician services and show that increasing the number of sellers leads to price increases. In the context of electronic marketplaces, Bakos (1997) views networks as an intermediary between the buyers and sellers in a market, creating a marketplace that lowers the costs to acquire information about seller price, product details, and product availability.

2.C Network Empirical Applications

There is little to no literature on how language and geography influence the probability of taking up Medicaid or having health insurance in general. However, there are several studies that explore how language, race/ethnicity, and geography influence different economic outcomes. There are also a myriad other studies, both within and outside the field of health economics, that test the role of networks, peer, or neighborhood effects in various markets.

The empirical method in this dissertation is motivated by studies that explore how language and geography-defined networks influence welfare use and health care utilization. The empirical framework for this paper is derived from Bertrand et al. (2000), who examine the role of networks in welfare participation using data on language spoken at home and geography to define networks. The authors hypothesize that, by reducing the stigma associated with welfare use and through the spread of information, being surrounded by high welfare-using contacts increases the individual's welfare recipiency more than being surrounded by low welfare-using contacts. They use the number of people in one's local area who speak one's language to measure contact availability and the mean welfare use of the language group as a proxy for network quality. The interaction between these two variables (network quality*network quantity) defines they key variable of interest. They also control for local area and language group fixed effects. Their results imply that networks would raise the responsiveness of welfare take-up policy shocks by 15-27%. Deri (2005) uses a similar method to estimate the impact of language-geography networks on health care utilization among immigrants in Canada. She finds that networks have an impact on health care utilization and that the utilization of services by immigrants increases with the number of physicians who speak their language in their neighborhood. One of her key results is that a policy that increases the use of regular doctors by 1 percentage point will increase the probability of having a regular doctor in the language group network by 4.9 percentage points.

Aizer and Currie (2004) analyze the effects of networks on the utilization of publicly-funded maternity care in California. They define networks using 5-digit zip codes and a woman's racial or ethnic group. The outcome they focus on is whether women who went on to have a public delivery used public services beginning in the first trimester of their pregnancies. The authors find correlations between individual use of publicly-funded maternity care and group use. They run various models, including one similar to Bertrand et al (2000), and find that the correlations still exist. However, the authors reject the hypothesis that the estimated network effects represent information sharing within groups. They find that network effects persist even among women who already knew about the services because they had used it in the past. Unfortunately, this study is limited by a loose definition of networks (ethnicity), lack of data on other Medicaid eligible groups besides pregnant woman, and lack of data beyond California.

Several other papers attempt to measure how language and racial/ethnic group behavior influence individual behavior. Gresenz, Rogowski, and Escarce (2007) find that for Mexican-American immigrants, living in an area populated by relatively more Hispanics, more immigrants, or more Spanish-speakers increases access to care (e.g., usual source of care and number of office visits). The authors believe that this is facilitated by the flow of information among people in the local area about where to go for care and what processes to use to get there. They also find that the network effects are stronger for more recent immigrants compared to those who are more established in the U.S., and find no effects on access to care for U.S. born Mexican-Americans. Devillanova (2008) uses a dataset with large sample of undocumented immigrants in Milan and contains a direct indicator of information networks-whether an immigrant was referred to health care opportunities by a strong social tie. The dependent variable in this analysis is the log of time spent in Italy before an immigrant first utilized health care. The key network variable is a dummy indicating whether or not the individual came in contact with Naga, a voluntary association which offers free primary care to irregular immigrants, through a strong social network of friends or relatives. Overall, the author finds that networks significantly accelerate health care utilization, reducing the time to visit by about 30%.

Borjas (1992, 1995) introduce ethnic capital into an economic model of intergenerational mobility. The author defines the dependent variable as the child's educational attainment level (or earnings) and defines the key independent network variable as the mean educational attainment of the ethnic group of the father's generation. Borjas and Hilton (1996) use a similar method to show that that types of public benefits used by an ethnic group's previous generation can predict those used in the current generation. However, this methodology does not sufficiently control for unobservable personal and ethnic group characteristics that might be correlated with the network variable.

To investigate the effect of ethnic capital in the context of this study, one can regress individual health insurance status on the mean health insurance status of the ethnic group in the previous generation (along with observable individual and ethnic group characteristics). However, this type of model, and the model used by Gresenz, Rogowski, and Escarce (2007), suffers from two omitted variable biases: (1) Omitted personal characteristics may be correlated with the network variable and (2) Omitted ethnic group characteristics may be correlated with the network variable (Bertrand et al. 2000). This study includes both neighborhood and language group fixed effects in order to avoid biases associated with omitted language/ethnic group and neighborhood characteristics. In addition, networks defined by language group as opposed to ethnic group provide a more precise measure of social links because ancestry can often include individuals who are loosely connected to their ethnic group (Alba 1991; Lazear 1995). These econometric concerns are further discussed in Chapter 4.

To the author's knowledge, only one other study explores the effects of networks on health insurance outcomes. Using panel data from the University of California, Sorensen (2006) quantifies the impact of social learning on individuals' choice of employer-sponsored health plans. To avoid simultaneity problems, the author focuses on the choices of newly hired employees and assumes their health plan choices are influenced by coworkers, but not vice versa. Sorensen finds that health plan choices are correlated across individuals within the same department. He also uses discrete choice models and finds large and statistically significant social learning effects that are robust across campuses and model specifications.

There are also several related neighborhood effects studies to note. One study finds that socioeconomic factors, including the racial composition of an area or its income level, can have independent effects beyond the sum of the effects of the race and income of individuals in the area (Subramanian and Kawachi 2004). This result is consistent with the Gautreaux Experiment (Rosenbaum 1995) and Moving to Opportunity Experiment (MTO, Katz et al. 2001; Ludwig et al. 2001). Gautreaux was a US housing desegregation projected initiated by court order. Public housing residents were essentially randomly assigned to neighborhoods (urban and suburban) in Chicago in order to mitigate high concentrations of poverty. Rosenbaum (1995) found that women allocated to better, typically suburban, neighborhoods experienced better outcomes, e.g., they were more likely to find employment and leave welfare. Due to its initial success, the Gautreaux experiment became a model for similar programs in various metropolitan areas and inspired the national MTO program. MTO was a true random assignment demonstration. Initial results are suggestive of strong neighborhood effects on child problem behaviors, child and adult health outcomes, and juvenile crime (Katz et al. 2001; Ludwig et al. 2001).

Pagan and Pauly (2006) find that community-level uninsurance rates are positively associated with having reported unmet medical needs, but only for insured adults. They find that, on average, a five percentage point increase in the local uninsured population is associated with a 10.5 percent increase in the likelihood that an insured adult will report having unmet medical needs during the 12-month period studied. Pauly and Pagan (2007) further expand and conclude that reducing the size of the uninsured population yields important spillover benefits to the insured population that go beyond a lower charity care burden e.g., the quality of care available to everyone locally as a result of the low demand for quality by the uninsured.

Finally, the existing literature shows that social networks play an important role influencing individual behavior in labor markets, as a means of matching workers and firms[9] (Ioannides and Loury 2004), crime, as a way of explaining the high variance of crime rates across time and space (Glaeser et al. 1996), retirement plan decisions (Duflo and Saez 2003), juvenile behavior (Gaviria and Raphael 2001), educational attainment (Sacerdote 2001; Evans et al. 1992) and obesity (Christakis and Fowler 2007). This chapter concludes with a discussion of the key findings and methodological contributions of some of these studies.

The use of randomized or natural experiments highlights how networks can alter the flow of information. Duflo and Saez (2003) used a randomized experiment to analyze the role of information and social interactions in employees' decisions to enroll in a Tax Deferred Account (TDA) retirement plan within a large university. The experiment provided financial incentives to a random sample of workers within a random subset of departments to attend a TDA information fair sponsored by the university. The nature of the experiment allowed the authors to compare results among treated individuals in treated departments, untreated individuals in treated departments, and untreated individuals in untreated departments. The experiment increased the attendance rate for treated individuals by five-fold relative to the controls, and tripled the attendance rate for untreated individuals within treated departments. The authors also found that effect on TDA enrollment is almost as large for individuals in treated departments who did not receive the financial incentive as for

[9] This study will not focus on employer-sponsored insurance (ESI) as an outcome because ESI rates are largely explained by labor market decisions

those who did. This result can be attributable to differential treatment effects, social network effects, and motivational reward effects. Sacerdote (2001) finds strong evidence for the existence of peer effects in student outcomes at Dartmouth College. Using data on freshman year roommates and dormmates, both of which are randomly assigned, he finds that peer effects in GPA occur at the roommate level and peer effects in fraternity membership occurs at the roommate and dorm level.

Christakis and Fowler (2007) and Fowler and Christakis (2008) use a panel of interconnected networks as part of the Framingham Heart Study to determine if obesity is spread through person-to-person interactions. They use a panel logistic regression models in which the “ego’s” (the individual) obesity status is a function of various personal attributes, including lagged obesity status, and the “alter’s’” (e.g., friend, sibling, or spouse) current and lagged obesity status. They use generalized estimating equations to account for multiple observations of the same ego across examinations and “ego-alter” pairs. They find that a person’s chances of becoming obese increased by 57% if he/she had a friend who became obese in a given time interval. They also obtained similar findings from siblings (40%) and spouses (37%).

Cohen-Cole and Fletcher (2008) respond by arguing that Christakis and Fowler (2007) fail to control for contextual effects, creating spurious inference on the social networks effect. The authors are able to replicate a similar model and obtain similar results from Christakis and Fowler (2007) using the Add Health panel dataset, a national sample of 7^{th}-12^{th} graders who transition into early adulthood. Their first model did not control for school-specific trends that account for any environmental factors shared by individuals at the same school. After including school-level fixed effects, the authors find a large drop in the magnitude and statistical significance of the coefficient of interest. Trogden et al. (2008) also use the Add Health panel to estimate peer effects for adolescent weight. They control for the endogeneity of peer groups by using a combination of school fixed effects, instrumental variables, and alternative (exogenous) definitions of peers. Even after controlling for endogenous peer effects, they find that mean peer weight is correlated with adolescent weight. The conflicting results from these obesity studies highlights the theoretical and econometric challenges associated with network-related studies.

Chapter 3: Theory

In this chapter, I build a model that illustrates how language and geography-defined networks shape health insurance choices. Better information on insurance states, gleaned through the network, can help consumers decide whether or not to take up a Medicaid benefit or remain uninsured.

3.A Assumptions

Networks can play an important role in providing consumers with information related to health insurance choices. Health insurance products can be complex and vary across geographic markets; consumers must make choices based on risk and can choose policies with various levels of benefits, price schedules, deductibles, networks, and/or coinsurance rates. Similarly, some individuals do not take-up Medicaid benefits because of administrative barriers (high transaction costs), lack of information and perceived benefits, and "stigma" associated with government programs (see Chapter 2 for more information). The model described below illustrates how language-geography networks influence individual outcomes by providing information on the value of health insurance choices among the individual's language group.

A simple theoretical model of expected utility maximization illustrates this behavioral effect. This model builds on the stylistic features (expected utility maximization model) from Herring (2005) and some of the social interaction mechanisms from Brock and Durlauf (2001). Herring (2005) developed a simple utility maximization model that predicts how the existence of charity decreases the propensity to purchase private health insurance. Brock and Durlauf (2001) study generalized logistic models of individual choice which incorporate terms reflecting the desire of individuals to conform to the behavior of others in an environment of non-cooperative decision-making. The following assumptions characterize the key features of the model:

- To clarify, this model deviates from the random utility framework in Brock and Durlauf 2001) and Brock and Durlauf (2002)[10], even though these models have nice econometric properties associated with logit and multinomial logit models, respectively. There are two main reasons for this. First, expected utility maximization models are widely used when dealing with choices related to risk and uncertainty and the predictions from this type of model are intuitive and clear-cut relative to the theoretical complications associated with the models in Brock and Durlauf. Second, while using a non-linear logistic framework might be more theoretically sound, it creates multiple complications for empirical implementation. The main network variable, an interaction of two continuous variables, is much easier to interpret with a linear probability model (LPM). LPMs also have much more flexibility in terms of using fixed effects relative to

[10] See McFadden (1974, 1981) for a discussion of random utility models.

the non-linear counterparts. Chapter 4 provides a more in-depth discussion and Chapter 5 compares the results from LPMs, logit, and multinomial logit models.

- This model focuses on the Medicaid-eligible population whom speak a language other than English at home and assumes these individual's face the choice of taking up Medicaid or being uninsured. These assumptions make the model more tractable, but can be relaxed to incorporate more than two choices or different populations.
- The model also assumes that individuals face a disutility from the total amount of medical care expenses and the valuation of risk associated with the variation in the amount of expenses. For simplicity, the model assumes each individual faces the full cost of medical expenses if they are uninsured. The existence of uncompensated care (in many instances, Medicaid will retroactively reimburse hospitals for treatment of those who are uninsured but eligible) reduces the realism of this assumption, however, the Medicaid eligible population does face non-zero medical expenses and risk and this parameter captures the benefit of coverage relative to being uninsured.
- Network formation is exogenous and in a state with imperfect information, each individual interacts with others in their common language group and local area. Exogenous network formation is a fair assumption due to the fact that individuals are born into language groups and differential selection across

geographic areas does not appear to be a major issue (See Chapter 4 for a more detailed discussion of this issue). The sociological and economic literature also strongly suggests that non-English speakers mainly interact with others in their common language group.

- Conditional upon eligibility, language groups with higher Medicaid/CHIP take-up rates are presumed to have a greater knowledge about the program. This assumption is sensible because it merely relates knowledge about Medicaid/CHIP with actual experience and encounters with the program. In other words, this assumption implies that language groups with higher take-up rates know more about important Medicaid/CHIP details such as eligibility rules, application requirements, and potential benefits of coverage.
- Each individual derives utility from the beliefs about the behavior of others in his language group. The model assumes individual i is influenced by what he thinks others in his group are doing via expectations derived by composition of his local area, not by their actual behavior per se.
- For comparison purposes, this model assumes there are two states of the world, one where consumers are perfectly informed about characteristics of each health insurance choice, and one where consumers are imperfectly informed. I assume that there are no social interaction effects in the former, whereas individuals rely on their language group for information in the latter.

- The model uses the following indices: individual i, language group k, and local area j

3.B The Model

The expected utility of an uninsured individual i facing uncertain costs of medical care can be expressed as

$$EU_{u,i} = I_i - EX_i - \frac{1}{2}AP_i var(X) \qquad (1)$$

Where I_i is i's income, X_i is the total amount of health expenses, and $\frac{1}{2}AP_i var(X)$ is the ex-ante valuation of risk due to the variation in the realization of X_i. AP_i is the Arrow-Pratt relative risk aversion coefficient and $var\ \ X$ is the variance of X. For further discussion on the valuation of risk, see Feldman and Dowd (1991) on the derivation and Herring (2005) and Pauly, Blavin, and Meghan (2009) for additional applications.

The expected utility of individual i if fully insured by Medicaid in this state of the world is

$$EU_{m,i} = I_i - EC_i \qquad (2)$$

Where C_i are the total costs associated with Medicaid take-up. C_i includes any premium and cost sharing that individual i may face and indirect costs (e.g., time and hassle costs) associated with taking-up Medicaid.

Individual i chooses to take-up Medicaid if $EU_{m,i}$> $EU_{u,i}$. Individual i's *propensity* for taking-up Medicaid, as defined as the difference between expected utilities, is expressed as

$$Y_{m,i}^{*} = EU_{m,i} - EU_{u,i} = EX_i + \frac{1}{2}AP_i var(X) - EC_i \qquad (3)$$

In other words, consumers are more likely to take-up Medicaid if the expected benefit of doing so, measured by the decrease in medical expenses and risk, is greater than the expected cost.

Now, suppose that each individual faces imperfect information related to the costs and benefits of Medicaid relative to being uninsured and relies on his/her language group for information. In the presence of social interactions, the sum of private and social utility for individual i if uninsured is:

$$\begin{aligned} EU_{u,i} &= [I_i - EX_i - \frac{1}{2}AP_i var(X)] + \\ &u_{j,k}^{e} * [I_i - EX_i - \frac{1}{2}AP_i var(X)] \\ &= 1 + u_{j,k}^{e} * [I_i - X_i - \frac{1}{2}AP_i var(X)] \end{aligned} \qquad (4)$$

Where

$$u_{j,k}^{e} = P_{j,k} * (u_k - u) \quad (5)$$

Equation (4) shows that each consumer receives more utility from being uninsured if they expect a higher proportion of their language group to be uninsured. The term $1 + u_{j,k}^{e}$ embodies a multiplicative interaction between the expected private utility associated with being uninsured and the expected social utility associated with being uninsured. $u_{j,k}^{e}$ is the expected average choices from individual i's perspective of the proportion of his/her language group that is uninsured and $(u_k - u)$ is the actual de-meaned group uninsurance rate, which proxies for the language group's valuation of being uninsured as a whole (e.g., the language group's cultural beliefs, physical characteristics, and experience with health insurance schemes in native country that shape the language group's proclivity towards being uninsured relative to taking-up Medicaid). As equation (5) illustrates, each individual does not directly observe the actual insurance choices of his/her language group. Rather, they receive a *signal* of this valuation, which depends on the proportion $P_{j,k}$ of the person's local area j that belongs to the same language group. This concept is referred to as contact availability throughout the paper.

Similarly, the expected total utility of taking-up Medicaid in the presence of social interactions is

$$EU_{m,i} = \; 1 + m_{j,k}^{e} \; * [I_i - EC_i] \quad (6)$$

Where

$$m_{j,k}^{e} = P_{j,k} * (m_k - m) \qquad (7)$$

Equations (6) and (7) show that the expected utility associated with Medicaid take-up has a multiplicative social interaction effect, with a converse explanation to the one described in the previous paragraph. Intuitively, as $m_{j,k}^{e}$ increases, more information related to Medicaid (e.g., eligibility and enrollment requirements for take-up) flows through a language-geography group and influences each individual's expected utility of take-up.

Each person chooses to take-up Medicaid if $EU_{m,i}$> $EU_{u,i}$. In addition, consistent with the method used in (3), individual i's propensity for taking-up Medicaid is defined as the difference between expected utilities:

$$Y_{m,i}^{*} = \quad 1 + m_{j,k}^{e} \quad * [I_i - EC_i] -$$

$$1 + u_{j,k}^{e} \quad * [I_i - EX_i - \frac{1}{2} AP_i var(X)] \qquad (8)$$

$$Y_{m,i}^{*} \quad m_k, u_k, P_{j,k} \quad = [1 + P_{j,k} * (m_k - m)] * [I_i - EC_i)] -$$

$$[1 + P_{j,k} * (u_j - u)] * [I_i - EX_i - \frac{1}{2} AP_i var(X)] \qquad (9)$$

Differentiating the propensity to take-up Medicaid, as expressed by (9), by m_k, u_k, and $P_{j,k}$, respectively, yields the following testable hypotheses:

$$dY^*_{m,i} \quad d\,m_k > 0 \tag{10}$$

$$dY^*_{m,i} \quad d\,u_k < 0 \tag{11}$$

$$\begin{gathered} dY^*_{m,i} \quad d\,P_{j,k} > 0, if \\ (m_k - m) * [I_i - EC_i] \\ > (u_k - u) * [I_i - EX_i - \frac{1}{2} AP_i var(X)] \end{gathered} \tag{12}$$

$$\begin{gathered} dY^*_{m,i} \quad d\,P_{j,k} < 0, if \\ (m_k - m) * [I_i - EC_i] \\ < (u_k - u) * [I_i - EX_i - \frac{1}{2} AP_i var(X)] \end{gathered} \tag{13}$$

Equation (11) shows that an exogenous increase in a language group's Medicaid take-up rate is associated with an increase in an individual's propensity towards participating in Medicaid. In contrast, (12) shows that an exogenous increase in a language group's uninsurance rate is associated with a decrease in an individual's propensity towards participating in Medicaid.

Equations (12) and (13) illustrate how contact availability has an ambiguous effect on the propensity to participate in Medicaid. An increase in a person's contact availability is associated with an increase in the propensity towards taking up Medicaid *if* the total utility (individual utility plus social utility that incorporates language group behavior) associated with taking-up Medicaid is greater than the total utility associated with being uninsured. Combining these first-order conditions leads to the following testable hypothesis: Being surrounded by a high Medicaid utilizing language group increases an individual's probability of taking-up Medicaid more than being surrounded by a low Medicaid utilizing language group.

It is important to note that this model does not explicitly predict what the impact of being surrounded by others who are part of another language group besides one's own. However, the results imply that being surrounded by a high concentration of any language group that has a high take-up rate could have a positive impact on the individual's probability of take-up. This is consistent with the "naïve" regression results mentioned in the introduction and further discussed in Chapter 5. These results show that the take-up rate in one's local area (overall, regardless of language) has a very strong and positive impact on the individual's probability of take-up. However, this independent variable captures several local area omitted variables that are correlated with the main independent variable.

Chapter 4: Empirical Framework

4A. Data

Household and Individual-Level Microdata

This analysis uses pooled cross-sections of the 2008-2009 public use microdata sample (PUMS) of the Census' American Community Survey (ACS). I downloaded and analyzed an augmented version of the survey, the Integrated Public Use Microdata Series (IPUMS), from the University of Minnesota Population Center (Ruggles et al. 2010). The ACS is part of the reengineered decennial census program and provides detailed information every year instead of every ten years. ACS data are collected continuously using independent monthly samples and are designed to produce nationally representative economic, social, demographic, and housing information (Turner et al. 2009). The ACS samples approximately 1.3 million housing and 3 million person records annually throughout the U.S. and Puerto Rico. The survey is administered using a mixed-mode approach-over half of the sample is completed by mail and the rest is completed by telephone or in person-and has a reported response rate of 98% in both years (Kenney et al. 2010; U.S. Census Bureau 2010).

This analysis is feasible because of the uniqueness and size of the ACS data. The U.S. Census Bureau has been conducting the ACS over the past decade, but only recently added a question related to health insurance status for each individual in the household in 2008. To the author's knowledge, it is the only nationally representative survey that contains information on health insurance coverage, language spoken at home, and detailed geographic area of residence. The health insurance question on the ACS questionnaire is in Figure 1.

Research suggests that the that ACS coverage estimates are valid and highly consistent with other federally-funded representative surveys such as the Current Population Survey (CPS), Medical Expenditure Panel Survey (MEPS), and National Health Interview Survey (NHIS) (Turner et al 2009). However, there is a concern that the survey understates Medicaid and CHIP coverage relative to the other surveys because the ACS does not specifically mention CHIP or provide names for state's particular Medicaid and CHIP programs (Kenney et al. 2010). This could create confusion between Medicaid and private non-group coverage, leading to an underestimate of the former and an overestimate of the latter. To address this underreporting, in addition to the known underreporting of public coverage on household surveys in general, I applied a modified set of logical edit rules that were developed by the U.S. Census Bureau (Lynch et al. 2010). These edits, displayed in Figure 2, have also been harmonized to analyze changes over time (Ruggles et al. 2010). However, it is important to note that the regression results in Chapter 5 are insensitive to the use of these edits[11], which provides evidence to favor the notion that Medicaid underreporting is uncorrelated with language and geography.

The ACS has a distinct advantage over the CPS, MEPS, and NHIS because of its large sample size (approximately 3 million individuals or 1% of the US population) which allows for estimation at the local area level. Whereas the other surveys only allow estimation at the national, state, or census region level, the ACS contains identifiers for public use microdata areas (PUMAs), which consist of populations of approximately 100,000 individuals, super-PUMAs, areas with approximately 400,000 individuals, and metropolitan statistical areas (MSAs), which are larger urban areas that could contain multiple PUMAs.[12]

In addition to the geographic identifiers and health insurance variables, the ACS also contains information on each individual's language spoken at home. According to both the 2008 and 2009 ACS data, approximately 20% of the non-elderly population speaks a language other than English at home (close to 100 different languages). The ACS also contains key individual, household, and family socio-demographic characteristics such as on citizenship status, year person came to live in the U.S.,

---

[11] Data not show. The results from the data without the logical edits were used in the dissertation proposal and are highly consistent with the final results.

[12] MSA identifiers are derived from a crosswalk created by the Missouri Census Data Center (http://mcdc.missouri.edu/webrepts/geography/). PUMAs that belong to multiple MSAs are assigned to the MSA with the largest Census population.

migration status, country of birth, age, sex, race, marital status, disability status[13], educational attainment, income, household size, level English fluency, occupation, industry and work status. Given the large sample size of the ACS, I am able to produce neighborhood characteristics, such as racial composition, income, and education, at the PUMA and MSA-level.

The ACS also contains individual and household-level sample weights that can be used to produce statistics representative of the population. This study uses the weights to produce descriptive statistics in 5.A and mean-level variables (e.g, language group take-up rate), but does not use them for the regression analysis. The reason for this is because when sampling weights are solely a function of independent variables included in the model, which they are in the ACS, unweighted ordinary least squares (OLS) estimates are preferred because they are unbiased, consistent, and have small standard errors than weighted OLS estimates (Winship and Radbill 1994). However, when weights are used in a sensitivity analysis, they produce results that are consistent with the unweighted models.

Other Data Sources

[13] The ACS contains six questions asking if the respondent has serious difficulty hearing, seeing, concentrating/remembering/making decisions, walking or climbing stairs, dressing or bathing, and doing errands.

I obtained 2008 and 2009 Medicaid/CHIP eligibility thresholds, determined by the maximum percent of the federal poverty level (FPL) within a state for Medicaid/CHIP or similar programs (e.g., premium assistance), for parents, non-parents and children from the Kaiser Family Foundation (Cohen Ross and Marks 2009a; Cohen Ross et al. 2009b). Figure 3 shows how eligibility thresholds vary by state and familial status. For example, eligibility for parents varies from 24% FPL in Alabama to 300% FPL in Connecticut, Maine, Massachusetts, and Vermont, and eligibility for children ranges from 160% FPL in North Dakota to 400% FPL in New York. In addition, childless adults are eligible for Medicaid or Medicaid-related benefits in 24 states.

I use state-level data from Zuckerman et al. (2009) and Hill et al. (2009) as part of the sensitivity analysis discussed in Chapter 4.C. Zuckerman et al. (2009) collect data on average fee by state and procedure and develop a state-level Medicaid Fee Index. This measure is used as a proxy state generosity and access to care for adult and child Medicaid beneficiaries. However, it is important to note that the empirical effects of Medicaid fee-for-serve payment generosity on access to care are modest (Shen and Zuckerman 2005). Hill et al. (2009) assessed state Medicaid program efforts to reach out to and enroll pregnant women into coverage. Their survey found that 30 states produce outreach materials in multiple languages in 2007 and I use this as a proxy for state outreach levels among those that speak a non-English language at home.

There are also several hypothetical or ideal data sources that could improve the strength of the empirical strategy described below. First, data at a more detailed geographic level (e.g., zip code or census track) could improve the precision of the estimation equations and provide a clearer picture of an individual's actual neighborhood or community relative to a PUMA. It would also be ideal to have better data related to other potential network definitions (e.g., characteristics of co-workers, church or community membership, etc…) and other health insurance outcomes (e.g., participation in Medicare advantage or private non-group coverage). For example, if the necessary data pieces were available in the Health and Retirement Study, it would be fruitful to test if language-geography groups influence participation in Medicare Advantage plans. This estimation strategy would also be more precise given the fact that 100 percent of the elderly population is eligible for Medicare.

4B. Sample and Sub-Sample Definitions

This dissertation conducts core and expanded sample analyses for adults and children. The core adult sample includes Medicaid-eligible parents and non-parents (determined by FPL thresholds in Figure 3), aged 19 to 64, who speak a language other than English at home and are either covered by Medicaid or are uninsured. The sample also excludes individuals that are part of smaller language groups, defined as those with less than 1,000 individuals in the ACS sample (e.g., Cebuano), in order to have sufficient sample size for PUMA-level contact availability estimates. The results are insensitive to the choice of this cutoff level. The core adult sample includes 59,300 individuals (7,755,281 weighted) among 41 language groups living in 1,877 PUMAs (522 Super-PUMAs and 283 MSAs). The expanded adult sample analysis includes those with private health insurance, (either employer-sponsored or directly-purchased nongroup coverage), increasing the sample size to 83,906 unweighted or 10,786,093 weighted adults.

The core child sample includes Medicaid/CHIP eligible children (determined by FPL thresholds in Figure 3) under 19 who live in a non-English household and are either covered by Medicaid/CHIP or are uninsured. The household language is determined by the language spoken by the child's mother because investments in children's health are made largely by a child's mother (Case and Paxson 2001). The sample also excludes children that are part of smaller language groups. The core sample includes 136,542 children (17,459,492 weighted) among 43 language groups living in 2,044 PUMAs (531 Super-PUMAs and 299 MSAs). The expanded sample includes those with private health insurance, increasing the sample size to 192,414 unweighted or 24,047,763 weighted children.

This study also focuses on several sub-samples as a part of the sensitivity analysis. An obvious concern is that the results could be exclusively driven by the behavior of Spanish-speakers because they comprise the majority of the sample. A simple solution is to exclude Spanish-speakers from the core and expanded samples. Similarly, I also exclude several outlier language groups, such as the Yiddish and Pennsylvania Dutch speakers. I also test how sensitivity the results are to changes in Medicaid/CHIP eligibility definitions e.g., focusing on all adults and children who are under 200% or 300% FPL.

This study also addresses several concerns related to immigration and citizenship status. The core sample includes citizens and non-citizens even though most immigrants are subject to a five-year ban on eligibility and undocumented immigrants are generally ineligible.[14] Legal permanent residents are ineligible for Medicaid/CHIP during their first five years in the U.S. and become eligible afterwards if they meet the programs' other eligibility requirements. However, some immigrants (e.g., refugees and humanitarian immigrants) are exempt from the bar and are eligible for Medicaid/CHIP regardless of their length of residence and 17 states and D.C. have used state funds to provide coverage to recent immigrants who would otherwise be ineligible (Cohen Ross and Marks 2009a; Cohen Ross et al. 2009b). The ACS contains information on citizenship status (but not undocumented vs. documented), years since entry in the U.S., and country of birth. As part of the sub-sample analysis, I exclude non-citizens, focus on the foreign born population (and define networks by country of birth), and focus on recent immigrants (<5 years in the U.S.). The latter two sub-sample analyses are related

[14] Some states use CHIP funds to prove prenatal care to pregnant women, regardless of immigration status. In addition, emergency treatment is available to all immigrants, regardless of status (Kaiser 2006).

to econometric identification concerns and are further discussed in 4.C. I also estimate a model that excludes recent immigrants (<5 years) who live in states where they would be presumably ineligible.

4C. Empirical Strategies

The Identification Problem

Several papers (Manski 1993; Manski 2000; Moffitt 2001; Hartmann et al. 2008) highlight the identification problems associated with network, peer, neighborhood, and social interaction effects studies. Networks are difficult to measure, as few data sets have information on actual contacts, and those that do are typically endogenous because most individuals choose their own contacts. Empirical researchers are therefore challenged to separate the correlations in observed behavior from the true causal effects of one agent (or agents) behavior on another. The "reflection problem", as noted by Manski (1993), occurs because it is difficult to disentangle the direction of causation between average group behavior and behavior of one of its members. The following OLS model highlights the identification problems associated with group behavior models:

$$Mcaid_{ij} = \overline{Mcaid_j} \propto +X_i\beta + \varepsilon_{ij} \qquad (14)$$

Where $Mcaid_{i,j}$ is a 0/1 binary variable indicating Medicaid take-up, $\overline{Mcaid_j}$ is the take-up rate for group language group *j* (or local area *j*), X_i is a vector of individual characteristics, and $\varepsilon_{i,j}$ is the error term. One might naively interpret ∝ as the effect of language group or local area Medicaid take-up behavior on individual take-up behavior. I ran this "naïve" model on the ACS data and, even after controlling for individual and household-level characteristics, the estimated coefficient ∝ was .77 for adults and .96 for children and were statistically significant on at the 1% level. When I defined $\overline{Mcaid_j}$ at the local area level, the estimated coefficient on ∝ was .76 for adults and .85 for children and were statistically significant on at the 1% level. However, this equation is plagued by the "reflection problem" and ∝ captures group behavior effects, what the researcher is interested in measuring, and unobservable effects, what the researcher needs to disentangle.

Correlation between group behavior and individual outcomes might be attributable to correlated unobservables that drive agents in the same reference group to behave similarly. These unobservables might drive *exogenous effects*, where the propensity of an individual to behave in some way varies with the exogenous characteristics of the group, and *correlated effects*, where individuals in the same group tend to behave similarly because they have similar individual characteristics or face similar institutional environments (Manski 2000). Correlated effects can include common levels of education and income or access to health insurance, whereas exogenous effects can include a genetic predisposition to heart disease among a language group that is correlated with the demand for health insurance. These unobservables are controlled for by using language group and local area fixed effects, along with several sensitivity models discussed later in this section.

The endogenous group formation problem is a subset of correlated effects and arises because agents with similar tastes tend to form social groups. As a result, correlation between group and individual behavior may reflect these common tastes, and not a causal effect of one on the other (Hartmann et al. 2008; Moffitt 2001). This problem is self-evident in studies that attempt to measure how peer behavior influences individual outcomes. For example, as mentioned in Chapter 2, Christakis and Fowler (2007) claim that obesity is a contagion that spreads through social networks. Studies like this might make it to the popular press (see "How Friends Make you Fat"[15]), but do a poor job controlling for endogenous group formation. Endogenous group formation is not a major problem in this dissertation because individuals are generally born into language groups and because I am not attempting to measure a direct peer effect; I merely assume that others who are part of a common language group are potential social contacts as opposed to claiming that individual A is friends with individual B, and individual B's behavior is causally influencing individual A's behavior. However,

[15] Rushin, Steve. "How Friends Make you Fat." *Time* August 2, 2007: http://www.time.com/time/magazine/article/0,9171,1649321,00.html

differential geographic sorting within a language group could create an upward bias in the main network variable coefficient. This problem is addressed later in the chapter.

The rest of this chapter describes the empirical strategy and how this dissertation deals with these identification issues.

Estimation Equations for Core and Expanded Samples

This dissertation uses two years of pooled cross-sectional data and a similar empirical strategy as Bertrand et al. (2000) and Deri (2005). The main independent variable in the model varies by each local area-language group combination; the variable is defined by the interaction term between contact availability (the density of each individual's language group in their local area) and the Medicaid take-up rate for each individual's language group. Because this variable is unique to each local area-language group combination, the model can include dummy variables for each local area and language group (fixed effects), which controls for biases associated with omitted local area and language group characteristics, respectively[16].

To reiterate, the interaction term (contact availability*language group take-up) measures the differential effect (between low and high take-up language groups) of

[16] To clarify, I am including dummy variables for each language group and local area. I am not using panel data.

living in areas of high concentration of a common language group (relative to low concentration areas) on an individual's probability of taking-up Medicaid. For an individual that is part of a high Medicaid/CHIP take-up language group (e.g., above the mean), living among a high concentration of his/her language group can increase the person's probability of taking-up Medicaid. In contrast, for those that are part of a low take-up group (e.g., below the mean), living among a high concentration of the language group can decrease the person's probability of taking-up Medicaid relative to living among a low concentration of the language group. These potential contacts might believe that costs of enrollment outweigh the benefits (e.g., it is more convenient to remain uninsured and utilize necessary care from safety net providers), and could discourage the individual from enrolling. It is also possible that living among a high concentration of the language group increases the probability of take-up, regardless if the person is from a low or high take-up group. However, the *differential effect* on the probability of take-up will be larger among those that are part of a higher take-up language group, as these groups might possess more practical knowledge (e.g., information related to eligibility and necessary documentation) that could help the individual enroll in Medicaid.

To measure the effects of language-geography defined groups on Medicaid take-up, I estimate the following OLS model for the adult and child core samples:

$$Mcaid_{ijk} = \tau + Netw_{jk} \propto +CA_{jk}\theta + X_i\beta + \gamma_j + \delta_k + \varepsilon_{ijk} \quad (15)$$

Where

$$Netw_{jk} = CA_{jk} * (Mca\imath d_k - Mca\imath d) \quad (16)$$

$$CA_{jk} = \ln \frac{\%Lang_{jk}}{\%Lang_{UnitedStatesk}} \quad (17)$$

In equation (16), $Mcaid_{ijk}$ is a binary variable equal to one if eligible individual *i*, who lives in local area *j* and is part of language group *k*, takes-up Medicaid and equal to zero if they are uninsured. $Netw_{jk}$ is the main network variable, CA_{jk} (“contact availability") is the direct effect for the quantity of contacts available to individual *i* (this variable also varies over time, but the results are insensitive to this choice), X_i is a vector of individual and household characteristics (including year in survey), γ_j is the local area (PUMA, Super-PUMA, or MSA) fixed effect, and δ_k is the language group fixed effect. The direct effect $\mathrm{Mca\imath d_k}$ drops out of the estimation equation because of the language group fixed effects. For all models, I use robust standard errors clustered at the local area and language group level. A positive estimate of $\propto$ provides evidence in favor of the causal effect of networks on Medicaid take-up.

Equation (17) shows that the main network variable is defined as the interaction between contact availability CA_{jk} and network quality, as measured by the mean Medicaid take-up rate $Mca\imath d_k$for language group *k*. As discussed in Chapter 3,

$Mcaid_k$proxies for the knowledge, information, and attitudes towards Medicaid of others from the language group *k* that live in area *j*. $Mcaid_k$ is calculated for each language group (it does *not* vary by local area) within the sample and is taken in deviation from the sample global take-up rate. The coefficient on the network variable is the same either way, but subtracting the global mean facilitates interpretation of the coefficient on the CA measure base effect. In addition, in order to increase sample size and precision, I estimate the language take-up variable over the combined two-year ACS file. It is also important to note that I do not have enough sample size to define language group take-up at the PUMA-level or the state-level in some cases. Even if there was sufficient sample size to do so, this measure would be endogenous to unobservable differences across local areas; Language group take-up defined at the national level is comparatively more exogenous, however, results are robust to the specification where take-up is defined at the more local level (see Chapter 5). One limitation of this study is that I am unable to identify a potential exogenous source of increased Medicaid take-up. While a couple of states changed their Medicaid eligibility rules between 2008 and 2009, these changes were relatively small and took place in states with relatively small non-English speaking populations.

The numerator for the contact availability measure in equation (18) is the share of the population in area *j* that are part of language group *k* and the denominator is the share of the total United States' population that is part of language group *k*. Contact availability is calculated among all individuals in the ACS file, not just among the Medicaid eligible sample. The denominator in (4) prevents under-weighting of smaller groups; without it, small groups would appear to have very small contact availability because even at full concentration, they would never be a large fraction of any area (Bertrand et al. 2000). However, the results are robust to the specification without the denominator. I use the natural log transformation so that the CA variable has a normal distribution. Otherwise, distribution for the untransformed variable is heavily skewed with a large spike close to 0 and with a long right tail.

I also test for multicollinearity between $CA_{jk} * (Mcaid_k - \overline{Mcaid})$ and the base effect CA_{jk} by calculating condition indices. An informal rule of thumb is that if the condition number is 15, multicollinearity is a concern and if it is greater than 30, multicollinearity is a very serious concern. I find that multicollinearity is not a concern for these variables, as further discussed in Chapter 5, because the condition index is sufficiently low.

In the adult sample, the X_i's include the following individual and household characteristics: Year in sample, gender, age, educational attainment, race and ethnicity, marital status, family structure and size, income relative to poverty, work status, self-employed status, occupation, English fluency, MSA status, citizenship status, number of functional limitations, welfare use, and foreign born status. The child sample includes similar covariates, but educational attainment and work status are defined at the household level.

I estimate a similar OLS model for the expanded sample analysis:

$$Insured_{ijk} = \tau + Netw_{jk} \propto +CA_{jk}\theta + X_i\beta + \gamma_j + \delta_k + \varepsilon_{ijk} \quad (18)$$

Where $Insured_{ijk}$is a binary variable equal to one if the individual has any health insurance and zero otherwise. $Netw_{jk}$ is the same as in (17), but is estimated among the entire expanded sample as opposed to the core sample that excludes those with private health insurance. CA_{jk} is defined in the same manner as (18).

Addressing Identification Concerns

Under an ideal scenario, I would use an experiment to causally identify the effects of language and geography on Medicaid take-up. For example, similar to the Gautreaux experiment, I would randomly assign non-English speakers to neighborhoods with varying levels of own language group contact availability. This type of experiment would identify the causal effects of contact availability (e.g., neighborhood effects), but it would not perfectly identify the effects of language group quality; the true ideal experiment that would randomize individuals into neighborhoods and language groups. These experiments are not feasible for this dissertation.

Alternatively, if the ACS health insurance questions went back further than 2008, I could take advantage of the early Medicaid and SCHIP expansions (e.g., 1997 to 2001 when states were implementing SCHIP) as a natural experiment to identify exogenous participation in public programs among various language groups. Similarly, as part of future research, I can expand on the methods from this paper to examine if there are any language-geography network effects associated with the ACA's Medicaid expansion for adults under 138 percent of the FPL. However, this natural experiment would provide an exogenous shock in Medicaid participation, but it would not exogenously influence an individual's geographic location. A potential natural experiment would be to further explore migration patterns (e.g., natural disaster in country X leads to an exogenous displacement of a population in city Y of country Z) to identify how random shocks in contact availability affect Medicaid take-up.

Given the limitations of the data at hand, empirical identification relies on two assumptions. First, an individual's PUMA residency is exogenous, or at least uncorrelated with the decision to participate in Medicaid. Second, language groups exogenously form (e.g., individuals are born into them) and language group take-up rates serve as a proxy for each language group's knowledge and proclivity towards participating in Medicaid relative to being uninsured. The main strength of this model it includes both language group and local area fixed effects, which is possible because the main network variable is an interaction term. The PUMA dummies control for omitted local area characteristics and unobserved differences between areas, the language group dummies control for omitted language group characteristics and unobserved differences between groups, and the main CA effect controls for other unobserved characteristics e.g., ambition, which may reduce the likelihood of having insurance and the probability of living among one's language group. The rest of this section highlights specific examples identification concerns and discusses how the empirical strategy address each of these issues.

There is potential for differential geographic sorting, where people who live in areas of high density of their language group are different in some unobservable way from people who live in low density areas, but in a way that is correlated with health insurance rates. In other words, there are omitted individual characteristics that are correlated with the key network variable. One solution to differential sorting is to construct the network and CA variables at the larger super-PUMA and MSA levels and use the corresponding local area fixed effects. Comparable estimates between these models and the main estimation model provides evidence that differential sorting is not driving the main results, assuming that MSA or super-PUMA location is exogenous, whereas the exact location within the MSA or super-PUMA (e.g., PUMA) can be a choice variable. If differential sorting is driving the results, the estimates from (19) and (19) are biased upwards relative to super-PUMA and MSA model. I also address sorting issues by comparing the network effects among those who moved in the past year vs. those who stayed in the same home and those who are recent immigrants vs. those who have been in the U.S. for longer periods of time.

A second method to address differential sorting is to include controls for number of years since entry (YSE) into the U.S. and interaction terms between YSE and language group. This model controls for the omitted variable biases associated with immigrant behavior over time. For example, recent immigrants might be more likely to initially locate in high CA areas and might be more homogenous relative to other immigrants. Over time, immigrants start to relocate in a way that is consistent with differential sorting (Deri 2005). Similarly, a regression model limited to recent immigrants (e.g., those who have been in the U.S. for less than two or five years) can avoid the same type of omitted variable bias.

Similarly, comparing the results with $Mcaid_k$defined at the national level versus at the MSA-level provides insight into whether or not language-group take-up is exogenous. In (16) and (19), take-up is defined at the language-group level among the Medicaid/CHIP eligible population. If the results drastically change when $Mcaid_k$ is defined at the MSA-language group or super PUMA-language group level, there might be some concerns that there is selection taking places across MSAs.

There are several other potential alternative explanations of the results that I address. First, individuals might sort based on health status in a way that is correlated with take-up. I address this concern by running separate models for the disabled and non-disabled populations. Another alternative explanation of the results might be due to supply-side behavior of Medicaid offices. A large concentration of a high Medicaid-using language group in an area may lead to CMS to hire more people in that area who speak that language, or to start an advertising or outreach campaign targeted towards specific language groups. Individuals in that language group and area will face lower search costs and might be more likely to take-up Medicaid benefits. This alternative explanation also predicts a positive coefficient on the main network variable. To avoid this problem, I limit the sample to Spanish speakers only and define the CA and network variables based on country of birth.[17] This model assumes that those who are born in the same country are more likely to interact with one another, and the empirical model predicts a positive coefficient on the main network variable. However, the supply-side explanation predicts no effect because everyone in the sample speaks the same language. The only way that supply-side behavior can drive the results is if CMS starts to differentially target individuals based on their country of birth as opposed to language. Another potential solution is to exclude states (e.g., NY and CA) that are known to have extensive Medicaid/CHIP outreach efforts or states that are known to have outreach efforts in multiple languages to see if the coefficient on the main network variable drastically changes.

[17] The country groups for this model include those from Puerto Rico, Spain, Mexico, Belize, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, Panama, Cuba, Dominican Republic, Argentina, Bolivia, Brazil, Chile, Columbia, Ecuador, Guyana, Paraguay, Peru, Uruguay, and Venezuela.

Local Area Characteristics

While the fixed effects models can effectively deal with identification concerns, models (16) and (19) hide how local area characteristics are correlated with the Medicaid take-up decision. The fixed effects model also prevents the researcher from exploring how network strength varies depending on local area characteristics or state policy choices. To provide insight into what is going on behind the scenes, I remove the PUMA fixed effect γ_j and replace it with a vector of PUMA characteristics. These characteristics include average age, the percent of the PUMA that is non-white, the percent of the PUMA living in poverty, and the percent of the PUMA that is foreign born. I also calculate the condition index associated with these variables to determine if multicollinearity is a concern. Given that the condition index is close to 50, I run the model with one characteristic at a time.

I also test if network strength varies by type of neighborhood by dividing the sample into quintiles based on PUMA characteristics. For example, I explore if network effects are stronger in lower-income or geographically smaller neighborhoods. I also use the state-level Medicaid fee index and language outreach dummy and interact each with the network variable to test if networks are stronger in states that have less generous Medicaid programs or weaker outreach efforts, respectively.

Model Specification Checks and Other Sensitivity Tests

I also check to see if the results differ when using non-linear binary models (logit, probit). For the expanded sample analysis, I use a multinomial logit similar to (19) that includes private health insurance as an additional choice outcome. However, there are two major disadvantages with non-linear models. First, it is practically difficult to estimate nonlinear models with fixed effects (they do not merge in Stata) and methodologically, the incidental parameters problem raises questions about the statistical properties of the estimators (Greene 2002). Second, while the multionomial logit might be the more theoretically “correct” model, interpreting an interaction term with two continuous variables is not a straightforward process (See Ai and Norton 2003). In Chapter 5, I compare the OLS results with various non-linear models and find that the results are consistent and comparable.

I also explored potential private health insurance network effects by adding an additional network variable to the expanded sample in equation (19). However, given the theory in Chapter 3 and the fact that my sample focuses on the Medicaid-eligible population, there is no a priori reason to believe that private health insurance defined networks will have an impact on the probability that an individual is insured. The majority of the privately insured obtain coverage from their employers; networks can either indirectly influence this choice through labor market decisions or directly through the choice of plan type once already employed, as shown in Sorenson (2003). However, I do not have the data to explore these options and the private insurance network variable may suffer from multiple omitted variable biases, such as employer-offer status. In addition, some individuals may rely on networks to obtain information on health insurance products in the nongroup market. However, coverage rates in the private nongroup are relatively small, especially among the low-income population. Overall, I find no statistically significant effects associated with the private health insurance network variable (results not shown).

Interpretation

This section describes how to interpret the coefficients obtained in equation (16).[18] This is not a straightforward process because the key variable of interest is the interaction between two continuous variables. Using the same interpretation used in Bertrand et al. (2000) and Deri (2005), τ can be viewed as a policy such that a 1-percentage point increase in τ leads to a 1-percentage point increase in Medicaid use in the absence of networks. Changes in policy lead to a direct effect on Medicaid (τ) and an indirect effect via networks. Intuitively, an increase in the policy variable τ raises

[18] A parallel explanation can be used to interpret equation (19).

$Mcaid_k$, which in turn raises each individual's Medicaid probability through a feedback or network effect. Examples include a Medicaid eligibility expansion or a targeted advertising/outreach campaign designed to increased Medicaid take-up.

Mathematically, averaging both sides (2) by language group *k* and differentiating with respect to τ yields:

$$\frac{dMcaid_k}{d\tau} = 1 + CA_k * \frac{dMcaid_k}{d\tau}\alpha \tag{19}$$

Solving for the change in Medicaid use for each language group for a policy change t and subtracting 1 for the direct effect yields:

$$\frac{dMcaid_k}{d\tau} = \frac{1}{(1 - \alpha CA_k)} \tag{20}$$

Overall, the multiplier effect is stronger (positive or negative) for higher average levels of contact availability and higher coefficients on the network variable. If the coefficient on the network variable is negative, higher contact availability has as stronger negative effect on an individual's probability of take-up, whereas if the coefficient on the network variable is positive, higher contact availability has as stronger positive effect on an individual's probability of take-up. The regression result tables in the next chapter include policy multiplier estimates for each OLS model and sub-sample.

The following chapter contains the results for all of the models described in this chapter and interprets the main variable of interest. There are also a few additional sensitivity tests that were not mentioned in this chapter.

Figure 1: American Community Survey Question on Health Insurance

Is this person CURRENTLY covered by any of the following types of health insurance or health coverage plans? *Mark "Yes" or "No" for EACH type of coverage in items a – h.*

a. Insurance through a current or former employer or union (of this person or another family member)
b. Insurance purchased directly from an insurance company (by this person or another family member)
c. Medicare, for people 65 and older, or people with certain disabilities
d. Medicaid, Medical Assistance, or any kind of government-assistance plan for those with low incomes or a disability
e. TRICARE or other military health care
f. VA (including those who have ever used or enrolled for VA health care)
g. Indian Health Service
h. Any other type of health insurance or health coverage plan – *Specify*

Source: U.S. Census Bureau, American Community Survey, 2008-2009.

Figure 2: Details of the IPUMS Health Insurance Edits

*Persons who did not claim to be covered by Medicaid were assigned Medicaid coverage if they were:

1. Less than 19 years old and the unmarried child of a parent with public assistance and/or Medicaid;
2. A citizen parent with public assistance;
3. A citizen parent married to a citizen with public assistance and/or Medicaid;
4. A foster child; or
5. A Supplemental Security Income (SSI) enrollee living in a state where SSI enrollees are automatically enrolled in Medicaid and who satisfies one of the following three additional conditions:

    Does not have children

    Has children but is disabled and/or not working

    Group quarters resident.

*Persons who did not claim to be covered by Medicare were assigned Medicare coverage if they were at least 65 years old and satisfied at least one of the following conditions:

1. Reported Social Security or Railroad Retirement Benefits
2. Reported Medicaid coverage.

*Persons who did not claim to be covered by TRICARE or other military insurance were assigned such coverage if they were:

1. Active duty military;
2. The spouse of an active duty military person and did not report other private coverage; or
3. Less than 21 years old, lacking in other private coverage, and the unmarried child of an active duty military person.

*Persons who gave direct reports (i.e., unallocated) of employer-based, privately purchased, military, Medicaid, and Medicare coverage had:

*VA coverage changed to "No" if the person was not a veteran; and

*IHS coverage changed to "No" if the person did not identify American Indian / Alaska Native as their only race.

Notes:

(1) IPUMS-USA, Health Insurance Variables in the American Community Survey, http://usa.ipums.org/usa/acs_healthins.shtml. Accessed on February 23rd, 2011.

(2) Lynch V, Bourdreaux m, and Davern M. Applying and Evaluating Logical Coverage Edits to Health Insurance Coverage in the American Community Survey. Suitland (MD): U.S. Census Bureau, July 2010.

(3) Unless otherwise noted, "parent" refers to a person with a child under age 18.

Figure 3

2008-2009 Medicaid Income Limits as a Percent of FPL

	Parents	**Childless Adults**	**Children (2008/2009)**
Alabama	24%	N/A	200%
Alaska	81%	N/A	175%
Arizona	106%	110%	200%
Arkansas	200%	200%	200%
California	106%	N/A	250%
Colorado	66%	N/A	205%
Connecticut	300%	300%	300%
Delaware	121%	110%	200%
DC	207%	211%	300%
Florida	53%	N/A	200%
Georgia	50%	N/A	235%
Hawaii	200%	200%	300%
Idaho	185%	185%	185%
Illinois	185%	N/A	200%
Indiana	200%	200%	250%
Iowa	250%	250%	200%/300%
Kansas	32%	N/A	241%
Kentucky	62%	N/A	200%
Louisiana	25%	N/A	250%
Maine	300%	300%	200%
Maryland	116%	116%	300%
Massachusetts	300%	300%	300%
Michigan	64%	45%	200%
Minnesota	275%	250%	280%
Mississippi	44%	N/A	200%
Missouri	25%	N/A	300%
Montana	56%	N/A	175%
Nebraska	58%	N/A	185%
Nevada	200%	N/A	200%
New Hampshire	49%	N/A	300%
New Jersey	200%	N/A	350%
New Mexico	250%	250%	235%
New York	150%	100%	400%

North Carolina	49%	N/A	200%
North Dakota	59%	N/A	160%
Ohio	90%	N/A	200%
Oklahoma	200%	200%	185%
Oregon	185%	185%	300%
Pennsylvania	208%	213%	300%
Rhode	181%	N/A	250%
South Carolina	89%	N/A	200%
South Dakota	52%	N/A	200%
Tennessee	129%	129%	250%
Texas	26%	N/A	200%
Utah	150%	150%	200%
Vermont	300%	300%	300%
Virginia	29%	N/A	200%
Washington	200%	200%	250%/300%
West Virginia	33%	N/A	220%/250%
Wisconsin	200%	200%	300%
Wyoming	52%	N/A	200%

Sources: (1) Kaiser Family Foundation, statehealthfacts.org (2) Cohen Ross, Jarlenski, Artiga, and Marks (2009)
Note: Thresholds are the maximum among Medicaid or Medicaid look-alike programs, programs more limited than Medicaid, and premium assistance with work-related eligibility requirements.

Figure 4
State-Level Data for Sensitivity Analyses

State	Medicaid Fee Indexes for All Services (2008)	Has Outreach in Multiple Languages (2007)?
US	1.00	
Alabama	1.10	No
Alaska	2.05	Yes
Arizona	1.45	Yes
Arkansas	1.10	Yes
California	0.83	Yes
Colorado	1.19	No
Connecticut	1.44	Yes
Delaware	1.44	No
District of Columbia	0.87	Yes
Florida	0.89	Yes
Georgia	1.21	Yes
Hawaii	1.04	No
Idaho	1.33	No
Illinois	0.90	Yes
Indiana	0.90	No
Iowa	1.22	No
Kansas	1.20	Yes
Kentucky	1.10	No
Louisiana	1.24	Yes
Maine	0.81	Yes
Maryland	1.27	Yes
Massachusetts	1.30	Yes
Michigan	0.90	Yes
Minnesota	0.98	No
Mississippi	1.14	No
Missouri	0.94	No
Montana	1.33	No

Nebraska	1.24	Yes
Nevada	1.46	Yes
New Hampshire	0.98	No
New Jersey	0.58	No
New Mexico	1.42	Yes
New York	0.62	No
North Carolina	1.27	Yes
North Dakota	1.30	Yes
Ohio	0.94	Yes
Oklahoma	1.28	No
Oregon	1.18	No
Pennsylvania	0.98	Yes
Rhode Island	0.59	Yes
South Carolina	1.24	No
South Dakota	1.19	Yes
Tennessee	N/A	Yes
Texas	1.01	Yes
Utah	1.08	Yes
Vermont	1.25	Yes
Virginia	1.23	No
Washington	1.28	Yes
West Virginia	1.12	No
Wisconsin	1.07	Yes
Wyoming	1.81	No

Source: Zuckerman et al. (2009); Hill et al. (2009)

Chapter 5: Results

5A. Descriptive Statistics

Table 1A shows weighted descriptive statistics for the core and expanded adult samples. Overall, there are 59,377 unweighted adults (7,755,281 weighted) in the core sample and 83,906 unweighted adults (10,786,093) in the expanded sample. Overall, a non-trivial proportion of the sample has access to private health insurance: 36.7% of the core sample has Medicaid and 63.3% is uninsured, whereas 26.4% of the expanded sample has Medicaid, 28.1% has any private coverage, and 45.5% is uninsured. Relative to the core sample, the expanded sample includes a higher proportion of adults with at least some college education (30.5% vs. 23.6%), a lower proportion of Hispanics (65.4% vs. 71.4%), and a lower proportion of individuals with family income below the poverty rate (61.8% vs. 68.7%).

Table 1B is the child equivalent to Table 1A. The child sample is substantially larger than the adult sample because Medicaid/CHIP eligibility rules are more generous for children. There are 136,542 unweighted children (17,459,492 weighted) in the core sample and 192,414 unweighted children (24,047,763) in the expanded sample. Overall, children have higher rates of Medicaid coverage compared to the adult samples: 71.6% of the core child sample has Medicaid and 28.4% is uninsured, and 52% of the expanded child sample has Medicaid, 27.4% has any private coverage, and 20.6% is uninsured. In addition, consistent with the statistics in Table 1A, the expanded child sample has higher levels of education (defined at the family-level) and income compared to the core child sample.

Appendix 1A and 1B include the same descriptive statistics for the adult and child expanded samples, but broken out type of health insurance coverage. There are considerable differences in observable characteristics among the different coverage categories. One interesting thing to note is that those with private non-group coverage look more similar, on average, to those with employer-sponsored insurance. However, Lynch et al. (2010) finds that there might be considerable measurement error in private non-group coverage e.g., respondent confusion between Medicaid and private-nongroup, which is supported by the fact some individuals report having non-group coverage even though it is most likely "unaffordable" given their low income levels. These tables provide some insight into the differences between the core and expanded sample and help explain some of differences in the regression results discussed in the next section.

As mentioned in the previous chapter, this study includes both citizens and non-citizens in the sample (non-citizens are dropped in a sensitivity analysis), even though most immigrants are subject to a five-year ban on eligibility and undocumented immigrants are generally ineligible. Legal permanent residents are ineligible for Medicaid/CHIP during their first five years in the U.S. and become eligible afterwards if they meet the programs' other eligibility requirements. However, some immigrants (e.g., refugees and humanitarian immigrants) are exempt from the bar and are eligible for Medicaid/CHIP regardless of their length of residence and 17 states and D.C. have used state funds to provide coverage to recent immigrants who would otherwise be ineligible. The decision to include non-citizens had a larger impact on the adult sample sizes compared to the child samples: 56.1% of adults in the core sample are non-citizens (52.2% in the expanded sample) and only 14.1% of children in the core sample are non-citizens (12.8% in the expanded sample). As a result of this generous definition of eligibility, the sample provides an upper-bound on the number of individuals who are eligible for Medicaid by including some individuals who are technically not eligible due to their immigration status.

Table 2 compares Medicaid coverage rates between citizens and non-citizens in the adult and child core samples. As expected, given the eligibility rules, citizens have higher Medicaid coverage rates than non-citizens. 23.6% of non-citizen adults and 39% of non-citizen children have Medicaid compared to 53.3% and 76.9% of adult and child citizens, respectively. There are only small differences Medicaid coverage rates among non-citizen immigrants who have been in the country for less than 5 years (those that should have a higher proportion ineligible for Medicaid) compared to those who have been in the U.S. for 5 to 10 years. However, there is a spike in Medicaid coverage among adults who have been in the U.S. for 5 years compared to those who have been in the U.S. for only 4 years, which is consistent with the eligibility rules described in the prior paragraph. Figure 1 shows a plot of this data and bootstrap results show that there is a statistically significant discontinuity at the 5 year mark in the U.S. (coefficient=.0468, bootstrap standard error=.0015, p-value=.002). These results can be further explored in future research.

Tables 3A and 3B highlight the differences across language groups in the core adult and child samples, respectively.[19] Medicaid take-up for an individual's language group as a whole, interacted with the contact availability, is the key independent variable of interest in this paper. As such, it is necessary to have sufficient variation in Medicaid take-up across language groups to identify network heterogeneity; variation in take-up across language groups proxies for the inherent differences in preferences and attitudes towards health insurance across language groups, as determined by culture, experiences with government insurance in one's native country, etc…In both the adult and child samples, Yiddish speakers have the highest (over 90%) and Pennsylvania Dutch have the lowest (under 10%) Medicaid take-up rates among all language groups. As discussed in the next section, the results are insensitive to including or excluding these groups. Table 3A shows that Medicaid take-up rates among the other language in the adult core sample vary widely, ranging from 74% among Cantonese, 68% among Armenian, and 67% among Bengali speakers to 22% among Korean, 21% among

[19] It is important to note that some language groups have small sample size (e.g., Hungarian), but the language group as a whole has at least 1,000 unweighted individuals in the ACS during each data year.

Japanese, and 14% among Dutch speakers. The Medicaid take-up among Spanish speakers, who comprise 73% of the sample, is near the low end at 33%. Table 3B also shows that Medicaid/CHIP take-up rates in the child core sample (where the language group is defined by the language that the child's mother speaks at home) vary widely, ranging from 93% among Hungarian, 90% among Miao/Hmong, and 89% among Hebrew speakers to 48% among Korean, 39% among German, 18% among Dutch speakers. Spanish speakers comprise 81% of the core child sample and have a take-up rate of 72%.

Appendix 2A and Appendix 2B contain the same information for the expanded adult and child samples. These tables contain an additional column indicating the rate of private health insurance for each language group. With a few exceptions, the ordering of language groups by proportion covered by Medicaid is consistent with the ordering in Tables 2A and 2B. However, there is considerable variation in private health insurance rates across language groups. For example, 69% of adult Japanese speakers have private health insurance compared to 23% of Spanish speakers and 14% of Navajo speakers. Similar trends prevail in the expanded child sample.

The regression models in the next section handle the privately insured in the expanded sample in two ways. First, I estimate the same OLS model used for the core sample, but change the dependent variable to "any insurance" as opposed to Medicaid take-up. This model determines the effect of Medicaid network variable on the probability of obtaining any insurance type, which can include Medicaid, employer-sponsored insurance (ESI), and private non-group. However, given the theoretical mechanisms described in Chapter 3 and the results shown among the core sample, any measurable effect of networks on "any insurance" should primarily be attributable to the Medicaid take-up effect. I verify this by estimating a more theoretically sound model (multinomial logit) that explicitly incorporates the choice of any private health insurance relative to the other insurance outcomes.[20] The tradeoffs associated with using the multinomial logit versus OLS are discussed in Chapter 4.

Contact availability, defined by equation (18) in Chapter 4, is the other key variable that comprises the network effect.[21] Tables 4A and 4B provide the mean, standard deviation, minimum, and maximum of various specifications of the contact availability variable. The main specification, which has a log transformation and adjusts for the under-weighting of small language groups, is in the first three rows in each table. Average contact availability defined at the PUMA-level has a higher mean and standard deviation compared to contact availability defined at the Super-PUMA and MSA level, implying that non-English speakers are more densely populated in smaller geographic areas. The bottom three rows of 4A and 4B contain the sensitivity specifications of contact availability. By removing the denominator, contact availability approaches zero for more individuals in the sample, and hence the average natural log of contact

[20] I also estimate models with any private is separated into ESI and private non-group.

[21] As a reminder: The numerator for the contact availability measure in equation (4) is the share of the population in area *j* that are part of language group *k* and the denominator is the share of the total United States' population that is part of language group *k.* This variable is then log transformed.

availability becomes negative. By removing the log transformation, contact availability becomes heavily skewed and has substantially larger range and standard deviation.

Finally, Table 5, along with Figure 2 and Figure 3, provide a more detailed picture on where adults and children in the sample live and the role of contact availability. Table 5 shows the state distribution of the core adult and child samples. More than half of Medicaid eligible adults that speak a non-English language at home live in 5 states: California (18.0%), New York (16.3), Arizona (6.0%), Massachusetts (6.0%), and Illinois (5.4%). In addition, more than half of Medicaid/CHIP eligible children that speak a non-English language at home live in 3 states: California (26.7%), Texas (18.4%), and New York (8.4%). Given the fact that the sample size in many states (e.g., Wyoming, West Virginia, South Dakota, North Dakota, etc…) is extremely low, Medicaid take-up for each language group is defined at the national level as opposed to the state or local level. Even without the sample size constraint, defining language group take-up at a local area level would most likely capture unobservable differences across areas as opposed to preferences and information that the language group as a whole possesses.

Figure 2 and Figure 3 show heat maps of the proportion of PUMAs that speak Spanish in the household for California and New York, respectively[22]. These maps provide some insight into how contact availability works: for each individual in a given PUMA, contact availability is a function of the proportion of the PUMA population that

[22] I would like to Thank Michael Huntress for his assistance in producing these maps.

speaks their common language. These maps provide a good visual snapshot because of the large share of the sample that speaks Spanish and live in New York or California. In both maps, the cutoffs for each color code correspond to the quartile distribution of PUMAs. For example, a quarter of PUMAs in California have between 36.6% and 81.7% of the population speaking Spanish in the household. There are two major common trends in these maps. First, there are high concentrations of Spanish speakers in Urban centers, such as New York City and Los Angeles. Most of the PUMAs outside of New York City have a relatively low proportion of Spanish speakers, with more urban areas such as Rochester and Syracuse being the exception. In contrast, California as a whole is more heavily concentrated with Spanish speakers. Most of the PUMAs in the lowest quartile are in sparsely populated areas, such as mountainous PUMAs in the north or the Mojave Desert towards the east. Second, there is considerable variation across the PUMAs within a given city. For example, within New York City, there are 3 PUMAs in Staten Island that have different proportions of Spanish speakers. There is also considerable variation between the Bronx (high proportion of Spanish speakers throughout) and Manhattan (mixed proportions).

5B. Regression Results

Illustrative Example

The main variable of interest is not very intuitive to interpret because it is an interaction between two continuous variables. An easier way to understand how the variable works is to create two binary variables (low vs. high take-up group and low vs. high contact availability) and divide the sample into four groups: Individuals from low and high Medicaid take-up language groups living in low and high contact availability areas, where the low/high cutoff for each variable is determined by the mean level. Using these binary variables instead of the continuous variables, one can view the results under the framework of a difference-in-differences (DD) model (this example uses unconditional means, but the results for regression-adjusted means are consistent). The goal of this exercise is to show the differential effect of living in high contact availability areas for individuals from low and high Medicaid take-up language groups or in other words, determine if living in a high contact availability area increases the probability of Medicaid take-up more for individuals that are part of higher Medicaid take-up language groups. Table 6 shows the DD estimates of unconditional means for the adult and child core samples. For individuals in the adult sample that are part of a low Medicaid take-up language group, the differential effect on Medicaid take-up of living in a high contact availability area relative to a low contact availability area is .070. However, the differential effect (.165) is higher for adults that are part of a high Medicaid take-up language group. The DD estimates for children are more succinct: living in a high contact availability area for those that are part of low take-up language groups has a *negative* effect (-.0580) on the individual's probability of taking up Medicaid, whereas living in a high contact availability area for those that are part of high take-up language groups has a positive effect (.0582) on an individual's probability of taking-up Medicaid. The overall DD estimate is .116 and is statistically significant at the 1% level.

Main Results

Table 7A and Table 7B compare the full regression results from the core and expanded sample OLS models. I estimated robust standard errors for all OLS models, although they are not shown in these two tables for sake of space. The core and expanded samples in Tables 7A/7B include PUMA and language group fixed effects, contact availability is defined at the PUMA level, Medicaid take-up at the language group level is defined among those in the sample, and robust standard errors are cluster corrected by PUMA and language group. The dependent variable in the "naïve" models and the core sample is a 0/1 indicator for Medicaid take-up, whereas the dependent variable in the expanded sample is a 0/1 indicator for having any health insurance type. Language group, PUMA, and occupation dummies are not shown, but are available upon request. The coefficients and significance levels for all other covariates are displayed in Tables 7A/7B.

Before discussing the main variable of interest, it is important to note that the covariates in Tables 7A and 7B have the signs that we would expect to see among the samples a priori. For both adults and children, being a non-citizen decreases the probability of taking-up Medicaid or having any health insurance, while being female increases the probability of take-up. Both of these results could be partially explained by both exogenous factors (e.g., the relationship between Medicaid eligibility and citizenship status and pregnancy status) and endogenous factors (e.g., non-citizens lack access or information related to coverage and women in the sample have a higher expected benefits associated with insurance). In addition, older adults and younger children are more likely to have Medicaid or any coverage relative to their counterparts and Hispanics are less likely to have Medicaid or any coverage relative to whites. English fluency, welfare use, and number of disabilities also has a positive impact on the probability take-up and having coverage. Another interesting pattern is that in the core sample, adults with higher levels of education and income are less likely to take-up Medicaid and are more likely to remain uninsured. In contrast, the signs flip in the expanded sample and those with higher levels of education and income are more likely to have any insurance type. This is due to the fact that the expanded sample includes those with private health insurance, which is positively correlated with higher levels of education and income.

The two columns in Tables 7A/7B display the main coefficients of interest and correspond to equations (16) and (19) in Chapter 4 for the core and expanded samples, respectively. For adults, the coefficient on the main network variable (take-up rate of language group*contact availability) is .100 for the core sample and .118 in the expanded sample, and for kids, the network coefficients are .071 and .100. All four coefficients are statistically significant at the 1% level and provide strong evidence for the existence of language and geography-defined network effects. By including neighborhood and language group fixed effects, these models address any biases associated with omitted neighborhood characteristics and omitted language group characteristics that could potentially be correlated with the network variable. In addition, the base effect for contact availability controls for any unobservable individual characteristics that could be correlated with network size.

These results imply that for a policy change that increases Medicaid take-up by 1 percentage point, the network will increase the probability of taking-up Medicaid in these language groups by 9.9 percentage points for adults and 7.4 percentage points for kids. In the expanded sample, for a policy change that increases Medicaid take-up by 1 percentage point, the network will increase the probability of having any health insurance in these language groups by 11.0 percentage points for adults and 10.4 percentage points for kids. The remainder of this chapter addresses specific identification issues, potential mechanisms associated language-geography networks, and various sensitivity analyses.

It is important to note that when I used the child's language spoken at home to define networks, I found relatively weak evidence of network effects among Medicaid/CHIP eligible children. There was also a much smaller sample as children are more likely to speaker English at home (results not shown). However, Table 7B and the other tables in this section show that the results are greatly strengthened when the mother's language is used to define networks.

Addressing Differential Selection and Identifying Network Mechanisms

Chapter 4 described how this there could be potential for differential geographic sorting, where people who live in areas of high density of their language group are different in some unobservable way from people who live in low density areas, but in a way that is correlated with health insurance rates. In other words, there could still be omitted individual characteristics that are correlated with the key network variable. In preliminary results for this paper, I addressed this problem by using the network and contact availability variables constructed at the larger MSA levels as instrumental variables (IVs) for the same variables constructed at the more detailed PUMA level.[23] A necessary condition for this model is that the MSA IVs need to be highly correlated with the PUMA-level variables. In addition, the IV model must assume that MSA location is exogenous, whereas exact location within the MSA (e.g., PUMA) can be a choice variable that can be biased due to differential sorting. If differential sorting is driving the results, the OLS estimates are biased upwards relative to the IV estimates. In contrast, comparable OLS and IV estimates provide evidence that differential sorting is not driving the main results. The null hypothesis that these two instruments are jointly zero in the first stage was easily rejected: In preliminary core model results, the joint F-statistic was 56,794 (p-value=0.000) for the network variable first stage regression and 43,045 (p-value=0.000) for the contact availability variable first stage regression. In addition, the magnitude and statistical significance of the network IV was nearly identical to the main OLS result.

The first three columns of Table 8A (adults) and Table 8B (children) compare the core and expanded models defined with PUMA variables[24], Super-PUMA variables, and MSA variables. Given the preliminary results discussed above, comparing the coefficients of these models, without using the IV approach, provides a comparable

[23] This method was used by Bertrand et al. (2000) and Deri (2005). Other network papers use similar approaches.
[24] These are the same core and expanded model results displayed in 7A and 7B.

level of precision.[25] The Super-PUMA model (model 2) includes Super-PUMA fixed effects as opposed to PUMA fixed effects and defines contact availability within the larger Super-PUMA as opposed to the PUMA area. Model 3 does the same thing at the MSA-level, but excludes individuals who live in non-MSAs and do not have MSA identifiers. The coefficients across all three models are similar (within approximately 1 standard error of the average) and for the most part, the estimated coefficients in the Super-PUMA and MSA models are slightly higher than the coefficients in the PUMA model. However, the policy multiplier effect is either the same or slightly larger in the PUMA models because contact availability levels is larger in the PUMA compared to the super-PUMA and MSA. If differential geographic selection were driving the results, we would expect to see substantially stronger network effects in the PUMA model, as individuals would select across PUMAs within a given super-PUMA or MSA in the manner described above. Given the fact that these estimates are so similar, it appears that differential selection cannot the main driving force behind the main results.

The remaining models in Table 8A provide additional evidence against differential sorting and show that recent immigrants are more likely to rely on language group networks. Models 4 and 5 limit the sample to foreign born adults and model 6 only includes the non-foreign born population. Model 5 also includes years since entry in the U.S. (YSE) dummies and interactions terms between YSE and language group. This

---

[25] I would like to thank Jeremy Tobacman for this suggestion.

model provides additional controls for the omitted variable biases associated with immigrant behavior e.g., over time, immigrants start to relocate in a way that is consistent with differential sorting. In both the core and expanded adult samples, the network coefficient is positive and significant at the 1% level among the foreign born population and is statistically insignificant among the U.S. born population. To further investigate, I separated the foreign born population by number of years in the U.S.. Models 7, 8, and 9 show that the network effect is substantially stronger among recent immigrants compared to those who have been in the country for five or more years. In addition, the results show stronger network effects that those who have been in the U.S. for two years or less compared to those who have been in the U.S. for five years or less. This key result provides not only provides evidence against differential sorting, but also shows how recent immigrants are more likely to rely on networks to receive information related to the language group's valuation or preferences towards Medicaid.

The foreign born results for the child sample (Table 8B) are a bit misleading. The results imply that there are stronger network effects among children born in the U.S. as opposed to foreign born children in the core sample, whereas the network effects are stronger among foreign born and recent immigrant children in the expanded sample. It appears that this reversion across samples is due to the inclusion of those with private non-group insurance coverage: Language-defined Medicaid networks have a positive effect on having non-group coverage among the foreign born population in the expanded sample, which in turn has a positive impact on the probability of have any health insurance coverage. This result can be attributable to two factors. First, there is considerable measurement error associated with the private non-group health insurance variable and parents confuse private non-group with a state Medicaid/CHIP plan. Second, it is possible that those with private non-group in the expanded child sample are actually covered by non-group and are technically ineligible for Medicaid due to their immigration status. Networks inform these individuals on the value of being insured, but since they are ineligible for public insurance, the newly acquired information increases the individual's probability of seeking out health insurance through the private non-group market. This is further explored in Table 19B with the multinomial logit model, and further research is needed to understand the underlying measurement error or behavioral mechanism.

The ACS also contains information on whether or not the person moved in the past year. I find that adults who moved locations in the past year have comparable network effects as those who lived in the same house, but children who moved in the past year had substantially larger network effects than those who stayed in the same house. However, the network effects remain positive and significant at the 1% level, even after removing the "mover" population (whom account for approximately 1/6 of the sample). A closer look at the data also shows that children who moved in the past year were more likely to be in poverty and more likely to be foreign born compared to those who lived in the same house (data not shown).

Tables 9A and 9B explore how network effects vary across levels of English fluency and health status. In 9A, I find that network effects are slightly stronger for adults in linguistically isolated households compared to individuals that are not in linguistically isolated households.[26] In contrast, among children (9B), the network coefficient is statistically significant in households that are not linguistically isolated, but insignificant among linguistically isolated households. There are also no network effects associated with adults and children that live in group quarters (e.g., those in institutions or non-institution group quarters, but not living in a household), which provides internal validation to the model because most of these individuals have little contact with others that are part of their language group. In addition, models 5 through 8 in Table 9A show that network effects are stronger among adults that are fluent in English relative to those that do not speak English or do not speak English well. This result is inconsistent with the finding in Bertrand et al. (2000), where the authors found that networks effects are weaker for people speaking better English in the context of welfare participation. The result from this study could be explained by the fact those who speak a non-English language at home, but are also fluent in English, have a more information and a

---

[26] "Linguistically isolated households" are households in which either no person age 14+ speaks only English at home, or no person age 14+ who speaks a language other than English at home speaks English "Very well" (Ruggles et al. 2010).

thorough understanding of complexities associated with Medicaid eligibility and the take-up process relative to those who are not fluent in English. In contrast, the welfare participation decision is relatively more simplistic compared to health insurance decision-making. The results in the child sample, where English fluency is defined at the mother's level, are fairly consistent with those in Table 9A.

Finally, the last two columns in 9A and 9B study whether networks are more important for individuals with a disability (respondent has serious difficulty hearing, seeing, concentrating/remembering/making decisions, walking or climbing stairs, dressing or bathing, or doing errands) compared to those without a disability. I find that network effects are non-existent among the disabled population, which provides evidence against the alternative hypothesis that individuals sort by health status in a manner which is correlated with language-geography networks.

Sample Sensitivity

The results in Table 10A and Table 10B show that the main results are relatively insensitive to various sample definitions. These tables include 9 models among the child (Table 10A) and adult (Table 10B) core and expanded samples. All of these models include PUMA and language group fixed effects and define contact availability at the PUMA-level. MSA and Super-PUMA model results are consistent and are available upon request.   All of the coefficients associated with the network variable in these models (with the sole exception being Model 5 for children) remain statistically significant at the 1% level and have robust magnitudes:

- Model 1: Main model (full results in 7A and 7B)
- Model 2 and Model 3: Sensitivity analysis over defining Medicaid/CHIP eligibility. Model 2 defines individuals as Medicaid/CHIP eligible if their family income is below 200% FPL and Model 3 defines individuals as Medicaid/CHIP eligible if their family income is below 300% FPL. These eligibility definitions are less precise, but the network effects are similar, but slightly stronger, compared to the main model.
- Model 4: Exclude Spanish speakers. All network effects are statistically significant and magnitudes are strong. However, the coefficient on the network variable for non-Spanish children in the core sample is only significant at the 10% level.
- Model 5: Spanish speakers only and network defined by country of birth. This model addresses the concern that supply-side forces (e.g., differential outreach across PUMAs) are driving the main results. Assuming that CMS does not differentially target individuals based on country of birth as opposed to language group, this alternative explanation cannot explain the results from this model. The results are positive and significant at the 1% level among adults, but are

statistically insignificant among the child sample, where country of birth is defined by the mother's country of birth. Further work is needed to explore the latter result, but this result could be explained by the fact that mother's country of birth is a weaker and more arbitrary network definition compared to language spoken at home in terms of obtaining information related to health insurance.

- Model 6: Exclude non-citizens. The results are insensitive to including or excluding non-citizens. I also found that that the coefficient on the network variable remains positive and statistically significant when immigrants who have been in the U.S. for less than 5 years, and live in states that do not have extended eligibility rules for this population, are excluded.
- Model 7 and Model 8: Exclude Yiddish and Pennsylvania Dutch speakers (potential outlier language groups). The network effects are slightly stronger after making these sample restrictions.
- Model 9: Exclude California and New York. As discussed earlier in this chapter, California and New York are two of the largest states in terms of sample size and could potentially devote more resources to outreach efforts. The results remain statistically significant at the 1% level after removing individuals who live in these states. However, the policy multiplier among the adult sample is slightly lower due to lower levels of contact availability in the states besides California and New York.

Local Area Characteristics

Table 11A and Table 11B compare the main PUMA fixed effects model with various models that replace fixed effects with PUMA characteristics. I use the percentage of the PUMA that is non-white, the percent of the PUMA under 100% FPL, the average age in the PUMA, and the percent of the PUMA that is foreign born as neighborhood characteristics. In column 2, I include all four characteristics in the model and find that the network variable remains positive and significant at the 1% level. In the child model (11B), I find that the network coefficient in this model is stronger in magnitude compared to the fixed effects model, which indicates that the fixed effects are capturing some unobservable differences across PUMAs. However, tests indicate that there is significant multicollinearity when all four characteristics are included in the model: The collinearity condition number for the four PUMA variables is 46.8 in the adult sample and 44.9 in the child sample. As an alternative, I add one variable at a time and find that the coefficient on the network variable remains consistent throughout. In addition, the coefficients on the PUMA characteristic variables imply that all four characteristics are positively correlated with an individual's probability of taking-up Medicaid. Tables 12 through 14 explore the strength of networks across various local area (PUMA-defined) characteristics.

Tables 12A and 12B analyze how PUMA-level poverty levels relate to the strength of networks. I divide both the core and expanded adult samples for adults (12A) and children (12B) into quintiles based on the average poverty ratio in the PUMA, where the 1^{st} quintile includes individuals in the lowest income PUMAs and the 5^{th} quintile includes those living in the highest income PUMAs. For adults, I find that network effects, based on the policy multiplier, are stronger lower income neighborhoods. I also find no network effect in the 5th quintile group in the core sample. For children, the trend is inconsistent in the core sample, but the policy multiplier effects are stronger in the lower PUMA income groups.

Tables 13A and 13B provide some evidence that network effects are stronger in smaller geographic PUMAs. I divided the sample into five quintiles based on total land area of the PUMA in square meters, where the 1^{st} quintile includes individuals in the smallest PUMAs and the 5^{th} quintile includes individuals in the largest PUMAs. A priori, one would expect that network effects would be larger in smaller geographic areas because individuals are more likely to have encounters (e.g., conversations at the grocery store or the smaller downtown area) with those others who are included in the contact availability measure. For adults (13A), I find positive and significant network effects among the smallest three quintile groups and statistically insignificant effects among the highest two quintile groups across both the core and expanded samples. Once again, the patterns among the child samples are odd: There are positive and significant effects among the 3rd quintile group only in the core sample, and positive and significant effects among all five quintile groups in the expanded sample with no discernable pattern in terms of magnitude. However, there is some measurement error in the PUMA size variable because it is based on the geographic size of the PUMA in 2000.[27] The geographic size of the PUMA changes over time because PUMAs are defined by the number of people that live in the area. This could partially explain some of the confusing patterns found in the child sample results.

The last PUMA-area characteristic that I explored is foreign born population. I would expect that network effects would be stronger in neighborhoods that have a higher proportion of the population that is foreign born. Once again, I divided the sample into quintile groups with the 1st quintile corresponding to individuals in PUMAs with a low foreign born population as a percent of the total area population. Practically, this is very similar to dividing the sample based on the average levels of contact availability because average contact availability substantially increases as the percent of the PUMA that is foreign born increases. For adults (14A), I find that network policy multiplier effect is 31.5 percentage points among the highest quintile group in the core sample. The coefficient on the network variable is statistically insignificant among the other quintile groups. In the expanded adult sample, I also find stronger network effects in PUMAs with a higher percentage of foreign born populations. However,

[27] This is the most recent information on PUMA size that I can find. I will update the results with new data if it becomes available.

smaller, but statistically significant (at the 10% level) network effects prevail in some of the lower quintile groups. Once again, the results for the child sample (14B) are the opposite, as network effects are stronger in neighborhoods in the lowest foreign born quintile group. This could be attributable to differences in income and other socio-demographic characteristics among the adult and child sample because the adult sample has lower income levels due to differences in Medicaid income eligibility thresholds for adults and children.

Networks and State Policy

The purpose of this section is to determine if network effects vary based on differences in state policy design. As discussed in the previous chapter, I use an index for state Medicaid fee-for-service reimbursement levels to proxy for the generosity of state Medicaid benefits and access to care. I also use an indicator for whether or not the state has outreach in multiple languages for pregnant women eligible for Medicaid to proxy for state outreach efforts.

For adults (Table 16A), I find some evidence that network effects are stronger in states that have less generous Medicaid fees, but comparable across multiple language outreach and non-outreach states. The former result suggests that networks effects are stronger in states that have worse access to providers for the low-income Medicaid population. I find mixed results among the child samples (Table 16B). Network effects appear stronger in states in the top 3 quintiles of Medicaid fee generosity. However, this might be a weak proxy for access because this index does not incorporate Medicaid/CHIP managed care capitation payments, and approximately 50% of Medicaid/CHIP children are enrolled in a managed care plans compared to only 25% of adults (Kaiser Family Foundation 2007).[28]

Specification Tests

The results in Tables 16, 17, 18, and 19 indicate that the main results from this study are insensitive to several variable and model definition specifications. To summarize, the results in these tables show that results are relatively unaffected by the following choices:

- The use of survey weights in the OLS regression model (Tables 16A/16B).
- Defining Medicaid take-up at the state-level and Super-PUMA level, as opposed to the national level, for quality component of the network variable (Tables 16A/16B).
- Removing the underweighting denominator component from the contact availability component of the network variable (Tables 16A/16B).

---

[28] http://www.statehealthfacts.org/comparemaptable.jsp?ind=200&cat=4, accessed on March 27th, 2011.

- The use of logit and probit models, with PUMA characteristics and Super-PUMA fixed effects (PUMA fixed effects models do not converge), as opposed to linear probability models (Tables 17A/17B)
- Placebo tests that change the left hand side variable to something (e.g., poverty level indicators, age, welfare take-up, and marital status) other than Medicaid take-up. For the majority of models, the Medicaid network variable does not have a statistically significant effect on the placebo LHS variable and where it does, the magnitude is small. This provides more confidence in the internal validity of the main network variable of interest (Tables 19A/19B).
- The use of multinomial logit models for the expanded samples (Tables 18A/18B).

The last bullet warrants further discussion. For the expanded adult sample (Table 18A), I ran three multinomial logit models with PUMA characteristics (PUMA and Super-PUMA fixed effects models do not converge) and language group fixed effects. The first two models include three choice outcomes, with non-group included in the private insurance choice in the first model and non-group combined with Medicaid in the second model. The third model includes private non-group as separate choice. For all models (the base choice is Medicaid), I find that the network variable has a negative effect on the probability of being uninsured relative to having Medicaid. I also find that this effect is larger when private non-group is combined with Medicaid. This could be attributable to the fact that there is considerable measurement error among those who report having private non-group (they might actually have Medicaid) or the fact that some individuals (most likely healthy individuals who can "afford" a relatively expensive non-group policy) who are actually ineligible for Medicaid purchase a non-group policy as an alternative. The latter point is supported by the fact that I found a positive and significant coefficient on the network variable for non-group vs. Medicaid choice among the foreign born subsample, but an insignificant coefficient among the non-foreign born population (results not shown). I also found similar results for the expanded child sample (Table 18B).

Language Group Distributional Effects

Finally, Tables 20A and 20B explore whether network effects are stronger among language groups with initially high versus low take-up of Medicaid. To address this area, I estimated the main regression as usual, but included language group take-up quintile dummies (instead of the continuous measure of take-up) interacted with the contact availability measure. In addition, for this empirical test I defined language group take-up at the state level (column 1) and the super-PUMA level (column 2) in order to create more balanced quintiles. Otherwise, the abundance of Spanish speakers would dominate a single quintile because take-up in the standard definition only varied across language groups. For each model, I excluded individuals in state-language group or super PUMA-language group cells with sample sizes under 30 in order to obtain more precise take-up estimates. For adults, this restriction reduced the sample size by approximately 8% in the state take-up models and over 25% in the super-PUMA models. For children, this restriction reduced the sample size by approximately 5% in the state take-up models and over 15% in the super-PUMA models. Tables 20A and 20B report the sample sizes and the the four coefficients (the 3rd quintile group is the excluded category) of the newly defined key variables of interest.

In all of the adult and child models where language group take-up is defined at the state level, I find that living in a high CA area has a positive and statistically significant impact on the individual's probability of take-up among those who are in the top quintile take-up group relative to those who are in the 3rd quintile. In three out of the four models (child expanded sample being the exception), I find that among those who are part of the lowest take-up quintile , living in a CA area has a negative and statistically significant effect on take-up compared to those in the 3rd quintile.

I find similar patterns, but with more varying results, when language group take-up is defined at the Super-PUMA level. For example, among adults, CA has a negative and statistically significant effect on take-up among the lowest quintile take-up group, but no effect among the highest take-up group. Among children in the expanded sample, there are statistically significant effects among both the top (positive) and bottom (negative) quintile groups. Among children in the core sample, there are positive and statistically significant effects among both the top quintile group.

Table 1A
Weighted Descriptive Statistics
Core vs. Expanded Adult Sample

	Core Sample	Expanded Sample
Unweighted N	59,377	83,906
Weighted N	7,755,281	10,786,093
Health Insurance Status		
Medicaid	36.7%	26.4%
Any Private	0.0%	28.1%
Uninsured	63.3%	45.5%
Foreign born	77.2%	75.5%
Fluent in English	54.3%	60.0%
Non-Citizen	56.1%	52.2%
MSA Status		
Non-MSA	9.3%	9.3%
MSA not identifiable	3.3%	3.6%
MSA, central city	33.3%	32.4%
MSA, outside central city	24.4%	25.2%
MSA, central city status unknown	29.7%	29.5%
Female	55.3%	54.6%
Age		
Age, 19-24	14.2%	15.6%
Age, 25-34	31.4%	29.8%
Age, 35-44	28.7%	28.7%
Age, 45-54	16.8%	17.0%
Age, 55-64	8.9%	8.9%
Education		
< High school	50.0%	43.4%
High school graduate	26.4%	26.1%
Some college	16.7%	20.1%
College+	6.9%	10.4%

Race and Ethnicity		
White, non-Hispanic	12.2%	14.9%
Black, non-Hispanic	3.4%	3.8%
Asian, non-Hispanic	10.4%	13.5%
Hispanic	71.4%	65.4%
Other and multiple races	2.6%	2.4%
Married	50.8%	51.3%
Family size	3.7	3.6
Number of own children in family	1.6	1.6
Income Relative to Poverty		
<=100% FPL	68.7%	61.8%
101-200% FPL	28.2%	32.4%
201-300% FPL	3.2%	5.8%
Work Status		
Worker, not self-employed	61.2%	65.6%
Worker, self-employed	8.9%	8.0%
Non-worker	29.9%	26.4%
Has Welfare Income	6.6%	5.2%
Number of disabilities	0.23	0.20
Year 2009	53.5%	52.5%

Source: 2008-2009 American Community Surveys

Table 1B
Weighted Descriptive Statistics
Core vs. Expanded Children Sample

	Core Sample	Expanded Sample
Unweighted N	136,542	192,414
Weighted N	17,459,492	24,047,763
Health Insurance Status		
Medicaid	71.6%	52.0%
Any Private	0.0%	27.4%
Uninsured	28.4%	20.6%
Foreign born	17.3%	16.5%
Fluent in English	65.4%	67.7%
Non-Citizen	14.1%	12.8%
MSA Status		
Non-MSA	8.3%	8.0%
MSA not identifiable	2.6%	2.6%
MSA, central city	27.1%	26.7%
MSA, outside central city	27.7%	29.7%
MSA, central city status unknown	34.4%	33.1%
Female	48.8%	48.9%
Age		
Infant	5.5%	5.1%
Age, 1-5	28.7%	27.3%
Age, 6-19	65.8%	67.5%
Number Family Members with At Least Some College		
0	66.2%	59.5%
1	24.1%	26.7%
2+	9.7%	13.8%
Race and Ethnicity		
White, non-Hispanic	8.9%	11.2%

Black, non-Hispanic	3.4%	3.9%
Asian, non-Hispanic	6.5%	8.5%
Hispanic	79.5%	74.3%
Other and multiple races	1.8%	2.1%
Family size	5.0	4.9
Two-Parent Family	65.1%	66.6%
Income Relative to Poverty		
<=100% FPL	50.9%	42.9%
101-200% FPL	41.9%	44.1%
201-300% FPL	6.7%	11.4%
301%+	0.5%	1.6%
Number of Workers in Family		
0	7.1%	6.1%
1	44.6%	43.4%
2+	48.3%	50.5%
Number of disabilities	0.06	0.05
Year 2009	52.6%	51.6%

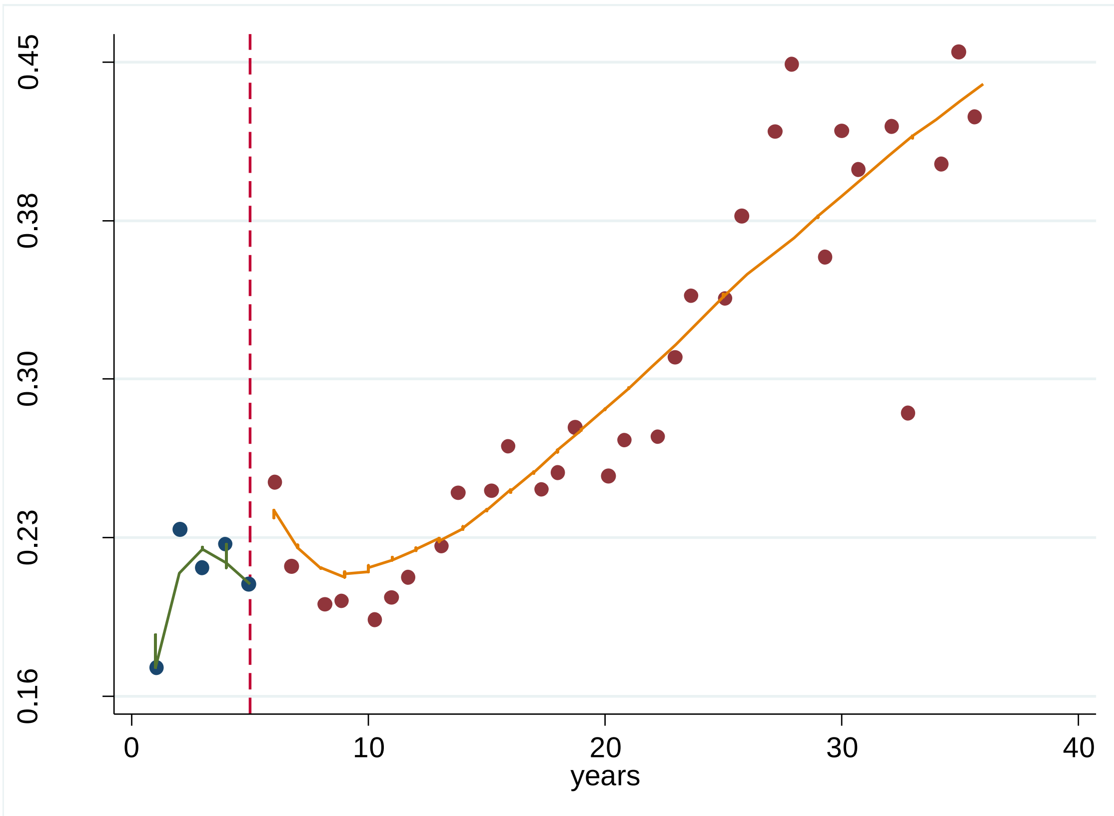
Source: 2008-2009 American Community Surveys

Table 2
Medicaid/CHIP Take-Up Rates by Citizenship Status and Number of Years in U.S.
Core Adult and Children Samples

	Adults		Children	
	Citizens	**Non-Citizens**	**Citizens**	**Non-Citizens**
Overall	53.3%	23.6%	76.9%	39.0%
Years in U.S.				
Not Foreign Born	49.8%	N/A	77.0%	N/A
<1	56.8%	19.2%	64.1%	33.8%
1	62.8%	21.6%	76.3%	39.5%
2	58.0%	20.7%	77.5%	37.3%
3	64.4%	20.4%	74.2%	38.9%
4	48.3%	20.2%	70.1%	41.8%
5	59.0%	24.3%	69.6%	41.8%
6	50.4%	19.7%	70.4%	39.9%
7	58.6%	18.7%	77.3%	37.4%
8	57.4%	18.7%	79.4%	38.4%
9	58.2%	18.7%	76.7%	37.3%
10	47.1%	19.0%	73.8%	37.1%
11-15	55.7%	22.7%	73.7%	40.3%
16-20	56.5%	25.3%	60.9%	38.1%
21+	58.8%	35.9%	n.a.	n.a.

Source: 2008-2009 American Community Surveys

Figure 5: Medicaid Take-Up Rates Among Non-Citizen Adults By Number of Years in U.S.



Source: 2008-2009 American Community Survey.

Table 3A
Select Weighted Descriptive Statistics
Core Adult Sample by Language Group

	Sample Size	Weighted Sample	Medicaid	Age	<=100% FPL	Family Size	Foreign Born	Non-citizen
Full Sample	59,377	7,755,281	0.37	37.1	0.69	3.7	0.77	0.56
Language Group								
Yiddish, Jewish	529	48,934	0.95	34.1	0.84	6.3	0.18	0.05
Cantonese	473	54,065	0.74	44.9	0.73	3.8	0.95	0.47
Armenian	159	21,148	0.68	41.6	0.91	3.6	0.97	0.62
Bengali	308	38,369	0.67	39.6	0.70	4.4	0.99	0.54
Miao, Hmong	349	48,810	0.66	35.1	0.69	5.5	0.79	0.41
Mon-Khmer, Cambodian	302	38,935	0.63	40.8	0.60	4.3	0.86	0.40
Hebrew, Israeli	154	15,853	0.63	36.7	0.83	3.6	0.45	0.16
Persian	214	26,841	0.57	41.0	0.74	3.7	0.94	0.50
Arabic	1,093	149,429	0.56	38.2	0.79	4.1	0.86	0.47
French or Haitian Creole	499	70,690	0.54	38.2	0.66	3.5	0.88	0.52
Vietnamese	1,163	136,242	0.53	41.4	0.66	3.7	0.95	0.40
Greek	131	15,784	0.53	42.8	0.57	3.2	0.43	0.20
Russian	872	112,629	0.51	40.6	0.66	3.2	0.92	0.54
Hungarian	18	1,817	0.49	45.9	0.75	3.3	0.74	0.28
Italian	294	32,536	0.48	42.9	0.56	2.6	0.36	0.14
Chinese	1,278	144,500	0.47	42.2	0.66	3.4	0.94	0.54
Turkish	66	7,796	0.47	36.6	0.46	3.6	0.94	0.66
Urdu	378	50,734	0.46	41.1	0.61	4.9	0.97	0.44
Laotian	133	18,353	0.46	37.2	0.60	3.9	0.76	0.43
Serbo-Croatian	86	10,576	0.46	40.4	0.65	3.1	0.95	0.41
Mandarin	336	46,106	0.44	41.3	0.70	3.1	0.96	0.69
Portuguese	821	123,390	0.44	36.8	0.36	2.7	0.88	0.70
French	771	96,058	0.42	39.4	0.68	2.6	0.55	0.40
Amharic, Ethiopian, etc.	169	23,834	0.41	37.7	0.69	2.8	0.95	0.51
Ukrainian	187	20,870	0.41	39.4	0.46	4.2	0.95	0.68
Albanian	84	13,907	0.40	38.0	0.41	4.4	0.92	0.50

Tamil, Malayalam and Telugu	73	8,530	0.37	37.0	0.74	3.7	1.00	0.80
Hindi and Punjabi	440	58,934	0.36	41.4	0.62	3.9	0.96	0.58
Polish	381	48,052	0.36	41.4	0.49	2.9	0.86	0.54
Navajo	1,297	96,999	0.35	39.2	0.70	3.8	0.00	0.00
Filipino, Tagalog	357	42,497	0.35	40.8	0.60	2.9	0.91	0.46
Kru	237	36,205	0.34	37.4	0.61	3.3	0.94	0.63
Spanish	43,070	5,817,727	0.33	36.3	0.70	3.8	0.77	0.60
Rumanian	109	15,779	0.30	36.4	0.66	3.0	0.78	0.38
German	765	67,996	0.29	37.9	0.60	3.8	0.24	0.12
Guajarati	158	21,279	0.28	43.1	0.63	3.7	0.97	0.49
Thai	79	10,296	0.25	39.7	0.79	3.1	0.91	0.60
Korean	792	96,062	0.22	41.3	0.65	3.0	0.93	0.61
Japanese	164	20,304	0.21	35.1	0.75	2.0	0.67	0.57
Dutch	158	13,181	0.14	35.9	0.53	4.8	0.20	0.12
Pennsylvania Dutch	430	33,234	0.05	35.9	0.45	6.6	0.00	0.00

Source: 2008-2009 American Community Surveys

Table 3B
Select Weighted Descriptive Statistics
Core Children Sample by Language Group

	Sample Size	Weighted Sample	Medicaid/CHIP	Age	<=100% FPL	Foreign Born	Non-citizen
Full Sample	136,542	17,459,492	0.72	8.6	0.51	0.17	0.14
Language Group							
Yiddish, Jewish	1,180	114,830	0.99	7.5	0.72	0.02	0.00
Hungarian	37	4,333	0.93	11.3	0.48	0.07	0.07
Miao, Hmong	783	97,823	0.90	10.1	0.57	0.25	0.20
Hebrew, Israeli	304	29,238	0.89	7.9	0.46	0.10	0.03
Cantonese	632	67,360	0.89	10.5	0.44	0.26	0.18
Serbo-Croatian	119	17,853	0.87	8.8	0.35	0.40	0.29
Albanian	201	30,636	0.86	8.7	0.35	0.31	0.18
Bengali	466	62,489	0.83	8.1	0.49	0.35	0.17
Ukrainian	371	37,970	0.82	9.8	0.30	0.52	0.43
Amharic, Ethiopian, etc.	277	33,305	0.82	7.8	0.49	0.33	0.24
Arabic	2,059	288,215	0.81	8.2	0.60	0.30	0.19
Armenian	245	33,595	0.81	11.1	0.53	0.39	0.34
Mandarin	302	36,927	0.81	9.2	0.39	0.31	0.26
Persian	332	40,672	0.80	9.7	0.49	0.35	0.25
Polish	364	43,765	0.79	8.8	0.31	0.19	0.13
French	1,157	133,986	0.79	9.2	0.45	0.17	0.14
Vietnamese	2,135	228,572	0.77	9.3	0.46	0.22	0.12
Russian	1,075	122,433	0.76	8.8	0.41	0.41	0.29
Mon-Khmer, Cambodian	478	54,309	0.76	9.9	0.48	0.10	0.07
Urdu	842	107,877	0.75	8.8	0.47	0.34	0.18
Kru	490	65,451	0.75	7.2	0.38	0.27	0.20
Italian	327	33,488	0.74	10.0	0.39	0.10	0.04
Filipino, Tagalog	958	96,040	0.73	9.1	0.30	0.36	0.25
Chinese	1,498	161,692	0.73	9.4	0.43	0.27	0.21
Rumanian	153	21,936	0.72	8.3	0.38	0.17	0.14
Spanish	109,919	14,468,963	0.72	8.5	0.52	0.16	0.13

Laotian	213	28,042	0.71	8.6	0.46	0.09	0.06
Llocano, Hocano	61	5,295	0.71	10.4	0.22	0.32	0.19
Greek	138	17,412	0.71	10.0	0.40	0.07	0.02
Japanese	202	21,979	0.69	8.4	0.34	0.27	0.08
Turkish	177	19,433	0.68	8.1	0.33	0.49	0.45
Portuguese	689	93,706	0.66	8.9	0.33	0.31	0.27
Hindi and Punjabi	669	86,987	0.66	8.9	0.33	0.31	0.19
Navajo	1,056	79,506	0.63	9.9	0.64	0.00	0.00
Other American Indian	337	29,863	0.63	9.4	0.62	0.08	0.07
Thai	118	14,346	0.62	10.0	0.40	0.32	0.24
French or Haitian Creole	1,572	219,590	0.61	9.0	0.47	0.27	0.22
Guajarati	213	24,249	0.57	10.0	0.41	0.37	0.30
Tamil, Malayalam and Telugu	75	8,786	0.55	9.5	0.53	0.49	0.44
Korean	1,173	128,283	0.48	9.9	0.36	0.38	0.34
German	1,617	135,941	0.39	8.8	0.48	0.09	0.05
Dutch	406	28,403	0.18	7.7	0.44	0.03	0.01
Pennsylvania Dutch	1,122	83,913	0.09	7.9	0.39	0.00	0.00

Source: 2008-2009 American Community Surveys

Table 4A
Contact Availability Variable
Core Adult Model

	Obs	Mean	Std. Dev	Min	Max
Log Transformed Contact Availability					
PUMA level	59377	0.89	1.56	-4.18	6.37
Super-PUMA level	59377	0.64	1.38	-4.78	5.49
MSA level	57119	0.41	1.05	-4.40	5.46
Sensitivity Specifications					
No Denominator	59377	-2.42	1.66	-10.02	-0.11
No Log Transformation	59377	16.89	71.85	0.02	583.30
No Denominator, No Log Transformation	59377	0.20	0.20	0.000	0.89

Source: 2008-2009 American Community Surveys

Table 4B
Contact Availability Variable
Core Child Model

	Obs	Mean	Std. Dev	Min	Max
Log Transformed Contact Availability					
PUMA level	136,542	0.97	1.30	-5.76	6.34
Super-PUMA level	136,542	0.75	1.17	-4.43	5.48
MSA level	132,390	0.56	0.97	-5.83	5.52
Sensitivity Specifications					
No Denominator	136,542	-1.88	1.53	-11.50	-0.06
No Log Transformation	136,542	10.88	49.40	0.00	565.92
No Denominator, No Log Transformation	136,542	0.30	0.26	0.000	0.94

Source: 2008-2009 American Community Surveys

Table 5
Tabulation of State Variable
Core Adult and Child Samples

	Adults			Children		
	Sample Size	Weighted Sample Size	%	Sample Size	Weighted Sample Size	%
Total	59,377	7,755,281	100.0	136,542	17,459,492	100.0
Alabama	33	3,130	0.0	534	71,824	0.4
Alaska	13	1,366	0.0	84	10,097	0.1
Arizona	3,786	468,646	6.0	5,283	700,221	4.0
Arkansas	637	81,116	1.1	694	88,950	0.5
California	11,733	1,392,424	18.0	38,492	4,656,373	26.7
Colorado	311	49,680	0.6	1,840	271,326	1.6
Connecticut	1,919	259,209	3.3	1,342	170,668	1.0
Delaware	158	20,095	0.3	186	25,177	0.1
DC	124	14,935	0.2	77	8,690	0.1
Florida	972	120,258	1.6	8,528	1,069,004	6.1
Georgia	331	44,838	0.6	3,157	439,133	2.5
Hawaii	230	26,244	0.3	205	19,411	0.1
Idaho	468	57,342	0.7	508	56,741	0.3
Illinois	2,748	416,365	5.4	4,467	657,816	3.8
Indiana	1,378	171,106	2.2	1,431	164,771	0.9
Iowa	415	62,897	0.8	363	51,095	0.3
Kansas	28	5,754	0.1	709	98,186	0.6
Kentucky	84	11,193	0.1	448	49,720	0.3
Louisiana	25	4,051	0.1	570	67,805	0.4
Maine	188	22,022	0.3	87	8,230	0.1
Maryland	742	95,353	1.2	1,277	163,262	0.9
Massachusetts	3,428	468,042	6.0	2,221	289,504	1.7
Michigan	606	81,500	1.1	1,536	194,355	1.1
Minnesota	916	151,952	2.0	801	124,329	0.7
Mississippi	24	2,839	0.0	271	32,755	0.2
Missouri	50	6,894	0.1	876	98,478	0.6
Montana	6	917	0.0	59	6,975	0.0

Nebraska	37	6,642	0.1	347	58,783	0.3
Nevada	940	124,912	1.6	1,343	179,207	1.0
New Hampshire	9	1,635	0.0	95	14,843	0.1
New Jersey	2,246	284,600	3.7	3,905	504,103	2.9
New Mexico	2,906	339,975	4.4	1,978	235,619	1.4
New York	9,280	1,261,785	16.3	10,828	1,460,670	8.4
North Carolina	237	29,621	0.4	2,674	361,723	2.1
North Dakota	0	0	0.0	30	4,639	0.0
Ohio	342	44,578	0.6	1,275	146,795	0.8
Oklahoma	990	140,012	1.8	851	111,465	0.6
Oregon	1,443	205,470	2.7	1,383	197,752	1.1
Pennsylvania	2,693	360,867	4.7	2,551	326,039	1.9
Rhode Island	239	32,886	0.4	451	56,755	0.3
South Carolina	209	25,001	0.3	771	96,769	0.6
South Dakota	6	1,340	0.0	49	7,408	0.0
Tennessee	844	105,181	1.4	1,044	136,268	0.8
Texas	1,223	159,235	2.1	25,101	3,218,383	18.4
Utah	662	94,799	1.2	758	104,970	0.6
Vermont	57	6,638	0.1	49	6,635	0.0
Virginia	62	8,787	0.1	1,071	146,808	0.8
Washington	2,552	328,889	4.2	2,850	348,625	2.0
West Virginia	4	404	0.0	53	5,386	0.0
Wisconsin	1,038	151,417	2.0	980	128,258	0.7
Wyoming	5	439	0.0	59	6,693	0.0

Source: 2008-2009 American Community Surveys

Figure 6

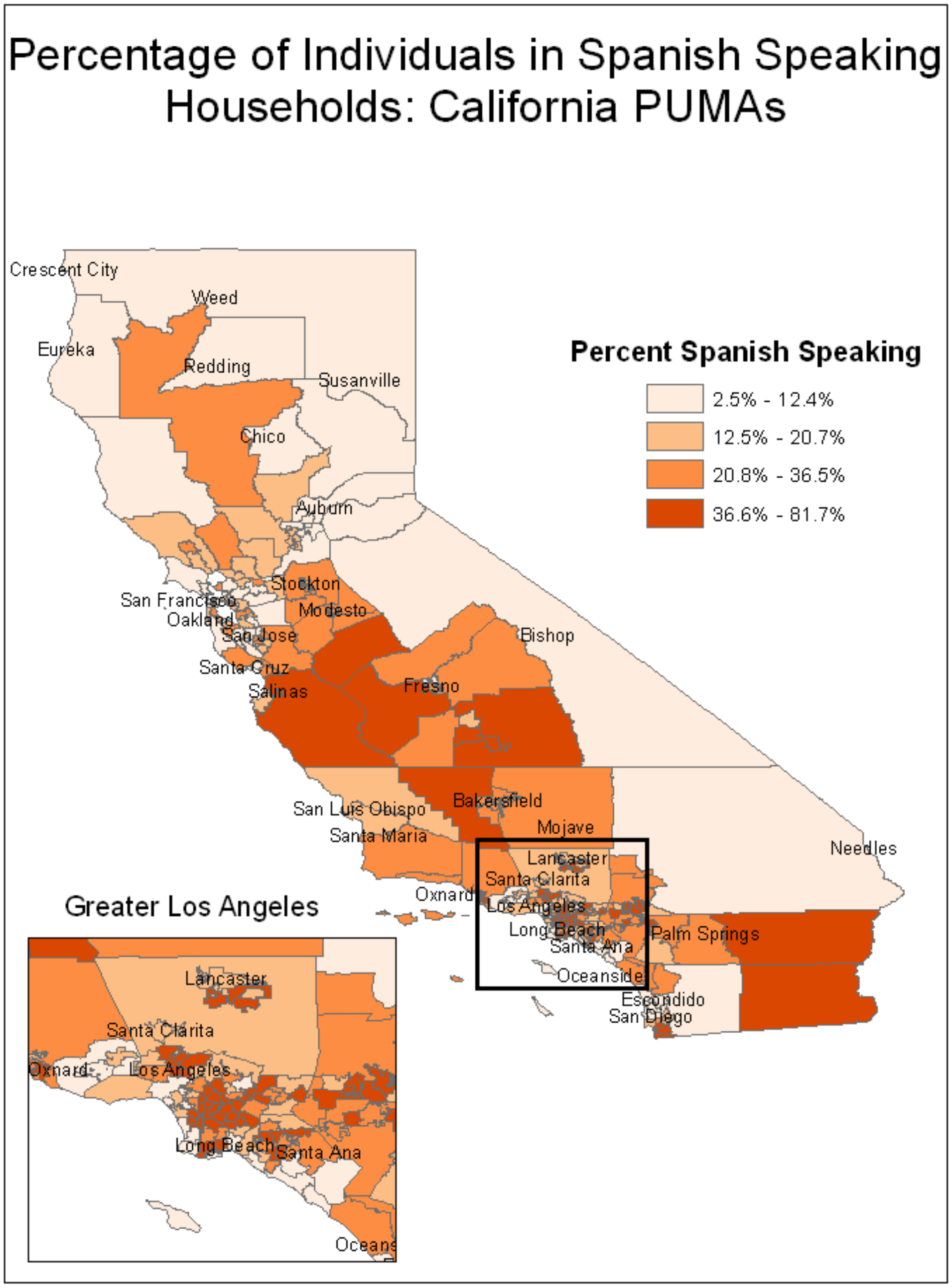

Percentage of Individuals in Spanish Speaking Households: California PUMAs
Percent Spanish Speaking
2.5% - 12.4%
12.5% - 20.7%
20.8% - 36.5%
36.6% - 81.7%
Crescent City
Weed
Eureka
Redding
Susanville
Chico
Auburn
San Francisco
Oakland
Stockton
Modesto
San Jose
Santa Cruz
Salinas
Fresno
Bishop
San Luis Obispo
Santa Maria
Bakersfield
Mojave
Lancaster
Santa Clarita
Oxnard
Los Angeles
Long Beach
Santa Ana
Oceanside
Palm Springs
Escondido
San Diego
Needles
Greater Los Angeles
Lancaster
Santa Clarita
Oxnard
Los Angeles
Long Beach
Santa Ana

Figure 7

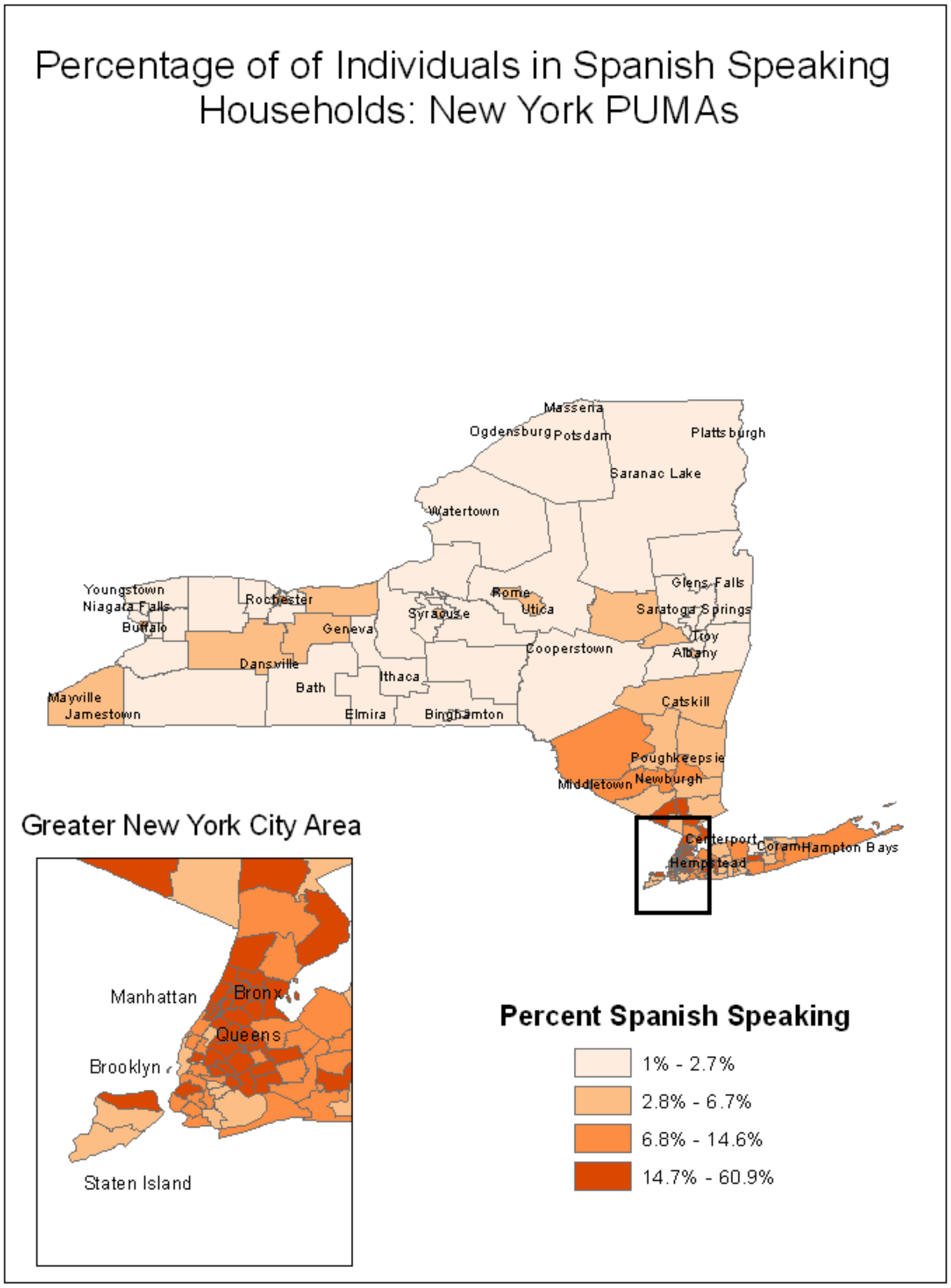

Percentage of of Individuals in Spanish Speaking Households: New York PUMAs
Massena
Ogdensburg
Potsdam
Plattsburgh
Saranac Lake
Watertown
Youngstown
Niagara Falls
Buffalo
Rochester
Geneva
Syracuse
Rome
Utica
Glens Falls
Saratoga Springs
Troy
Albany
Cooperstown
Dansville
Bath
Ithaca
Mayville
Jamestown
Elmira
Binghamton
Catskill
Poughkeepsie
Middletown
Newburgh
Centerport
Coram
Hampton Bays
Hempstead
Greater New York City Area
Manhattan
Bronx
Queens
Brooklyn
Staten Island
Percent Spanish Speaking
1% - 2.7%
2.8% - 6.7%
6.8% - 14.6%
14.7% - 60.9%

Table 6
Core Sample: Difference-in-Differences for Unconditional Weighted Means
Low/High Contact Availability and Low/High Medicaid Take-Up Groups

	Low Contact Availability	High Contact Availability	ΔCA
Dependent Variable=Medicaid Take-Up			
Adults			
Low Take-Up	0.300 (0.000)	0.370 (0.000)	0.070 *** (0.000)
High Take-Up	0.409 (0.001)	0.573 (0.000)	0.165 *** (0.001)
Difference-in-difference			**0.095 *** (0.012)**
Children			
Low Take-Up	0.574 (0.001)	0.516 (0.001)	-0.0580 *** (0.001)
High Take-Up	0.698 (0.000)	0.757 (0.000)	0.0582 *** (0.000)
Difference-in-difference			**0.116 *** (0.013)**

Source: 2008-2009 American Community Surveys

Notes: (1) Standard errors are in parenthesis. (2) Low (high) refers to individuals that belong to a language group with below (above) mean Medicaid take-up. (3) Low (high) CA refers to individuals that live in PUMAs with below (above) mean CA.

Table 7A
Full OLS Regression Results
Adult Samples

	Core Sample		Expanded Sample	
Medicaid Network: Mean Take-Up of Language Group*lnCA	***0.100***	******	***0.118***	******
Contact Availability (lnCA)	***-0.034***	******	***-0.035***	******
MSA not identifiable	-0.108	**	0.134	
MSA, central city	0.232	***	0.162	
MSA, outside central city	-0.252	***	-0.027	
MSA, central city status unknown	0.332	***	0.354	**
Non-citizen	-0.187	***	-0.161	***
Female	0.072	***	0.044	***
Age, 19-24	0.007		0.020	***
Age, 25-34	-0.006		-0.033	***
Age, 45-54	0.031	***	0.026	***
Age, 55-64	0.069	***	0.067	***
HS dropout	0.010	**	-0.015	***
Some college	0.019	***	0.052	***
College+	-0.028	***	0.065	***
Black	0.019		-0.039	***
Asian	0.020		0.037	*
Hispanic	-0.030	**	-0.072	***
Other and multiple races	0.014		-0.037	**
Foreign born	-0.019	***	0.000	
Fluent in English	0.023	***	0.062	***
Married	-0.004		0.016	***
Family size	-0.004	***	-0.015	***
Number of own children in family	0.037	***	0.046	***
101-200% FPL	-0.036	***	0.053	***
201-300% FPL	-0.110	***	0.091	***
Welfare	0.269	***	0.191	***
Number of disabilities	0.095	***	0.065	***
Year 2009	0.017	***	-0.004	
Work, not self-employed	-0.080		0.011	
Work, self-employed	-0.121		-0.096	**
_cons	0.123	***	-0.682	***
N	59,377		83,906	
Adjusted R-Squared	0.345		0.258	
Dependent Variable	Medicaid		Insured	
Language Group Fixed Effects?	Yes		Yes	
PUMA Fixed Effects?	Yes		Yes	
Medicaid Multiplier Effect	9.9%		11.0%	

Source: 2008-2009 American Community Surveys
Notes:(1) Robust standard errors are in parenthesis and are cluster corrected by local area and language group. (2) Language, local area, and occupation dummies not shown. (3) The collinearity condition number for lnCA and the Medicaid network variable (lnCA*LG take-up rate) in the core sample is 5.2.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 7B
Full OLS Regression Results
Child Samples

	Core Sample	Expanded Sample
Medicaid Network: Mean Take-Up of Language Group*lnCA	***0.071 ******	***0.100 ******
Contact Availability (lnCA)	-0.057 ***	0.007 ***
MSA not identifiable	0.058	-0.088
MSA, central city	0.137	0.169
MSA, outside central city	-0.014	0.032
MSA, central city status unknown	0.134	0.123
Non-citizen	-0.282 ***	-0.250 ***
Female	0.004 *	0.004 **
Age, 0	0.054 ***	0.039 ***
Age, 6-19	-0.071 ***	-0.052 ***
# some college in family	-0.001	0.019 ***
Black	0.083 ***	0.035 ***
Asian	0.020	0.030 ***
Hispanic	-0.017 *	-0.047 ***
Other and multiple races	0.036 **	0.011
Foreign born	-0.036 ***	-0.025 ***
Fluent in English	0.013 ***	0.023 ***
Number of parents	-0.012 ***	-0.009 ***
Family size	0.013 ***	0.005 ***
# own children	-0.029 **	-0.050 ***
101-200% FPL	-0.055 ***	-0.008 **
201-300% FPL	-0.097 ***	0.014 ***
>300% FPL	-0.126 ***	0.031 ***
Number of workers in family	-0.017 ***	-0.018 ***
Number of disabilities	0.065 ***	0.042 ***
Year 2009	0.035 ***	0.018 ***
_cons	0.382 ***	-0.403 ***
N	136,542	192,414
Adjusted R-Squared	0.213	0.163
Dependent Variable	Medicaid	Insured
Language Group Fixed Effects?	Yes	Yes
PUMA Fixed Effects?	Yes	Yes
Medicaid Multiplier Effect	7.4%	10.4%

Source: 2008-2009 American Community Surveys

Notes:(1) Robust standard errors are in parenthesis and are cluster corrected by local area and language group. (2) Language, local area, and occupation dummies not shown, but are available upon request. (3) The collinearity condition number for lnCA and the Medicaid network variable (lnCA*LG take-up rate) is 9.3

*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 8A
Adult Sample OLS Regression Results
Addressing Differential Geographic Selection

	1	2	3	4	5	6	7	8	9
Fixed Effects	PUMA, language	Super-PUMA, language	MSA, language	PUMA, language	PUMA, YSE*lang	PUMA, language	PUMA, language	PUMA, language	PUMA, language
Sample	Full Sample	Full Sample	MSA Only	Foreign Born	Foreign born	Not Foreign Born	Foreign Born, 0-2 YSE	Foreign Born, 0-5 YSE	Foreign Born, 5+ YSE
Core Sample (Dependent Variable=Medicaid)									
Medicaid Network	0.100 ***	0.089 ***	0.125 ***	0.134 ***	0.136 ***	0.032	0.268 ***	0.175 ***	0.122 ***
	(0.020)	(0.019)	(0.031)	(0.026)	(0.025)	(0.030)	(0.083)	(0.060)	(0.028)
Sample size	59,377	59,377	49,731	44,005	44,005	15,372	3,238	8,022	35,983
Policy Multiplier	9.9%	6.0%	7.2%	12.8%	13.0%	N/A	29.6%	16.3%	11.7%
Expanded Sample (Dependent Variable=Insured)									
Medicaid Network	0.118 ***	0.104 ***	0.168 ***	0.149 ***	0.148 ***	0.031	0.379 ***	0.242 ***	0.098 ***
	(0.021)	(0.021)	(0.034)	(0.027)	(0.027)	(0.033)	(0.071)	(0.054)	(0.030)
Sample size	83,906	83,906	70,306	61,049	61,049	22,857	5,200	11,580	49,469
Policy Multiplier	11.0%	6.4%	9.5%	14.4%	14.3%	N/A	37.7%	22.0%	9.3%

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level;

Table 8B
Child Sample OLS Regression Results
Addressing Differential Geographic Selection

	1	2	3	4	5	6	7	8	9
Fixed Effects	PUMA, language	Super-PUMA, language	MSA, language	PUMA, language	PUMA, YSE*lang	PUMA, language	PUMA, language	PUMA, language	PUMA, language
Sample	Full Sample	Core	MSAs Only	Foreign Born	Foreign born	Not Foreign	Foreign Born, 0-2	Foreign Born, 0-5	Foreign Born, 5+
Core Sample (Dependent Variable=Medicaid)									
Medicaid Network	0.071 ***	0.071 ***	0.104 ***	0.050	0.077	0.067 ***	0.158	0.141 *	0.077
	(0.024)	(0.021)	(0.036)	(0.058)	(0.058)	(0.026)	(0.126)	(0.083)	(0.058)
Sample size	136,542	136,542	117,832	23,707	23,707	112,835	4,583	10,701	23,707
Policy Multiplier Effect	7.4%	5.7%	7.8%	N/A	N/A	7.1%	N/A	15.8%	7.7%
Expanded Sample (Dependent Variable=Insured)									
Medicaid Network	0.100 ***	0.089 ***	0.141 ***	0.266 ***	0.265 ***	0.059 ***	0.325 ***	0.276 ***	0.266 ***
	(0.015)	(0.015)	(0.021)	(0.033)	(0.033)	(0.017)	(0.075)	(0.051)	(0.047)
Sample size	192,414	192,414	166,708	31,786	31,786	160,628	6,474	14,348	17,438
Policy Multiplier Effect	10.4%	6.9%	10.5%	33.3%	33.2%	5.9%	43.5%	34.6%	32.8%

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10%

Table 9A
Adult Sample OLS Regression Results
Exploring Network Mechanisms

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Fixed Effects	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language
Sample	Full Sample	Ling. Isolated Household	Not Ling. Isolated Household	In Group Quarters	Does Not Speak English	Speaks English, But Not Well	Speaks English Well	Speaks English Very Well	Disabled	Non-Disabled
Core Sample (Dependent Variable=Medicaid)										
Medicaid Network	0.100 ***	0.118 ***	0.093 ***	0.031	-0.006	0.052	0.093 ***	0.109 ***	-0.016	0.109 ***
	(0.020)	(0.037)	(0.026)	(0.070)	(0.085)	(0.042)	(0.034)	(0.028)	(0.045)	(0.023)
Sample size	59,377	21,327	34,179	3,871	9,344	16,392	12,600	21,041	8,139	51,238
Policy Multiplier	9.9%	12.3%	10.4%	N/A	N/A	N/A	9.8%	10.3%	N/A	11.0%
Expanded Sample (Dependent Variable=Insured)										
Medicaid Network	0.118 ***	0.173 ***	0.093 ***	0.076	-0.060	0.096 *	0.118 ***	0.106 ***	0.069	0.120 ***
	(0.021)	(0.048)	(0.025)	(0.060)	(0.110)	(0.052)	(0.036)	(0.027)	(0.050)	(0.023)
Sample size	83,906	27,071	51,013	5,822	10,625	21,048	18,684	33,549	10,063	73,843
Policy Multiplier	11.0%	18.8%	9.4%	N/A	N/A	9.4%	12.1%	8.8%	N/A	11.5%

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 9B
Child Sample OLS Regression Results
Exploring Network Mechanisms

	1	2	3	4	5	6	7	8	9
Fixed Effects	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language
Sample	Full Sample	Ling. Isolated Household	Not Ling. Isolated Household	In Group Quarters	Mom Does Not Speak English or Speak English Well	Mom Speaks English Well	Mom Speaks English Very Well	Disabled	Non-Disabled
Core Sample (Dependent Variable=Medicaid)									
Medicaid Network	0.071 ***	-0.002	0.084 ***	0.436	-0.009	-0.045	0.071 **	0.013	0.075 ***
	(0.024)	(0.040)	(0.029)	(0.517)	(0.049)	(0.049)	(0.033)	(0.100)	(0.024)
Sample size	136,542	45,196	90,415	931	56,783	23,791	42,736	5,861	130,681
Policy Multiplier Effect	7.4%	N/A	9.2%	N/A	N/A	N/A	7.6%	N/A	8.0%
Expanded Sample (Dependent Variable=Insured)									
Medicaid Network	0.100 ***	0.033	0.092 ***	0.088	0.079 **	0.003	0.066 ***	-0.003	0.103 ***
	(0.015)	(0.033)	(0.017)	(0.077)	(0.035)	(0.034)	(0.020)	(0.049)	(0.016)
Sample size	192,414	55,325	134,786	2,303	70,013	35,148	68,947	7,608	184,806
Policy Multiplier Effect	10.4%	N/A	9.6%	N/A	8.1%	N/A	7.2%	N/A	10.8%

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 10A
Adult Sample OLS Regression Results
Sample Sensitivity

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Fixed Effects	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, Country of Birth	PUMA, language	PUMA, language	PUMA, language	PUMA, language
Sample	Full Sample	<=200% FPL	<=300% FPL	Exclude Spanish	Spanish Speakers	Exclude Non-Citizens	Excludes Yiddish	Excludes P. Dutch	Exclude CA/NY
Core Sample (Dependent Variable=Medicaid)									
Medicaid Network	0.100 ***	0.113 ***	0.121 ***	0.069 ***	0.067 ***	0.061 **	0.131 ***	0.110 ***	0.081 ***
	(0.020)	(0.015)	(0.014)	(0.022)	(0.015)	(0.026)	(0.021)	(0.022)	(0.027)
Sample size	59,377	165,341	210,754	16,307	32,307	28,367	58,848	58,947	38,364
Policy Multiplier	9.9%	11.3%	12.4%	17.4%	8.0%	6.5%	12.6%	10.7%	5.4%
Expanded Sample (Dependent Variable=Insured)									
Medicaid Network	0.118 ***	0.095 ***	0.129 ***	0.108 ***	0.092 ***	0.075 ***	0.146 ***	0.144 ***	0.08 ***
	(0.021)	(0.016)	(0.017)	(0.022)	(0.019)	(0.025)	(0.024)	(0.024)	(0.031)
Sample size	83,906	234,753	343,099	27,577	40,778	43,135	83,219	83,437	57,614
Policy Multiplier	11.0%	8.9%	12.8%	24.8%	11.2%	7.4%	13.5%	13.7%	5.1%

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 10B
Child Sample OLS Regression Results
Sample Sensitivity

	1	2	3	4	5	6	7	8	9
Fixed Effects	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, Country of Birth	PUMA, language	PUMA, language	PUMA, language	PUMA, language
Sample	Full Sample	<=200% FPL	<=300% FPL	Exclude Spanish	Spanish Speakers	Exclude Non-Citizens	Excludes Yiddish	Excludes P. Dutch	Exclude CA/NY
Core Sample (Dependent Variable=Medicaid)									
Medicaid Network	0.071 ***	0.083 ***	0.063 ***	0.049 *	-0.022	0.073 ***	0.086 ***	0.112 ***	0.095 ***
	(0.024)	(0.024)	(0.023)	(0.026)	(0.040)	(0.027)	(0.026)	(0.028)	(0.032)
Sample size	136,542	126,374	148,926	26,623	76,142	117,806	135,362	135,420	87,222
Policy Multiplier	7.4%	8.7%	6.5%	12.1%	N/A	7.8%	8.7%	11.9%	7.6%
Expanded Sample (Dependent Variable=Insured)									
Medicaid Network	0.100 ***	0.121 ***	0.091 ***	0.078 ***	0.018	0.052 ***	0.106 ***	0.108 ***	0.112 ***
	(0.015)	(0.018)	(0.015)	(0.020)	(0.051)	(0.016)	(0.016)	(0.014)	(0.023)
Sample size	192,414	165,947	222,870	46,206	97,368	168,437	190,846	191,163	119,840
Policy Multiplier	10.4%	12.6%	9.1%	18.2%	N/A	5.2%	10.7%	11.1%	8.6%

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 11A
Adult Sample OLS Regression Results
Replacing Fixed Effects with PUMA Characteristics

			1	2	3	4
PUMA Characteristic(s)	Fixed Effects	All Four	% Non-White	% FPL	Age	% Foreign Born
Language group fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Medicaid Network	0.100 ***	0.092 ***	0.105 ***	0.108 ***	0.074 ***	0.12 ***
	(0.020)	(0.022)	(0.022)	(0.021)	(0.021)	(0.021)
lnCA	-0.034 ***	-0.025 **	-0.02 *	-0.02 **	-0.001	-0.04 ***
	(0.009)	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)
% PUMA non-white	-	-0.064 **	0.020	-	-	-
		(0.030)	(0.024)			
% PUMA <100% FPL	-	0.335 ***	-	0.130 **	-	-
		(0.064)		(0.058)		
Average age in PUMA	-	0.008 ***	-	-	0.006 ***	-
		(0.001)			(0.001)	
% PUMA foreign born	-	0.327 ***	-	-	-	0.268 ***
		(0.051)				(0.042)
Sample Size	59,377	59,377	59,377	59,377	59,377	59,377
Policy Multiplier Effect	9.9%	9.0%	10.4%	10.7%	7.1%	12.0%

Source: 2008-2009 American Community Surveys
Notes:(1) Robust standard errors are in parenthesis and are cluster corrected by local area and language group. (2) Full regression results available upon request. (2) Language, local area, and occupation dummies not shown. (3) The collinearity condition number for the four PUMA characteristic variables is 46.8.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 11B
Child Sample OLS Regression Results
Replacing Fixed Effects with PUMA Characteristics

PUMA Characteristic(s)	Fixed Effects	All Four	**1** % Non-White	**2** % FPL	**3** Age	**4** % Foreign Born
Language group fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Network	0.071 ***	0.128 ***	0.138 ***	0.141 ***	0.145 ***	0.128 ***
	(0.024)	(0.027)	(0.026)	(0.026)	(0.025)	(0.027)
lnCA	-0.057 ***	-0.094 ***	-0.096 ***	-0.093 ***	-0.090 ***	-0.097 ***
	(0.018)	(0.020)	(0.019)	(0.019)	(0.019)	(0.020)
% PUMA non-white	-	0.013	0.041 ***	-	-	-
		(0.022)	(0.016)			
% PUMA <100% FPL	-	0.092 *	-	0.043	-	-
		(0.049)		(0.042)		
Average age in PUMA	-	0.004 ***	-	-	0.003 ***	-
		(0.001)			(0.001)	
% PUMA foreign born	-	0.17 ***	-	-	-	0.183 ***
		(0.038)				(0.031)
Sample Size	136,542	136,542	136,542	136,542	136,542	136,542
Policy Multiplier Effect	7.4%	14.2%	15.5%	15.9%	16.4%	14.2%

Source: 2008-2009 American Community Surveys

Notes:(1) Robust standard errors are in parenthesis and are cluster corrected by local area and language group. (2) Full regression results available upon request. (2) Language and local area dummies not shown. (3) The collinearity condition number for the four PUMA characteristic variables is 44.9.

*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 12A
Adult Sample OLS Regression Results
By PUMA Poverty Level Quintile

Fixed Effects?	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language
Sample	Full Sample	Lowest Quintile	2	3	4	Highest Quintile
Core Sample (Dependent Variable=Medicaid)						
Medicaid Network	0.100 ***	0.133 *	0.137 ***	0.125 ***	0.107 ***	0.061
	(0.020)	(0.069)	(0.039)	(0.041)	(0.040)	(0.043)
Sample size	59,377	11,892	11,896	11,855	11,913	11,821
Policy Multiplier	9.9%	28.7%	14.6%	9.5%	7.8%	N/A
Expanded Sample (Dependent Variable=Insured)						
Medicaid Network	0.118 ***	0.248 ***	0.150 ***	0.101 *	0.101 **	0.139 ***
	(0.021)	(0.073)	(0.049)	(0.051)	(0.039)	(0.046)
Sample size	83,906	16,791	16,782	16,864	16,702	16,767
Policy Multiplier	11.0%	62.6%	14.4%	6.2%	8.1%	7.8%

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 12B
Child Sample OLS Regression Results
By PUMA Poverty Level Quintile

Fixed Effects?	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language
Sample	Full Sample	Lowest Quintile	2	3	4	Highest Quintile
Core Sample (Dependent Variable=Medicaid)						
Medicaid Network	0.071 ***	0.083	0.020	0.107 **	0.078	0.103 **
	(0.024)	(0.103)	(0.047)	(0.044)	(0.053)	(0.046)
Sample size	136,542	27,526	27,102	27,307	27,308	27,299
Policy Multiplier	7.4%	N/A	N/A	9.2%	N/A	6.7%
Expanded Sample (Dependent Variable=Insured)						
Medicaid Network	0.100 ***	0.211 ***	0.111 ***	0.134	0.081 **	0.071 ***
	(0.015)	(0.045)	(0.035)	(0.037)	(0.032)	(0.026)
Sample size	192,414	38,499	38,493	38,602	38,347	38,473
Policy Multiplier	10.4%	50.5%	11.9%	11.6%	6.5%	4.9%

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 13A
Adult Sample OLS Regression Results
By PUMA Geographic Size Quintile

Fixed Effects?	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language
Sample	Full Sample	Lowest Quintile	2	3	4	Highest Quintile
Core Sample (Dependent Variable=Medicaid)						
Medicaid Network	0.100 ***	0.069 *	0.147 ***	0.166 ***	0.077	-0.062
	(0.020)	(0.038)	(0.054)	(0.056)	(0.051)	(0.051)
Sample size	59,377	11,901	12,081	11,645	11,886	11,864
Policy Multiplier	9.9%	12.7%	20.8%	13.2%	N/A	N/A
Expanded Sample (Dependent Variable=Insured)						
Medicaid Network	0.118 ***	0.123 ***	0.145 ***	0.212 ***	0.067	-0.020
	(0.021)	(0.036)	(0.055)	(0.057)	(0.056)	(0.074)
Sample size	83,906	16,829	16,796	16,750	16,759	16,772
Policy Multiplier	11.0%	24.0%	19.4%	18.2%	N/A	N/A

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 13B
Child Sample OLS Regression Results
By PUMA Geographic Size Quintile

Fixed Effects?	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language
Sample	Full Sample	Lowest Quintile	2	3	4	Highest Quintile
Core Sample (Dependent Variable=Medicaid)						
Medicaid Network	0.071 ***	0.011	0.045	0.194 ***	0.05	0.063
	(0.024)	(0.050)	(0.072)	(0.069)	(0.055)	(0.040)
Sample size	136,542	27,328	27,377	27,582	26,960	27,295
Policy Multiplier	7.4%	N/A	N/A	21.5%	N/A	N/A
Expanded Sample (Dependent Variable=Insured)						
Medicaid Network	0.100 ***	0.073 ***	0.112 ***	0.127 ***	0.087 **	0.159 ***
	(0.015)	(0.026)	(0.034)	(0.037)	(0.043)	(0.038)
Sample size	192,414	38,611	38,760	38,582	38,062	38,399
Policy Multiplier	10.4%	13.3%	15.4%	13.0%	3.5%	10.9%

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 14A
Adult Sample OLS Regression Results
By PUMA Foreign Born Quintile

Fixed Effects?	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language
Sample	Full Sample	Lowest Quintile	2	3	4	Highest Quintile
Core Sample (Dependent Variable=Medicaid)						
Medicaid Network	0.100 ***	0.032	0.048	0.067	0.044	0.137 ***
	(0.020)	(0.043)	(0.050)	(0.068)	(0.048)	(0.052)
Sample size	59,377	11,909	11,912	11,870	11,875	11,811
$\ln CA_k$ Mean	0.89	0.16	0.27	0.95	1.36	1.75
Policy Multiplier	9.9%	N/A	N/A	N/A	N/A	31.5%
Expanded Sample (Dependent Variable=Insured)						
Medicaid Network	0.118 ***	-0.025	0.094 *	0.095	0.080 *	0.102 **
	(0.021)	(0.058)	(0.054)	(0.063)	(0.044)	(0.045)
Sample size	83,906	16,819	16,780	16,779	16,751	16,777
$\ln CA_k$ Mean	0.85	0.05	0.26	0.73	1.28	1.73
Policy Multiplier	11.0%	N/A	2.5%	N/A	11.4%	21.6%

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 14B
Child Sample OLS Regression Results
By PUMA Foreign Born Quintile

Fixed Effects?	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language
Sample	Full Sample	Lowest Quintile	2	3	4	Highest Quintile
Core Sample (Dependent Variable=Medicaid)						
Medicaid Network	0.071 ***	0.074 **	0.280 ***	-0.003	-0.035	-0.031
	(0.024)	(0.033)	(0.073)	(0.076)	(0.100)	(0.058)
Sample size	136,542	27,310	27,358	27,320	27,320	27,234
Policy Multiplier	7.4%	0.7%	17.7%	N/A	N/A	N/A
Expanded Sample (Dependent Variable=Insured)						
Medicaid Network	0.100 ***	0.099 **	0.130 ***	0.117 **	0.010	0.041
	(0.015)	(0.039)	(0.042)	(0.051)	(0.048)	(0.031)
Sample size	192,414	38,494	38,563	38,563	38,336	38,458
Policy Multiplier	10.4%	0.0%	7.3%	14.7%	N/A	N/A

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 15A
Adult Sample OLS Regression Results
By State Medicaid Fee Generosity Quintile and Multiple Language Outreach States

Fixed Effects?	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language
Sample	Full Sample	Lowest Quintile	2	3	4	Highest Quintile	Outreach States	Non-Outreach States
Core Sample (Dependent Variable=Medicaid)								
Medicaid Network	0.100 ***	0.053	0.211 ***	0.123 ***	0.068	0.004	0.120 ***	0.0798 ***
	(0.020)	(0.034)	(0.073)	(0.042)	(0.044)	(0.124)	(0.028)	(0.029)
Sample size	59,377	11,953	11,857	12,011	13,355	10,201	41,056	18,321
Policy Multiplier Effect	9.9%	N/A	39.5%	5.0%	N/A	N/A	12.8%	6.7%
Expanded Sample (Dependent Variable=Insured)								
Medicaid Network	0.118 ***	0.127 ***	0.284 ***	0.023	0.015	0.137	0.126 ***	0.119 ***
	(0.021)	(0.034)	(0.087)	(0.051)	(0.053)	(0.119)	(0.031)	(0.031)
Sample size	83,906	30,279	8,537	11,903	18,807	14,380	57,679	26,227
Policy Multiplier Effect	11.0%	20.0%	14.8%	N/A	N/A	N/A	12.7%	9.6%

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 15B
Child Sample OLS Regression Results
By State Medicaid Fee Generosity Quintile and Multiple Language Outreach States

Fixed Effects?	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language
Sample	Full Sample	Lowest Quintile Fee State	2	3	4	Highest Quintile Fee State	Outreach States	Non-Outreach States
Core Sample (Dependent Variable=Medicaid)								
Medicaid Network	0.071 ***	-0.005	-0.065	0.056	0.193 **	0.112 *	0.042	0.104 ***
	(0.024)	(0.037)	(0.209)	(0.038)	(0.086)	(0.063)	(0.032)	(0.036)
Sample size	136,542	53,763	8,605	39,177	7,924	27,073	110,004	26,538
Policy Multiplier Effect	7.4%	N/A	N/A	N/A	1.5%	7.4%	N/A	9.4%
Expanded Sample (Dependent Variable=Insured)								
Medicaid Network	0.100 ***	0.034 *	N/A	0.108 ***	0.157 ***	0.102 ***	0.102 ***	0.101 ***
	(0.015)	(0.020)		(0.036)	(0.051)	(0.034)	(0.022)	(0.022)
Sample size	192,414	80,297	0	62,140	12,152	37,825	149,360	43,054
Policy Multiplier Effect	10.4%	4.6%	N/A	9.9%	2.6%	6.6%	11.0%	9.1%

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 16A
Adult Sample OLS Regression Results
Network Variable Specification Checks

Fixed Effects	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language
Specification	Full Sample	Survey Weights	Take-Up @ State-Level	Take-Up @ Super-PUMA Level	Remove lnCA Denom.
Core Sample (Dependent Variable=Medicaid)					
Medicaid Network	0.100 ***	0.117 ***	0.131 ***	0.191 ***	0.101 ***
	(0.020)	(0.024)	(0.011)	(0.008)	(0.020)
Sample size	59,377	59,377	59,377	59,377	59,377
Expanded Sample (Dependent Variable=Insured)					
Medicaid Network	0.118 ***	0.133 ***	0.081 ***	0.093 ***	0.130 ***
	(0.021)	(0.025)	(0.011)	(0.007)	(0.020)
Sample size	83,906	83,906	83,906	83,906	83,906

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 16B
Child Sample OLS Regression Results
Network Variable Specification Checks

Fixed Effects	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language
Specification	Full Sample	Survey Weights	Take-Up @ State-Level	Take-Up @ Super-PUMA Level	Remove lnCA Denom.
Core Sample (Dependent Variable=Medicaid)					
Medicaid Network	0.071 ***	0.085 ***	0.184 ***	0.246 ***	0.073 ***
	(0.024)	(0.029)	(0.011)	(0.007)	(0.024)
Sample size	136,542	136,542	136,542	136,542	136,542
Expanded Sample (Dependent Variable=Insured)					
Medicaid Network	0.100 ***	0.129 ***	0.077 ***	0.078 ***	0.100 ***
	(0.015)	(0.017)	(0.008)	(0.005)	(0.015)
Sample size	192,414	192,414	192,414	192,414	192,414

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 17A
Adult Sample Non-Linear Binary Models

Area Fixed Effects or Characteristics?	PUMA Characteristics	Super-PUMA FE	PUMA Characteristics	Super-PUMA FE	PUMA Characteristics	Super-PUMA FE
Language Group Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes
Model	OLS	OLS	Logit	Logit	Probit	Probit
Core Sample (Dependent Variable=Medicaid)						
Medicaid Network	0.092 ***	0.089 ***	0.604 ***	0.522 ***	0.318 ***	0.276 ***
	(0.022)	(0.019)	(0.137)	(0.122)	(0.078)	(0.069)
Sample size	59,377	59,377	59,377	59,177	59,377	59,377
Expanded Sample						
Medicaid Network	0.099 ***	0.104 ***	0.739 ***	0.714 ***	0.390 ***	0.383 ***
	(0.024)	(0.011)	(0.150)	(0.129)	(0.085)	(0.073)
Sample size	83,906	83,906	83,906	83,814	83,906	83,814

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 17B
Adult Sample Non-Linear Binary Models

Area Fixed Effects or Characteristics?	PUMA Characteristics	Super-PUMA FE	PUMA Characteristics	Super-PUMA FE	PUMA Characteristics	Super-PUMA FE
Language Group Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes
Model	OLS	OLS	Logit	Logit	Probit	Probit
Core Sample (Dependent Variable=Medicaid)						
Medicaid Network	0.128 ***	0.071 ***	1.149 ***	0.748 ***	0.595 ***	0.376 ***
	(0.027)	(0.021)	(0.184)	(0.155)	(0.107)	(0.082)
Sample size	136,542	136,542	136,542	136,364	136,542	136,364
Expanded Sample						
Medicaid Network	0.127522 ***	0.089 ***	0.699 ***	0.672 ***	0.353 ***	0.371 ***
	(0.027)	(0.015)	(0.148)	(0.135)	(0.080)	(0.071)
Sample size	192,414	192,414	192,414	192,105	192,414	192,105

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 18A
Multinomial Logit Regressions
Expanded Adult Sample

Area Fixed Effects or Characteristics?	PUMA Characteristics	PUMA Characteristics	PUMA Characteristics
Language Group Fixed Effects?	Yes	Yes	Yes
Check	NG=PHI	NG=Medicaid	NG=Separate Category
Expanded Sample: Medicaid Network Variable			
Medicaid Network: PHI vs. Medicaid	-0.351 ** (0.024)	-0.617 *** (0.131)	N/A
Medicaid Network: ESI vs. Medicaid	N/A	N/A	-0.492 *** (0.024)
Medicaid Network: Nongroup vs. Medicaid	N/A	N/A	-0.117 (0.191)
Medicaid Network: Uninsured vs. Medicaid	-0.736 *** (0.174)	-0.875 *** (0.159)	-0.720 *** (0.174)
Sample size	83,906	83,906	83,906

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 18B
Multinomial Logit Regressions
Expanded Child Sample

Area Fixed Effects or Characteristics?	PUMA Characteristics	PUMA Characteristics	PUMA Characteristics	Super-PUMA Fixed Effects
Language Group Fixed Effects?	Yes	Yes	Yes	Yes
Check	NG=PHI	NG=Medicaid	NG=Separate Category	NG=Separate Category
Expanded Sample: Medicaid Network Variable				
Medicaid Network: PHI vs. Medicaid	0.079 (0.108)	-0.183 * (0.103)	N/A	N/A
Medicaid Network: ESI vs. Medicaid	N/A	N/A	-0.021 (0.114)	0.079 (0.117)
Medicaid Network: Nongroup vs. Medicaid	N/A	N/A	0.353 ** (0.145)	0.287 * (0.151)
Medicaid Network: Uninsured vs. Medicaid	-0.558 *** (0.156)	-0.721 *** (0.151)	-0.540 *** (0.156)	-0.574 *** (0.146)
Sample size	192,414	192,414	192,414	192,414

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 19A
Adult Sample OLS Regression Results
Placebo Tests: Changing Left Hand Side Variable

Fixed Effects	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language	PUMA, language
LHS Variable	Medicaid Take-Up	<100% FPL	100-200% FPL	200-300% FPL	Age	Welfare Take-Up	Married
Core Sample (Dependent Variable=Medicaid)							
Medicaid Network	0.100 ***	0.035 *	-0.028	-0.007	-0.361	0.011	0.086 ***
	(0.020)	(0.018)	(0.018)	(0.006)	(0.432)	(0.008)	(0.018)
Sample size	59,377	59,377	59,377	59,377	59,377	59,377	59,377
Expanded Sample (Dependent Variable=Insured)							
Medicaid Network	0.118 ***	0.044 **	-0.040 *	-0.004	-0.428	0.026 ***	0.033 *
	(0.021)	(0.020)	(0.020)	(0.009)	(0.467)	(0.008)	(0.018)
Sample size	83,906	83,906	83,906	83,906	83,906	83,906	83,906

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 19B
Child Sample OLS Regression Results
Placebo Tests: Changing Left Hand Side Variable

Fixed Effects	PUMA, language		PUMA, language	PUMA, language		PUMA, language		PUMA, language	
LHS Variable	Medicaid Take-Up		<100% FPL	100-200% FPL		200-300% FPL		Welfare Take-Up	
Core Sample (Dependent Variable=Medicaid)									
Medicaid Network	0.071	***	0.040	0.008		-0.042	**	-0.001	
	(0.024)		(0.030)	(0.030)		(0.017)		(0.013)	
Sample size	136,542		136,542	136,542		136,542		136,542	
Expanded Sample (Dependent Variable=Insured)									
Medicaid Network	0.100	***	0.034 *	0.017		-0.030	**	-0.025	**
	(0.015)		(0.019)	(0.020)		(0.015)		(0.011)	
Sample size	192,414		192,414	192,414		192,414		192,414	

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 20A
Adult Sample OLS Regression Results
Quintile Take-Up Regressions

Fixed Effects	PUMA, language		PUMA, language	
Level for Language Group Take-Up	State Take-Up		Super-PUMA Take-Up	
Core Sample (Dependent Variable=Medicaid)				
lnCA*Take-up quintile 1	-0.027	**	-0.082	***
	(0.013)		(0.025)	
lnCA*Take-up quintile 2	-0.003		-0.014	
	(0.013)		(0.018)	
lnCA*Take-up quintile 4	-0.004		-0.043	
	(0.012)		(0.027)	
lnCA*Take-up quintile 5	0.033	***	-0.004	
	(0.010)		(0.020)	
Sample size	54,254		44,415	
Expanded Sample (Dependent Variable=Insured)				
lnCA*Take-up quintile 1	-0.016	**	-0.059	***
	(0.008)		(0.011)	
lnCA*Take-up quintile 2	-0.011		-0.014	
	(0.007)		(0.009)	
lnCA*Take-up quintile 4	0.007		-0.030	***
	(0.006)		(0.008)	
lnCA*Take-up quintile 5	0.014	**	-0.006	
	(0.007)		(0.013)	
Sample size	77,459		61,574	

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates. (3) Small sample size area-language cells (<30) are removed from the sample.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Table 20B
Child Sample OLS Regression Results
Quintile Take-Up Regressions

Fixed Effects	PUMA, language	PUMA, language
Level for Language Group Take-Up	State Take-Up	Super-PUMA Take-Up
Core Sample (Dependent Variable=Medicaid)		
lnCA*Take-up quintile 1	-0.031 ***	-0.001
	(0.008)	(0.015)
lnCA*Take-up quintile 2	0.005	0.006
	(0.030)	(0.017)
lnCA*Take-up quintile 4	0.005	0.020
	(0.013)	(0.016)
lnCA*Take-up quintile 5	0.015 **	0.035 ***
	(0.006)	(0.013)
Sample size	129,665	115,669
Expanded Sample (Dependent Variable=Insured)		
lnCA*Take-up quintile 1	0.009	-0.017 ***
	(0.011)	(0.004)
lnCA*Take-up quintile 2	0.024 **	0.008
	(0.011)	(0.005)
lnCA*Take-up quintile 4	0.026 **	0.012 ***
	(0.012)	(0.004)
lnCA*Take-up quintile 5	0.023 ***	0.011 **
	(0.011)	(0.005)
Sample size	183,583	160,149

Source: 2008-2009 American Community Surveys
Notes: (1) Robust and cluster corrected standard errors are in parenthesis. (2) See full regression table for a list of all other covariates. (3) Small sample size area-language cells (<30) are removed from the sample.
*** The coefficient is significant at the 1% level; ** 5% level; * 10% level.

Appendix 1A
Descriptive Statistics
Expanded Adult Sample, by Health Insurance Status

	Employer Sponsored Insurance	Private Non-Group	Medicaid	Uninsured
Unweighted N	19,008	5,521	22,980	36,397
Weighted N	2,375,098	655,714	2,843,547	4,911,734
Foreign born	70.7%	73.7%	69.0%	81.9%
Fluent in English	74.1%	77.6%	62.1%	49.7%
Non-Citizen	40.9%	46.8%	36.1%	67.6%
MSA Status				
Non-MSA	9.5%	8.6%	6.9%	10.7%
MSA not identifiable	4.5%	4.1%	2.5%	3.7%
MSA, central city	28.4%	36.7%	43.7%	27.3%
MSA, outside central city	27.9%	24.6%	21.4%	26.1%
MSA, central city status unknown	29.8%	26.0%	25.6%	32.1%
Female	52.4%	53.9%	65.0%	49.8%
Age				
Age, 19-24	16.9%	28.3%	13.2%	14.7%
Age, 25-34	26.1%	22.9%	26.4%	34.4%
Age, 35-44	31.1%	20.8%	27.6%	29.3%
Age, 45-54	17.7%	16.7%	20.1%	14.9%
Age, 55-64	8.1%	11.4%	12.8%	6.7%
Education				
< High school	28.4%	19.9%	46.6%	52.0%
High school graduate	26.8%	19.6%	26.8%	26.2%
Some college	27.8%	32.0%	20.1%	14.7%
College+	17.0%	28.5%	6.5%	7.1%

Race and Ethnicity				
White, non-Hispanic	20.2%	27.2%	15.7%	10.3%
Black, non-Hispanic	5.1%	4.4%	4.5%	2.8%
Asian, non-Hispanic	17.5%	35.4%	13.3%	8.7%
Hispanic	55.1%	31.2%	63.6%	75.9%
Other and multiple races	2.1%	1.8%	2.9%	2.3%
Married	55.4%	42.5%	45.8%	53.7%
Family size	3.6	2.7	3.8	3.7
Number of own children in family	1.6	1.0	1.7	1.6
Income Relative to Poverty				
<=100% FPL	38.1%	66.3%	74.5%	65.3%
101-200% FPL	47.5%	27.7%	22.4%	31.5%
201-300% FPL	14.4%	5.9%	3.1%	3.2%
Has Welfare Income	1.3%	2.3%	14.5%	2.0%
Number of disabilities	0.11	0.15	0.46	0.10
Year 2009	49.5%	51.1%	54.6%	52.9%
Work Status				
Worker, not self-employed	81.7%	59.5%	55.9%	64.2%
Worker, self-employed	4.2%	12.1%	6.8%	10.1%
Non-worker	14.2%	28.4%	37.3%	25.7%

Source: 2008-2009 American Community Surveys

Appendix 1B
Descriptive Statistics
Expanded Child Sample, by Health Insurance Status

	Employer Sponsored Insurance	Private Non-Group	Medicaid	Uninsured
Unweighted N	46,064	9,808	98,348	38,194
Weighted N	5,500,313	1,087,958	12,497,181	4,962,311
Foreign born	12.8%	21.0%	11.0%	33.2%
Fluent in English	73.8%	74.1%	62.6%	72.5%
Non-Citizen	8.2%	15.8%	7.7%	30.1%
MSA Status				
Non-MSA	7.4%	6.7%	7.5%	10.3%
MSA not identifiable	2.6%	2.2%	2.3%	3.3%
MSA, central city	25.2%	27.9%	30.0%	19.7%
MSA, outside central city	35.0%	33.6%	26.9%	30.0%
MSA, central city status unknown	29.8%	29.6%	33.4%	36.8%
Female	48.9%	49.6%	49.0%	48.2%
Age				
Infant	4.2%	3.9%	6.6%	2.7%
Age, 1-5	24.0%	22.1%	32.1%	20.2%
Age, 6-19	71.8%	74.1%	61.3%	77.1%
Number Family Members with At Least Some College				
0	41.4%	43.7%	65.9%	66.9%
1	33.7%	33.4%	24.7%	22.6%
2+	24.9%	23.0%	9.5%	10.5%
Race and Ethnicity				
White, non-Hispanic	16.9%	20.0%	8.7%	9.4%

Black, non-Hispanic	5.7%	4.0%	3.5%	3.1%
Asian, non-Hispanic	12.4%	20.6%	6.6%	6.1%
Hispanic	62.2%	52.5%	79.4%	79.7%
Other and multiple races	2.8%	2.9%	1.8%	1.7%
Family size	4.71	4.34	4.96	4.94
Two-Parent Family	71.5%	67.3%	63.3%	69.4%
Income Relative to Poverty				
<=100% FPL	19.4%	34.3%	53.7%	43.8%
101-200% FPL	50.9%	46.2%	39.3%	48.2%
201-300% FPL	25.0%	16.8%	6.4%	7.5%
301%+	4.7%	2.6%	0.5%	0.6%
Number of Workers in Family				
0	2.7%	7.6%	7.9%	5.1%
1	39.8%	42.1%	45.7%	41.9%
2+	57.6%	50.3%	46.5%	53.0%
Number of disabilities	0.039	0.043	0.067	0.028
Year 2009	48.9%	49.5%	54.0%	49.0%

Source: 2008-2009 American Community Surveys

Appendix 2A
Select Weighted Descriptive Statistics
Expanded Adult Sample by Language Group

	Sample Size	Weighted Sample	Medicaid	Any Private	Uninsured	Age	<=100% FPL	< High School	Family Size	Foreign Born	Non-citizen
Full Sample	83,906	10,786,093	0.26	0.28	0.46	37.0	0.62	0.43	3.6	0.76	0.52
Language Group											
Yiddish, Jewish	687	67,279	0.69	0.27	0.04	34.8	0.80	0.24	6.3	0.18	0.05
Armenian	206	26,619	0.54	0.21	0.25	41.3	0.85	0.13	3.4	0.94	0.53
Cantonese	764	84,731	0.47	0.36	0.16	42.6	0.65	0.50	3.6	0.90	0.43
Bengali	480	57,818	0.44	0.34	0.22	38.9	0.65	0.28	4.3	0.98	0.48
Mon-Khmer, Cambodian	482	59,796	0.41	0.35	0.24	39.4	0.45	0.46	4.4	0.86	0.39
Arabic	1,555	210,711	0.40	0.29	0.31	37.2	0.71	0.21	3.9	0.86	0.48
Miao, Hmong	554	80,086	0.40	0.39	0.21	34.4	0.50	0.40	5.6	0.80	0.34
Persian	341	41,771	0.36	0.36	0.28	39.1	0.68	0.11	3.4	0.89	0.47
Vietnamese	1,749	204,477	0.35	0.33	0.31	40.4	0.56	0.43	3.7	0.94	0.38
Russian	1,299	166,864	0.34	0.33	0.33	39.1	0.59	0.11	3.1	0.91	0.49
Urdu	532	68,863	0.34	0.26	0.40	40.2	0.56	0.19	4.6	0.95	0.41
French or Haitian Creole	825	117,605	0.32	0.40	0.28	38.1	0.53	0.25	3.6	0.86	0.47
Navajo	1,485	112,751	0.30	0.14	0.56	39.3	0.65	0.33	3.7	0.00	0.00
Hebrew, Israeli	335	35,495	0.28	0.55	0.17	36.2	0.74	0.09	3.5	0.50	0.21
Ukrainian	281	30,698	0.28	0.32	0.40	39.2	0.41	0.10	4.0	0.93	0.64
Spanish	56,329	7,506,950	0.26	0.23	0.52	36.4	0.64	0.52	3.8	0.75	0.56
Laotian	227	32,966	0.26	0.44	0.30	37.7	0.41	0.40	3.8	0.82	0.39
Portuguese	1,521	211,259	0.25	0.42	0.33	37.4	0.30	0.33	2.8	0.84	0.61
Amharic, Ethiopian, etc.	275	38,793	0.25	0.39	0.36	37.2	0.54	0.23	2.9	0.96	0.52
Turkish	118	14,819	0.25	0.47	0.28	35.7	0.36	0.21	3.4	0.93	0.67
Mandarin	633	80,850	0.25	0.43	0.32	36.8	0.69	0.28	2.6	0.89	0.64
Chinese	2,393	280,153	0.24	0.48	0.27	38.1	0.64	0.27	2.9	0.92	0.56
Greek	311	34,457	0.24	0.54	0.22	42.5	0.46	0.22	2.8	0.50	0.21
Albanian	148	23,533	0.24	0.41	0.35	38.1	0.32	0.14	4.3	0.88	0.48
Hungarian	34	3,783	0.23	0.52	0.25	41.8	0.61	0.29	3.1	0.73	0.37
Polish	670	79,119	0.22	0.39	0.39	41.1	0.41	0.10	2.9	0.83	0.48
Italian	675	73,269	0.21	0.56	0.23	42.1	0.48	0.17	2.6	0.39	0.16
French	1,609	193,527	0.21	0.50	0.29	38.3	0.58	0.10	2.4	0.51	0.36
Serbo-Croatian	176	23,235	0.21	0.54	0.25	39.4	0.42	0.22	3.4	0.94	0.41
Kru	422	59,220	0.21	0.39	0.41	37.4	0.54	0.12	3.2	0.93	0.60
Hindi and Punjabi	827	107,837	0.20	0.45	0.35	37.3	0.62	0.20	3.3	0.95	0.62
Rumanian	176	24,332	0.20	0.35	0.45	36.2	0.56	0.20	3.1	0.82	0.36
Filipino, Tagalog	836	91,792	0.16	0.54	0.30	40.6	0.47	0.09	3.3	0.92	0.45
German	1,321	126,456	0.15	0.46	0.38	37.0	0.58	0.24	3.1	0.31	0.18
Guajarati	311	41,379	0.14	0.49	0.37	39.5	0.57	0.13	3.4	0.85	0.36
Korean	1,624	194,603	0.11	0.51	0.38	37.5	0.69	0.05	2.7	0.91	0.64
Thai	192	23,956	0.11	0.57	0.32	36.6	0.67	0.19	2.5	0.90	0.61
Tamil, Malayalam and Telugu	248	30,056	0.10	0.72	0.18	32.7	0.59	0.05	2.8	0.94	0.68
Dutch	242	21,738	0.08	0.39	0.52	36.2	0.48	0.46	4.2	0.35	0.24
Japanese	544	65,670	0.06	0.69	0.24	33.7	0.67	0.03	2.1	0.75	0.65
Pennsylvania Dutch	469	36,777	0.04	0.10	0.86	36.4	0.42	0.83	6.5	0.00	0.00

Source: 2008-2009 American Community Surveys

Appendix 2B
Select Weighted Descriptive Statistics
Expanded Child Sample by Language Group

	Sample Size	Weighted Sample	Medicaid/ CHIP	Any Private	Uninsured	Age	<=100% FPL	Family Size	Foreign Born	Non-citizen
Full Sample	192,414	24,047,763	0.52	0.27	0.21	8.9	0.43	4.9	0.16	0.13
Language Group										
Albanian	388	60,607	0.43	0.49	0.07	9.1	0.25	4.9	0.29	0.15
Amharic, Ethiopian, etc.	490	61,380	0.44	0.46	0.10	8.5	0.33	4.4	0.28	0.21
Arabic	2,983	411,786	0.57	0.30	0.13	8.3	0.49	5.3	0.29	0.18
Armenian	359	49,190	0.55	0.32	0.13	10.5	0.44	4.0	0.33	0.27
Bengali	729	95,104	0.54	0.34	0.11	8.4	0.41	4.9	0.34	0.15
Cantonese	1,023	106,952	0.56	0.37	0.07	11.0	0.37	4.5	0.24	0.16
Chinese	2,885	303,065	0.39	0.47	0.15	9.8	0.36	4.2	0.25	0.19
Dutch	594	44,893	0.11	0.37	0.52	7.9	0.38	6.2	0.07	0.04
Filipino, Tagalog	2,538	254,255	0.28	0.62	0.10	9.3	0.20	4.7	0.31	0.19
French	2,494	272,810	0.39	0.51	0.11	10.2	0.34	4.2	0.15	0.11
French or Haitian Creole	2,453	334,939	0.40	0.34	0.26	9.2	0.37	4.8	0.24	0.19
German	2,636	225,846	0.24	0.40	0.37	9.6	0.40	5.5	0.12	0.07
Greek	368	41,905	0.30	0.58	0.12	10.2	0.23	4.4	0.06	0.01
Guajarati	471	54,980	0.25	0.56	0.19	10.3	0.32	4.3	0.31	0.24
Hebrew, Israeli	743	82,349	0.32	0.64	0.04	8.5	0.32	5.7	0.15	0.07
Hindi and Punjabi	1,180	149,298	0.38	0.42	0.20	9.4	0.26	4.8	0.31	0.19
Hungarian	92	9,915	0.41	0.56	0.03	10.2	0.37	3.8	0.15	0.10
Italian	1,014	106,865	0.23	0.69	0.08	10.4	0.22	4.3	0.08	0.03
Japanese	684	72,606	0.21	0.70	0.09	8.7	0.29	3.8	0.29	0.17
Korean	2,202	244,026	0.25	0.47	0.27	10.1	0.36	3.8	0.39	0.34
Kru	950	117,180	0.42	0.44	0.14	7.9	0.28	4.9	0.29	0.19
Laotian	411	51,518	0.39	0.46	0.16	9.1	0.30	4.6	0.07	0.04
Ilocano, Hocano	210	18,148	0.21	0.71	0.08	9.5	0.10	6.1	0.30	0.21
Mandarin	619	73,816	0.40	0.50	0.10	10.7	0.42	3.8	0.30	0.22
Miao, Hmong	1,334	168,254	0.52	0.42	0.06	9.9	0.38	6.8	0.19	0.13
Mon-Khmer, Cambodian	784	87,071	0.47	0.38	0.15	9.7	0.34	5.1	0.08	0.05
Navajo	1,238	95,034	0.52	0.16	0.31	9.9	0.59	5.2	0.00	0.00
Other American Indian	418	35,972	0.52	0.17	0.31	9.8	0.55	5.1	0.08	0.06
Pennsylvania Dutch	1,251	94,189	0.08	0.11	0.81	7.9	0.36	8.2	0.00	0.00
Persian	574	69,702	0.47	0.42	0.11	9.6	0.38	4.5	0.26	0.18
Polish	779	86,646	0.40	0.49	0.10	9.0	0.24	4.0	0.19	0.11
Portuguese	1,345	170,154	0.37	0.45	0.19	9.2	0.26	3.9	0.27	0.23
Rumanian	276	36,007	0.44	0.39	0.17	8.4	0.27	5.4	0.20	0.14
Russian	1,809	208,440	0.45	0.41	0.14	8.9	0.33	4.8	0.36	0.23
Serbo-Croatian	295	40,567	0.38	0.56	0.06	9.6	0.20	4.2	0.42	0.25
Spanish	146,208	18,889,927	0.55	0.23	0.22	8.7	0.45	4.9	0.15	0.12
Tamil, Malayalam and Telugu	309	34,356	0.14	0.74	0.12	8.8	0.25	4.2	0.46	0.39
Thai	253	28,442	0.31	0.50	0.19	9.7	0.34	4.0	0.29	0.24
Turkish	307	36,064	0.37	0.46	0.17	8.8	0.27	4.3	0.44	0.39
Ukrainian	527	54,272	0.58	0.30	0.12	9.4	0.23	6.1	0.46	0.37
Urdu	1,180	145,359	0.56	0.26	0.19	9.0	0.41	5.3	0.31	0.17
Vietnamese	3,443	363,927	0.49	0.37	0.14	9.2	0.35	4.6	0.20	0.11
Yiddish, Jewish	1,568	159,947	0.71	0.28	0.01	7.7	0.61	7.5	0.01	0.00

Source: 2008-2009 American Community Surveys

Chapter 6: Conclusion

Summarizing Research Objectives

This dissertation explores how group behavior influences individual economic decision-making in the context of health insurance choice. Broadly speaking, any paper that analyzes group behavior effects (e.g., network effects, peer effects, social interactions) ought to address four key areas. First, the study must define a relevant or interesting economic outcome. Medicaid take-up rates are below 100% and there is considerable uncertainty over why this is the case. This study provides insight into why some groups have lower take-up rates than others. In addition, the Medicaid take-up process is complicated (e.g. eligibility pathways that vary by income, age, and geographic location) and there is uncertainty over the benefits that certain populations might face (e.g., variation in covered benefits, payment rates, quality of care, access, etc…). Such an uncertain environment is conducive to non-market interactions among social contacts and is an interesting study for measure group effects

Second, the study must define the each individual's group as precisely as possible. Some studies, such as those that look at peer effects in high school cohorts or roommate effects in college, use rich data to precisely measure group behavior effects and social interactions. One major limitation of this study is that I do not have available data at this precise of a level. In contrast, I define each individual's group of potential contacts by using non-English language spoken at home and geographic PUMA of residence. Given this assumption, careful consideration is needed to determine how these groups operate and why they serve as a good proxy for the social interactions between individuals.

Third, the study ought to determine how the defined group can influence individual behavior. For this study, I chose language groups because previous sociological and economic studies show that persons who speak a non-English language at home interact mainly with others who speak that language. It is also likely that these individuals, especially the foreign born population, are less likely to have specific capital about the U.S health care system (relative to their English-speaking counterparts) and are more likely to rely on their language group for information related to Medicaid. While PUMAs are not the most precise geographic measures, they are the most detailed than anything available in comparable surveys, such as the CPS and MEPS. I also develop an economic model that shows how group behavior alters the individual's expected utility between taking-up Medicaid and being uninsured. This model shows how each person receives information through the available contacts in each person's local area. However, one limitation of this model is that it does not predict the specific type of information that language groups transfer. For example, groups can pass along information related to the existence or eligibility rules associated with Medicaid, and/or they can pass along information related to the relative benefits of Medicaid compared to being uninsured.

Finally, and most importantly, the study must empirically identify the *causal* effect of group behavior on individual outcomes. The empirical framework for this study is well-established (Bertrand et al. 2000; Deri 2005) and utilizes several important components. First, the left-hand side (LHS) variable in the core model is a 0-1 indicator for having Medicaid vs. being uninsured. I also use an expanded sample that includes those with private health insurance and define the LHS variable as a 0-1 indicator for having any insurance. Second, the main right-hand side variable interest is *an interaction term* between two continuous variables:

(1) The Medicaid take-up rate of the individual's language group. For the core model, for each individual, this variable is defined as number of persons in their common language group that are enrolled in Medicaid divided by the number of persons in the common language group that are either enrolled in Medicaid or are eligible for Medicaid but are uninsured . This is a standard definition of Medicaid take-up or participation. The concept being this variable is that it serves as a proxy for language group *quality*: language groups with higher take-up rates possess more knowledge or information related to Medicaid coverage or have a higher valuation of Medicaid relative to being uninsured.

(2) The proportion of the person's local area that are part of the person's common language group. This variable serves as a proxy the *contact availability* for each person. This variable captures the geographic component of group behavior.

What exactly does this interaction term mean? Intuitively, for an individual that is part of a high Medicaid/CHIP take-up language group (e.g., above the mean), living among a high concentration of his/her language group can increase the person's probability of taking-up Medicaid. For example, these potential contacts can provide information related to the benefits of enrollment relative to being uninsured. In contrast, for those that are part of a low take-up group (e.g., below the mean), living among a high concentration of the language group can decrease the person's probability of taking-up Medicaid. These potential contacts might believe that costs of enrollment outweigh the benefits (e.g., it is more convenient to remain uninsured and utilize necessary care from safety net providers), and could discourage the individual from enrolling. It is also possible that living among a high concentration of the language group can increase the probability of take-up, regardless if the person is from a low or high take-up group. However, the differential effect on the probability of take-up will be larger among those that are part of a high take-up group, as these groups might possess more practical knowledge (e.g., information related to eligibility and necessary documentation) that could help the individual enroll in Medicaid. In other words, this study analyzes the differential effect of living in areas of high concentration of a common language group on an individual's probability of taking-up Medicaid.

One of the main advantages of the interaction term is that the model can include dummy variables for each language group and for each PUMA (fixed effects), controlling for omitted language group and local area characteristics that could be correlated with main variable of interest. The model also allows the researcher to directly control for the contact availability for each individual. The "naïve" regression results in Chapter 5 show that models that fail to control for local area or language group characteristics will create group behavior coefficients that are biased upwards. However, it is possible that there are some remaining omitted variables that are correlated with the interaction term. For example, there is potential for differential geographic sorting, where people who live in areas of high concentration of their language group are different in some unobservable way from people who live in low concentration areas, but in a way that is correlated with Medicaid take-up (e.g., sorting based on health status). It is also possible omitted outreach effects that are not captured by the language group and local are dummies, could partially explain the results. For example, it is high concentrations of a Medicaid/CHIP utilizing language group in a local area could cause a school district or a Medicaid/CHIP office to implement policies that increase ease of enrollment. This effect would be capture by the network variable in the main model. Various sensitivity models in Chapter 5 address these concerns, and the main results are summarized below.

Main Empirical Results

For both adults and children, I find positive and statistically significant coefficients on the main variable of interest across multiple model specifications and sample restrictions. For the core sample, which is limited to Medicaid/CHIP eligibles without private health insurance, the coefficient on the network/group variable is .100 among the adult sample and .071 among the child sample, both of which are statistically significant at the 1% level. When those with private coverage are included in the sample (and the dependent variable is a 0-1 indicator for any coverage), the coefficients for the adult and child sample are .118 and .100, respectively, and remain statistically significant at the 1% level.

However, these coefficients by themselves are difficult to interpret due to the fact that the independent variable is an interaction between two continuous variables. Interpreting this variable as a policy multiplier, as described in Chapter 4, I find that for a hypothetical policy that increases Medicaid use by 1 percentage point, the network for these language groups will increase the probability of taking-up Medicaid by 9.9 percentage points for adults and 7.4 percentage points for children (averaged across all language groups). These multipliers are slightly larger (11.0 and 10.4 percentage points for adults and children, respectively) among the expanded sample.

The results from various models also indicate that these effects cannot be completely driven by alternative hypotheses driven by omitted variable biases. There are two main omitted variable biases to be concerned about. The first is differential geographic sorting, where people who live in areas of concentrations of their language group are different in some unobservable way from people who live in low concentration areas, but in a way that is correlated with Medicaid take-up. For example, suppose a recent immigrant that is part of a high take-up language group initially lives among in a high CA area. However, over time, this person's beliefs related to Medicaid change, and he/she moves away from this high CA area and behaves differently from the rest of the language group. In this model, this person would not have Medicaid, lives in a low CA area, and is part of a high take-up group. The network coefficient would be upward bias because it would assume that this person does not have Medicaid because he/she is in a low take-up area. Another example could be related how people could sort based on their health or disability status, where unobservable health characteristics of the individual create an upward bias on the network variable.

There are several results that provide evidence against these biases. First, this paper finds small differences across models when defining contact availability at the large super-PUMA or MSA levels. If geographic sorting were driving the results, we would expect to see more drastic differences across the models. Second, the results in tables 8A and 8B show that the language-geography group effect is stronger among recent immigrants compared to those that have been in the country for a longer period of time. I also find that the network variable remains positive and statistically significant even after controlling for years since entry in the U.S. and interaction terms between language group and years since entry, and when I limit the sample to those that have lived in the same house for the past year. I do not find evidence in favor of selection based on health status. I find that the network variable is statistically insignificant when limited to the portion of the sample that has at least one ACS-defined disability or limitation.

The second omitted variable bias concern is related to unobservable outreach efforts that are correlated with language and geography. This bias could partially explain some of the child sample results, but not the adult sample results. However, I find that network effects remain statistically significant after I exclude observations two states (CA and NY) that have well known outreach programs, and when I exclude states that do not have Medicaid outreach programs, targeted toward pregnant women, in foreign languages. I also address this concern by limiting the sample to Spanish speakers only and defining networks based on country of birth. If outreach works in manner that is correlated with language but not country of birth, the results from these models would only capture a potential country of birth network effect, but not an outreach effect. For adults, I find that the country of birth network variable is statistically significant at the 1% level and consistent in magnitude with the language network variable. However, I find statistically insignificant results among the Spanish-only child sample when defining networks based on the mother's country of birth. However, this result could just be attributable to the fact that a mother's country of birth is a weak measure of networks compared to language spoken at home.

The results from this study are also consistent with the hypothesis that networks operate through the spread of information. I find that language-geography defined group effects are strongest among the foreign born population and recent immigrants, whom are more likely to rely on social contacts to obtain information related to the U.S. health care system. I also find evidence that group effects are generally stronger in smaller geographic areas, where there could be more opportunities to run into potential contacts as opposed larger, sparsely population areas. Interestingly, I find that the effects are stronger among adults in linguistically isolated household, but that the opposite is true among children in linguistically isolated households.

Finally, I find that the results are relatively insensitive to various sample and model specification tests. I find that

- Language-geography group effects remain statistically significant even when excluding Spanish speakers and outlier language groups;

- Group effects are insensitive to the choice of including or excluding non-citizens, whether or not I include immigrants who might be ineligible for Medicaid/CHIP because they have been in the U.S. for under 5 years. Excluding non-citizens decreases the magnitude of the effect, but the results remain significant at the 1% level;
- The results are consistent across both core and expanded samples;
- The results are consistent across various definitions of Medicaid eligibility, language group take-up, contact availability;
- The results are consistent across the use of linear OLS and non-linear models, such as logits, probits, and multinomial logits.

Policy Implications

The main results from this paper have so far been interpreted through the broad lens of a policy multiplier effect: compared to a world without the existence of non-market interactions, on average, the presence of language-geography defined groups can increase the responsiveness to policies that aim to increase Medicaid take-up. However, one major limitation of this study is that the coefficient of interest can only be interpreted in terms of an average across all language groups; state policymakers must know the details of the populations that they are dealing with, in terms of the composition of language groups, where they generally live, and how the generally behave, in order to full realize the distributional effects of policy changes. The presence of group effects might make it easier to reach out to certain language groups with average to above-average take-up rates, but policymakers might face resistance among the low take-up language groups as a whole.

The results from this study could assist State and local policymakers that are looking for ways to spend the $100 million of approved CHIPRA outreach and enrollment funds. In order to maximize take-up among non-English speakers, the low hanging fruit for policymakers lies with the uninsured that are part of “moderate” to “high” take-up language groups. Policymakers might achieve “more bang-for-your-buck” with outreach efforts among this population. The results in Table 20 indicate that network effects are stronger among those in the top distribution of language take-up groups. As such, the message associated with outreach could be able to spread more quickly and efficiently among these groups.

Second, the eligible uninsured population that are part of low take-up language groups are more complicated, and there are different policy tools are needed to reach these populations depending on their characteristics. If state and local officials are convinced that these individuals are uninsured because they lack information related to Medicaid eligibility and enrollment procedures, improved outreach efforts can be use to provide practical information in terms of how and where to obtain coverage. However, low language group take-up rates might not be due to lack of practical information. It is possible that individuals that are part of these groups know about Medicaid eligibility rules, but decided that it's not worth the time and hassle costs to enroll. For these groups, simple outreach efforts may prove to be futile; policy-makers should focus on improving the value of Medicaid relative to being uninsured, either by decreasing the costs of enrollment (e.g., setting up more automatic processes) or increasing the benefits of coverage (e.g., improving the quality of care or increasing the network of providers that accept Medicaid patients).

Third, policymakers might want to target residents in low CA areas in order to maximize social welfare. People in high CA areas might have better informal insurance networks and information compared to those who are not surrounded by a high concentration of their language group. As such, outreach dollars could be used to enroll those uninsured who are in hard-to-reach places.

Finally, the results from this paper can be used to understand the newly eligible Medicaid population and analyze compliance patterns related to the individual mandate under the ACA. The ACA expands Medicaid eligibility to most adults under 138% of the federal poverty level. ACS data can be used to determine which language groups will be most affected by this expansion, and the results from this paper can be used to understand why newly eligible individuals from certain language groups are or are not enrolling in Medicaid. The results from this study could also shed some light on the behavior of non-English speakers between 138 and 400% of the FPL who could potentially participate in the health insurance Exchange. This paper supports the idea that individuals obtain information from others who are part of the same language group, regardless of their socio-economic status (contact availability is defined over the entire population). Moving forward, the patterns that we see with the Medicaid-eligible populations could be similar to the patterns that will emerge among those who are eligible for premium and cost-sharing subsidies in the Exchange. While overall take-up of Exchange benefits could be higher (most of these individuals must comply with the individual mandate or face a penalty and the quality of private coverage will most likely be higher) the distribution of of Exchange take-up could be similar to what we see among current Medicaid-eligible language groups.

There are some obvious limitations to this study that could hinder policymaker's ability to interpret or fully utilize these results. First, and most obviously, these results have little insight into the behavior of English speakers who are eligible for Medicaid benefits. Future research and data sources are needed to determine which networks (e.g., church participation) play an important role in influencing health insurance behavior for the majority of the population. Second, the coefficient on the main variable of interest is an average across all language groups, making it challenging to apply the same uniform information to all non-English language groups. It is also likely that the network mechanism (e.g., information related to product search vs. information related to product value) varies across different language groups. Third, the

ACS only captures point-in-time enrollment in Medicaid. Additional research can provide insight into how previous encounters with the Medicaid system influence the quality of information flowing through language-geography groups. Finally, this study only used age, income, and state or residence to determine Medicaid/CHIP eligibility; in reality, eligibility is more complicated. Future research will use a more precise eligibility simulation/imputation model to determine which populations might actually be eligible but were excluded from the sample (e.g., the disabled or medically needy) and which populations are not eligible but were included in the sample, such as undocumented immigrants.

Bibliography

Ai, Chunrong and Edward Norton, 2003. Interaction Terms in Logit and Probit Models. Economics Letters 80(1), 123−129.

Aizer, Anna, 2003. Low Take-Up in Medicaid: Does Outreach Matter and for Whom? AEA Papers and Proceedings 93(2), 238-241.

Aizer, Anna, 2007. Public Health Insurance, Program Take-Up, and Child Health. The Review of Economics and Statistics 89(3), 400-415.

Aizer, Anna and Janet Currie, 2004. Networks or Neighborhoods? Correlations in the Use of Publicly-Funded Maternity Care in California. Journal of Public Economics 88, 2573-2585.

Akerlof, George, 1980. A Theory of Social Custom, of Which Unemployment May be One Consequence. Quarterly Journal of Economics 94)4), 749-775.

Alba, Richard, 1990. Ethnic Identity: The Transformation of White America. Yale University Press.

Artiga, Samantha, 2009. Where are States Today: Medicaid and State-Funded Coverage Eligibility Levels for Low-Income Adults. Kaiser Commission on Medicaid and the Uninsured, Kaiser Family Foundation, #7993.

Bailey, N.T.J., 1975. The Mathematical Theory of Infectious Diseases and Its Applications Griffen, London.

Bakos, J. Yannis, 1997. Reducing Buyer Search Costs: Implications for Electronic Marketplaces. Management Science 43(12), 1676-1692.

Bala, Venkatesh, and Sanjeev Goyal, 1998. Learning from Neighbours. The Review of Economic Studies 65(3), 595-621.

Banerjee, Abhijit, 1992. A Simple Model of Herd Behavior. Quarterly Journal of Economics 107(3), 797-817.

Bansak, Cynthia, and Steven Raphael. 2006. The Effects of State Policy Design Features on Take-Up and Crowd-Out Rates for the State Children's Health Insurance Program. Journal of Policy Analysis and Management 26(1), 149-175.

Bertrand, Marianne, Erzo Luttm, and Sendhil Mullainathan, 2000. Network Effects and Welfare Cultures. Quarterly Journal of Economics 115, 1019-1055.

Bertrand, Marianne, Sendhil Mullainathan, and Eldar Shafir, 2004. Memos to the Council of Behavioral-Economics Advisors: A Behavioral-Economics View of Poverty. AEA Papers and Proceedings 94(2), 419-423.

Besley, Timothy, and Stephen Coate, 1992. Workfare versus Welfare: Incentive Arguments for Work Requirments in Poverty-Alleviation Programs. American Economic Review 82(1), 249-261.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, 1992. A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. Journal of Political Economy 100(5), 992-1026.

Borjas, George and Lynette Hilton, 1996. Immigration and the Welfare State: Immigrant Participation in Means-Tested Entitlement Programs. Quarterly Journal of Economics 111, 575-604.

Borjas, George, 1992. Ethnic Capital and Intergenerational Mobility. Quarterly Journal of Economics 107, 123-150.

Borjas, George, 1995. Ethnicity, Neighborhoods, and Human-Capital Externalities. American Economic Review 85, 365-390

Brock, William, and Steven Durlauf, 2001. Discrete Choice with Social Interactions. Review of Economic Studies 68(2), 235-260.

Brock, William, and Steven Durlauf, 2002. A Multinomial-Choice Model of Neighborhood Effects 92(2), 298-303.

Card, David and Lara Shore Sheppard, 2004. Using Discontinuous Eligibility Rules to Identify the Effects of the Federal Medicaid Expansions on Low Income Children. Review of Economics and Statistics 86(3), 752-766.

Case, Anne, and Christina Paxson, 2001. Mothers and Others: Who Invests in Children's Health. Journal of Health Economics 20, 301-328.

Christakis, Nicholas and James Fowler, 2007. The Spread of Obesity in a Large Social Network over 32 Years. The New England Journal of Medicine 357, 370-379.

Cohen-Cole, Ethan and Giulio Zanella, 2008. Unpacking Social Interactions. Economic Inquiry 46(1), 19-24.

Cohen-Cole, Ethan and Jason Fletcher, 2008 Is Obesity Contagious? Social Networks vs. Environmental Factors in the Obesity Epidemic. Journal of Health Economics 27, 1382-1387.

Congdon, William, Jeffrey Kling, and Sendhil Mullainathan, 2011. Policy Choice: Public Finance through the Lens of Behavioral Economics. Brookings Institution Press, c247pp.

Congressional Budget Office (CBO), 2010. H.R. 4872, Reconciliation Act of 2010 (Final Health Care Legislation). [Online] http://www.cbo.gov/ftpdocs/113xx/doc11379/AmendReconProp.pdf, April 23, 2011.

Currie, Janet, 2006. The Take-Up of Social Benefits. Ch. 3 in Poverty, The Distribution of Income , and Public Policy, eds. Alan Auerbach, David Card, and John Quigley, Russell Sage Foundation, 80-148.

Currie, Janet and Jonathan Gruber, 1996a. Health Insurance Eligibility, Utilization of Medical Care, and Child Health. Quarterly Journal of Economics 111, 431-466.

Currie, Janet and Jonathan Gruber, 1996b. Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women. Journal of Political Economy 04(6), 1263-1296.

Deri, Catherine, 2005. Social Networks and Health Service Utilization. Journal of Health Economics 24, 1076-1107.

Devillanova, Carlo, 2008. Social Networks, Information, and Health Care Utilization: Evidence from Undocumented Immigrants in Milan. Journal of Health Economics 27, 265-286.

Dubay, Lisa and Genevieve Kenney, 2001. Health Care Access and Use Among Low-Income Children: Who Fares Best? Health Affairs 20(1), 112-121.

Duflo, Esther and Emmanuel Saez, 2003. The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment. Quarterly Journal of Economics 118, 815-842.

Ellison, Glenn, and Drew Fudenberg, 1993. Rules of Thumb for Social Learning. Journal of Political Economy 101(4), 612-643.

Ellison, Glenn, and Drew Fudenberg, 1995. World-of-Mouth Communication and Social Learning. Quarterly Journal of Economics 110(1), 93-125.

Ettner Susan, 1997. Medicaid Participation Among the Eligible Elderly. Journal of Policy Analysis and Management 16, 237–255.

Feldman, Roger and Bryan Dowd, 1991. A New Estimate of the Welfare Loss of Excess Health Insurance. The American Economic Review 81(1), 297-301.

Fowler, James and Nicholas Christakis, 2008. Estimating Peer Effects on Health in Social Networks: A Response to Cohen-Cole and Fletcher; and Togden, Nonnemaker, and Pais. Journal of Health Economics 27, 1400-1405.

Gaviria, Alejandro and Steven Raphael, 2001. School-Based Peer Effects and Juvenile Behavior. The Review of Economics and Statistics 83, 257-268.

Glaeser, Edward, Bruce Sacerdote, and Jose Scheinkman, 1996. Crime and Social Interactions. Quarterly Journal of Economics 111(2), 507-548.

Granovetter, Mark, 1973. The Strength of Weak Ties. American Journal of Sociology 78(6), 1360-1380.

Granovetter, Mark, 2005. The Impact of Social Structure on Economic Outcomes. The Journal of Economic Perspectives 19(1), 33-50.

Greene, William., 2002. The Behavior of the Fixed Effects Estimator in Nonlinear Models. Working paper 02-05, Department of Economics, Stern School of Business, New York University, http://www.stern.nyu.edu/~wgreene/nonlinearfixedeffects.pdf, accessed May 25th, 2011.

Gresenz, Carol, Jeannette Rogowski, and Jose Escarce, 2007. Social Networks and Access to Health Care Among Mexican Americans. NBER Working Paper No 13460.

Gruber, Jonathan, 2003. “Medicaid”, in Robert Moffitt, ed., Means Tested Transfer Programs in the U.S. Chicago: University of Chicago Press, 2003, pp. 15-77

Hartmann, Wesley, Puneet Manchanda, Harikesh Nair, Matthew Bothner, Peter Dodds, David Godes, Kartik Hosanagar, and Catherine Tucker, 2008. Modeling Social Interactions: Identification, Empirical Methods, and Policy Implications. Marketing Letters 29(3-4), 284-304.

Hearne, Jean, 2005. Medicaid Eligibility for Adults and Children. Congressional Research Service Report for Congress, The Library of Congress, RL 33019.

Herring, Bradley, 2005. The Effect of the Availability of Charity Care to the Uninsured on the Demand for Private Health Insurance. Journal of Health Economics 24(2), 225-252.

Hill, Ian, Sara Hogan, Louise Palmer, Brigette Courtot, Shelly Gehshan, Dan Belnap, and Andrew Snyder, 2009. Medicaid Outreach and Enrollment for Pregnant Women: What Is the State of the Art? Urban Institute and National Academy for State Health Policy Report prepared for The March of Dimes Foundation.

Ioannides, Yannis, and Linda Loury, 2004. Job Information Networks, Neighborhood Effects, and Inequality. Journal of Economic Literature 42(4), 1056-1093.

Jackson, Mattew, 2005. The Economics of Social Networks. Proceedings of the 9th World Congress of the Econometric Society eds. Richard Blundell, Whitney Newey, and Torsten Persson, Cambridge University Press.

Jackson, Matthew and A. Wolinsky, 1996. A Strategic Model of Social and Economic Networks. Journal of Economic Theory, 71(1), 44-74.

Kaiser Commission on Medicaid and the Uninsured, 2006. Medicaid and SCHIP Eligibility for Immigrants. Kaiser Family Foundation #7492.

Katz, Lawrence, Jeffrey R. Kling, and Jeffrey B. Liebman, 2001. Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment. Quarterly Journal of Economics 116, 607-654.

Kenney, Genevieve, Victoria Lynch, Allison Cook, and Samantha Phong, 2010. Who And Where Are The Children Yet To Enroll In Medicaid And The Children's Health Insurance Program? Health Affairs 29(10), 1920-1929.

Klees, Barbara, Christian Wolfe, and Catherine Curtis, 2010. Brief Summaries of Medicare and Medicaid: Title XVIII and Title XIX of the Social Security Act as of November 1, 2010. Centers for Medicare & Medicaid Services, Office of the Actuary. [Online] https://www.cms.gov/MedicareProgramRatesStats/downloads/MedicareMedicaidSummaries2010.pdf, accessed May 25th, 2011.

Lazear, Edward, 1995. Culture and Language. Journal of Political Economy 107(S6), S95-S126.

Liebman, Jeffrey and Richard Zeckhauser, 2008. Simple Humans, Complex Insurance, Subtle Subsidies. NBER Working Paper 14330.

Lindbeck, Assar, Sten Nyberg, and Jorge Wibull, 1999. Social Norms and Economic Incentives in the Welfare State. Quarterly Journal of Economics 114(1), 1-35

LoSasso, Anthony, and Thomas Buchmueller. 2004. The Effect of the State Children's Health Insurance Program on Health Insurance Coverage. Journal of Health Economics 23, 1059–1082.

Lynch, Victoria, Michel Boudreaux, and Michael Davern, 2010. Applying and Evaluating Logical Coverage Edits to Health Insurance Coverage in the American Community Survey. U.S. Census Bureau Working Paper.

Ludwig, Jens, Greg J. Duncan, and Paul Hirschfield, 2001. Urban Poverty and Juvenile Crime: Evidence from a Randomized Housing-Mobility Experiment. Quarterly Journal of Economics 116, 655-679.

Manski, Charles, 1993. Identification of Endogenous Social Effects: The Reflection Problem. Review of Economic Studies 60(3), 531-542.

Manski, Charles, 2000. Economic Analysis of Social Interactions. Journal of Economic Perspectives 14(3), 115-136.

McFadden, Daniel, 1974. Conditional Logit Analysis of Qualitative Choice Behavior in P. Zarembka (editor), Frontiers in Econometrics, New York: Academic Press.

McFadden, Daniel, 1981. Econometric Models of Probabilistic Choice in C. Manski and D. McFadden (editors), Structural Analysis of Discrete Data with Econometric Applications, Cambridge, Mass.: M.I.T. Press.

Moffit, Robert, 1983. An Economic Model of Welfare Stigma. American Economic Review 73(5), 1023-1035.

Moffitt, Robert, 2001. Policy Interventions, Low-Level Equilibria and Social Interactions. Ch. 3 in Social Dynamics, eds. Steven Durlauf and Peyton Young, MIT Press, 45-81.

Pagan, Jose, and Mark Pauly, 2006. Community-Level Uninsurance and the Unmet Medical Needs of Insured and Uninsured Adults. Health Services Research 41(3p1), 788-803.

Pauly, Mark, and Mark Satterthwaite, 1981. The Pricing of Primary Care Physicians Services: A Test of the Role of Consumer Information. The Bell Journal of Economics 12(2), 488-506

Pauly, Mark, and Jose Pagan, 2007. Spillovers and Vulnerability: The Case of Community Uninsurance. Health Affairs 26(5), 1304-1314.

Pauly, Mark, Fredric Blavin, and Sudha Meghan, 2009. How Private, Voluntary Health Insurance Can Work in Developing Countries. Health Affairs 28(6), 1778-1787.

Remler, Dahlia, and Sherry Glied, 2003. What Other Programs Can Teach Us: Increasing Participation in Health Insurance Programs. American Journal of Public Health 93(1), 67-74.

Rosenbaum, James, 1995. Changing the Geography of Opportunity by Expanding Residential Choice: Lessons from the Gautreaux Program. Housing Policy Debate 6(1), 231-261.

Ross, Donna Cohen, Marian Jarlenski, Samantha Artiga, and Caryn Marks, 2009. A Foundation for Health Reform: Findings of a 50 State Survey of Eligibility Rules, Enrollment and Renewal Procedures, and Cost- Sharing Practices in Medicaid and CHIP for Children and Parents During 2009. Kaiser Commission on Medicaid and the Uninsured.

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek, 2010. Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota

Sebelius, Kathleen, 2010. Rising to the Challenge: Tools for Enrolling Eligible Children in Health Coverage. Health Affairs 29(10), 1930-1932.

Sacerdote, Bruce, 2001. Peer Effects with Random Assignment: Results for Dartmouth Roommates. Quarterly Journal of Economics 116, 681-704.

Selden, Thomas, Jessica Banthin, and Joel Cohen, 1998. Medicaid's Problem Children: Eligible But Not Enrolled. Health Affairs 17(3), 192-200.

Shen, Yu-Chu, and Stephen Zuckerman, 2005. The Effect of Medicaid Payment Generosity on Access and Use Among Medicaid Beneficiaries. Health Services Research 40(3), 723-744.

Sorensen, Alan, 2006. Social Learning and Health Plan Choice. RAND Journal of Economics 37(4), 929-945.

Stuber Jennifer, Kathleen Maloy, Sara Rosenbaum, and Karen Jones, 2000. Beyond Stigma: What Barriers Actually Affect the Decisions of Low-Income Families to Enroll in Medicaid? Center for Health Services Research and Policy Issue Brief, George Washington University.

Subramanian, SV, and Ichiro Kawachi, 2004. Income Inequality and Health: What Have We Learned So Far? Epidemiologic Reviews 26, 78-91.

Trogden, Justin, James Nonnemaker, and Joanne Pais. 2008. Peer Effects in Adolescent Overweight. Journal of Health Economics 27, 1388-1399.

Turner, Joanna, Michel Boudreaux, and Victoria Lynch, 2009. A Preliminary Evaluation of Health Insurance Coverage in the 2008 American Community Survey. U.S. Census Bureau, http://www.census.gov/hhes/www/hlthins/2008ACS_healthins.pdf.

U.S. Census Bureau, October (1992, 1995, and 1999) and November (1979 and 1989), Current Population Surveys, and 2000–2008 American Community Survey. [Online] Available http://www.childstats.gov/americaschildren/tables/fam5.asp, April 23, 2011.

U.S. Census Bureau, 2010. Response Rates-Data. [Online] Available http://www.census.gov/acs/www/methodology/response_rates_data/, April 23, 2011.

U.S. General Accounting Office (GAO), 1994. Health Care Reform: Potential Difficulties in Determining Eligibility for Low Income People. Washington D.C.: Government Printing Office, document number GAO/HEHS-94.

Winship, Christopher, and Larry Radbill, 1994. Sampling Weights and Regression Analysis. Sociological Methods and Research 23(2), 230-257.

Wolfe, Barbara, and Scott Scrivner. 2005. The Devil May Be in the Details: How the Characteristics of SCHIP Programs Affect Take-Up. Journal of Policy Analysis and Management 24(3), 499-522.

Yelowitz, Aaron. 2000. Using the Medicare Buy-In Program to Estimate the Effect of Medicaid on SSI Participation. Economic Inquiry 38(3), 419–441.

Zuckerman, Stephen, Aimee Williams, and Karen Stockley, 2009. Trends in Medicaid Physician Fees, 2003-2008. Health Affairs 28(3), w510-w519.