The Healthy Undocumented Immigrant Effect: Evidence from the US *

Hoa Vu†

January 12, 2020

Abstract

This paper uses residual approach to identify undocumented populations hidden in micro survey data. I then document what I term the "Healthy Undocumented Immigrant Effect": undocumented immigrants are healthier than legal immigrants. I show evidence that the paradox of the undocumented immigrants' health advantage can be attributed to the return-migrant effect.

Keywords: Health disparity, Healthy immigrant effect, Undocumented Immigrant, Return-migrant.

JEL Classification: I10, J10, J15

*I am grateful to David Slusky, Donna Ginther, Jennifer Hunt, Brent Zey, and seminar participants at IV Workshop on Migration, Health and Well-Being; the University of Kansas, for their help and comments. I am thankful to George Borjas for making available his undocumented status imputation algorithm. All errors are my own.

†Department of Economics, University of Kansas, Lawrence, KS 66045. Email: hoavu@ku.edu

1 Introduction

Relatively little is known about unauthorized immigrants' health outcomes. It is well documented that immigrants are on average healthier relative to comparable native-born populations (Moullan and Jusot, 2014; Neuman, 2014; Vang et al., 2017). This is known as the "Healthy Immigrant Effect" (Markides and Coreil, 1986; Palloni and Arias, 2004; Kennedy et al., 2015). However, past research often does not distinguish between legal and unauthorized immigrants. The reason is that surveys usually do not explicitly inquire about undocumented status or we did not have a valid method to detect unauthorized immigrants in micro datasets.[1]

This study adopts a newly developed method[2] constructed by George Borjas that imputes undocumented status in micro survey data to study the disparity between health outcomes of undocumented immigrants and legal immigrants which focuses on the question: Are undocumented immigrants relatively healthier than legal immigrants? Specifically, in my analysis, I identify individuals in the National Health Interview Survey (NHIS) that are likely undocumented immigrants. I then document the health difference across immigration groups what I term the "Healthy Undocumented Immigrant Effect," the notion that unauthorized immigrants are healthier than legal immigrants. Given this paradoxical finding, I evaluate possible expla-

[1]There are surveys, including (among others) the National Agricultural Workers Survey and the Survey of Income and Program Participation, that ask about documentation status. However, one might wonder whether immigrants would answer such questions or answer them honestly (Carter-Pokras and Zambrana, 2006).

[2]Details about this method are in the Appendix. Borjas (2016) developed his undocumented status imputation algorithm by adjust the assignment method used by Passel and Cohn (2014). Researchers have increasingly employed this method to assign legal status to immigrants in micro-data (Pourat et al., 2014; Borjas, 2017; Borjas and Slusky, 2018; Cohen and Schpero, 2018; Gunadi, 2019).

nations for this paradox. I find evidence suggesting that the return-migrant effect might account for the health disparity between undocumented immigrants and legal immigrants.

There is a fair amount of recent research on the health of undocumented immigrants (see a review in Martinez et al. 2015). My work is closest to works of Arbona et al. (2010) and Chavez (2012). Arbona et al. (2010) studied undocumented and documented Hispanic immigrants living in two major cities in Texas and found that undocumented immigrants reported higher levels of challenges from family separation and language difficulties than documented immigrants. Chavez (2012) found that undocumented immigrants were less likely to (1) have insurance, (2) experience stress, and (3) utilize medical services when compared to legal immigrants and citizens.

The NHIS undocumented immigrant sample used in this article has several desirable features. First, the sample is much larger than most previously used undocumented immigrant data sets, especially in the US. For example, Arbona et al. (2010) uses a sample of 416 Mexican and Central American immigrants, Chavez (2012) works with a sample of 1201 observations, and Poon et al. (2013) uses information on 1620 observations. The relatively large size of the NHIS sample, with 29,836 likely undocumented immigrants, is critical to draw robust estimations. Second, it is also noteworthy that the NHIS undocumented immigrants are derived from a national probability sample and therefore representative of the general undocumented immigrant population, whereas most existing US undocumented immigrants data sets are to varying degrees derived from convenience samples.

The rest of the paper proceeds as follows. Section 2 provides an overview of the data. Section 3 documents the healthy undocumented immigrant effect. In Section 4, I explore possible explanations for the health disparity between undocumented and legal immigrants. Section 5 offers some concluding remarks.

2 Data

I use data from the 1998-2017 National Health Interview Survey (NHIS) Integrated Public Use Microdata Series (Blewett et al., 2018). The NHIS is a cross-sectional household interview survey that collects annual data on the health status and medical conditions of a large, nationally representative sample of the US population. It samples approximately 35,000 households, containing about 87,500 persons per year. NHIS is suitable for this analysis because it contains information that helps detect undocumented status as well as health status. Also, the NHIS sample is large enough to allow a statistically reliable estimate of the undocumented population. I restrict the sample to foreign-born people aged 18-64 as few individuals aged 65 and older are imputed undocumented status. Therefore, I do not have the statistical power to draw robust conclusions for the elderly sample.

The algorithm that I use to impute undocumented status is the residual method developed by Warren and Passel (1987) and Passel and Cohn (2014). Borjas (2017) adapted the method to the Current Population Surveys. Roughly speaking, the algorithm classifies the foreign-born persons in the sample who are likely to be legal, the residual of foreign-born persons then are classified as likely undocumented immigrants.[3]

I classify the population into natives, legal immigrants (including naturalized citizens and legal residents), and undocumented immigrants. I next compare health outcomes between two immigrant groups: undocumented versus legal. Due to the fact that naturalized citizens have both immigrant status and the full protection of citizenship, I do robustness checks by comparing undocumented immigrant with naturalized citizen (Appendix Table A.2) and legal residents (Appendix Table A.3), separately.

In the NHIS data, my primary health outcomes are: 1) self-reported health status measured on the Likert scale score of 1 = Excellent, 2 = Very Good, 3 = Good, 4 = Fair, 5 = Poor and 2) psychological distress measured on the Kessler 6 (K6) scale. In my analysis, I define poor health as the bottom two measure on the Likert scale (Likert score is 4 or 5). In 2016, for example, for people aged 18-64, the definition implies ten percent of the population was reported to be in poor health. Self-reported health is a strong prediction of serious chronic conditions and mortality, even when controlling for objective measures of health status and health behaviors (Goldstein et al., 1984; Idler and Kasl, 1995; Burström and Fredlund, 2001; Case et al., 2002; Wu et al., 2013).

I construct a measure of psychological distress based on K6 score that ranges from 0 to 24, with a higher score indicating higher level of mental health distress.[4]

[3] Details about the algorithm are in the Appendix.

[4] In particular, I used the following six questions in the NHIS to measure mental health: (1) During the past 30 days, how often did you feel so sad that nothing could cheer you up? (2) During the past 30 days, how often did you feel nervous? (3) During the past 30 days, how often did you feel restless and fidgety? (4) During the past 30 days, how often did you feel hopeless? (5) During the past 30 days, how often did you feel that everything was an effort? (6) During the past 30 days,

I then recode the mental health score into a dummy variable: mental distress (1 indicating respondents K6 score is 5 or more, otherwise 0).

It is possible that self-reported measures of health status may be affected by measurement error or biases (Baker et al., 2004; Case et al., 2002; Currie and Stabile, 2003). Thus, together with self-reported health and mental distress, I present results for an indicator whether a person has any activity limitation, in which diagnosis and misreporting are unlikely.[5,6]

Table 1 shows the summary statistics for the main variables used in my analysis. The first row of Table 1 reports the fraction of the US population imputed as likely undocumented immigrants in the NHIS. This fraction is similar to the estimated fraction of undocumented immigrants by Pew Research Center and Borjas (2017) using the Current Population Survey. Undocumented populations tend to be younger, less educated, and less likely to be employed than legal immigrants. About

how often did you feel worthless? Respondents were asked to provide answers to these questions on a scale of 0 to 4 (none of the time, a little of the time, some of the time, most of the time, and all of the time).

[5] I use "lany" variable in the NHIS. "lany" is a recoded variable from several questions that indicates whether a person has any activity limitation: (1) Because of a physical, mental, or emotional problem, does [person] need the help of other persons with personal care needs, such as eating, bathing, dressing, or getting around inside this home? (2) Because of a physical, mental, or emotional problem, does [person] need the help of other persons in handling routine needs, such as everyday household chores, doing necessary business, shopping, or getting around for other purposes? (3) Does a physical, mental, or emotional problem now keep [person] from working at a job or business? (4) Is [person] limited in the kind or amount of work [he/she] can do because of a physical, mental or emotional problem? (5) Because of a health problem, does [person] have difficulty walking without using any special equipment? (6) Is [person] limited in any way because of difficulty remembering or because [he/she] experiences periods of confusion? (7) Is [person] limited in any way in any activities because of physical, mental or emotional problems?

[6] In the Appendix Table A.1, I present the results using an additional set of 9 potentially serious health conditions on which the NHIS collects information. These include whether the person has ever been diagnosed with chronic conditions (asthma, bronchitis, cancer, diabetes, heart disease, hypertension, liver condition, and ulcer), and information on whether the person is obese (BMI ≥ 30).

health outcomes, while 10 percent of legal immigrants reported to be in poor health, only 7 percent of undocumented immigrants reported to be in poor health. Similarly, 17 percent of legal immigrants reported suffering from mental distress while 14 percent of unauthorized immigrants reported having mental distress in the past 30 days. Undocumented also less likely to report having any activity limit, 2.5 percent compared with 7 percent that of legal immigrants.

3 Healthy Undocumented Immigrant Effect

To examine the differences in health outcomes between legal immigrants and undocumented immigrants, I estimated the following logit regression model:

$$log\frac{p_{irt}}{1-p_{irt}} = \alpha + \beta \cdot Undocumented_{irt} + year_t + region_r + \gamma X_{irt} + \epsilon_{irt} \tag{1}$$

where p_{irt} is a dummy variable denoting whether person i in region r at time t reports health condition C. $C = \{$poor health, mental distress, any activity limit$\}$. $Undocumented_{irt}$ is a dummy variable denoting whether person i is an undocumented immigrant. X_{irt} consists of person i's demographic characteristics such as age, age squared, sex, race, dummies for education, married dummy, dummy for health insurance coverage, number of persons in family, and number of years spent in the US.

Year fixed effects ($year_t$) are included to my preferred specification to capture the change in quality and origins of immigrants over time. Undocumented immigrants could settle in certain areas. I account for this possibility by including census region of residence fixed effects ($region_r$).[7] All specifications are estimated using sample weights,[8] and standard errors are clustered at the region level.

Table 2 reports estimates of equation (1) for all immigrants, Hispanic immigrants, and non-Hispanic immigrants using the NHIS sample of 1998-2017.[9] Column (1) of Table 2 shows the results for all immigrants, column (2) shows the results for Hispanic immigrants, column (3) shows the results for Non-Hispanic immigrants. As observed in Table 2, the marginal effects of Undocumented coefficients do not vary much when I stratify by Hispanic ethnicity.

The results in Panel A and B of Table 2 show that a lower proportion of undocumented immigrants reports poor health or mental distress. In particular, undocumented immigrants are 2.3 percent and 2.4 percent less likely reporting poor health and mental distress, respectively, than legal immigrants. It is surprising that undocumented migrants report lower levels of distress as access to needed services such as health, legal, educational and other social support services is non-existent or challenging for undocumented migrants in the US, and mental health services are even less accessible. Furthermore, anti-immigration policies in the US (Secure Communities, 287(g) program, E-Verify) have a negative impact on undocumented immigrants' mental health outcomes (Martinez et al., 2015; Wang and Kaushal, 2018). However, this is consistent with findings in Chavez (2012) that undocumented immigrants are less likely to experience stress in their lives than legal immigrants and citizen using

[7]Region is the smallest geographic unit identified in the IPUMS NHIS. The four regions are: Northeast, North Central/Midwest, South, and West.

[8]The results are similar between weighting and not weighting (Solon et al., 2015)

[9]Inspired by the fact that the majority of undocumented are Hispanic (Krogstad et al., 2019), I stratify the sample to Hispanic and non-Hispanic immigrants.

a random sample telephone survey of 805 Latinos and 396 non-Hispanic whites.

The results in Panel C indicate that among all immigrants, undocumented are 3.6 percent less likely having any activity limitation. Among Hispanic immigrants, undocumented Hispanics are 3.4 percent less likely reporting activity limitation than legal Hispanic immigrants. A similar disparity exists between undocumented non-Hispanic immigrants and legal non-Hispanic immigrants.

These results are robust with different group categorizations. In Table 2, I classify groups as legal immigrants (including both naturalized citizens and legal residents) and undocumented immigrants. The observed health advantage favoring the undocumented persists if I compare undocumented immigrants with naturalized citizens and legal residents separately. Results are presented in the Appendix Table A.2 and Table A.3. These results are also robust when comparing undocumented immigrants with a propensity score matched legal immigrants sample (Table A.4).[10]

4 Understanding the Disparity

The results above have shown that undocumented immigrants in the United States experience better health outcomes than legal immigrants. This phenomenon is paradoxical because undocumented immigrants generally have lower access to social benefits than legal immigrants. I evaluate some potential explanations for the paradox

[10]Specifically, I construct a legal immigrant sample with observables similar to the undocumented sample, I estimate individual propensity scores with a logit specification that models the probability of undocumented status as a function of the age, age squared, sex, race, education, marital status, health insurance coverage status, number of persons in family, and number of years spent in the US. This propensity score is used to match undocumented immigrants to their nearest neighbor in the legal immigrant sample (with replacement), and only these matched legal immigrant observations are used in this robustness analysis.

in this section.

Selection effect. This hypothesis says that the paradox of healthy undocumented immigrant effect can be explained by selection effect, whereby undocumented immigrants who enter the US are disproportionately drawn from groups at country of origin whose health status is better than those who migrate legally. Due to the fact that most of the undocumented immigrants who come to the US do not expect to receive any social benefits, only the ones with better physical and psychological health from the population migrate. Thus, undocumented immigrants are healthier than those who do not migrate and maybe healthier than the average legal immigrants in the receiving country.

Health outcomes for undocumented immigrants and legal immigrants may become increasingly similar because both groups might integrate themselves into a new country by adopting the native population's social, cultural, and behavioral factors (Waters and Jimnez, 2005; Antecol and Bedard, 2006). The consequence is that the health advantage of undocumented immigrants should be diluted as number of years spent in the US increases, when age effects are held constant. Thus, if the selection effect prevails, I should observe a decreasing advantage as the number of years spent in the US increases.

I use data from the NHIS to examine whether the selection effect accounts for the better health outcomes of undocumented immigrants than legal immigrants. The NHIS collects information on the number of years a person spent in the US. However, the NHIS only collects this data in intervals: less than 1 year, 1-5 years, 5-10 years, 10-15 years, and 15 years or more. In Table 3, I show results of the effects of a dummy variable reflecting undocumented status relative to legal immigrants with different duration of stay.

The results in Table 3 do not support the selection effect. In fact, results in column (1) indicate insignificant effects at the shortest duration of stay. Results in columns (2) to (4) indicate larger significant health advantages for undocumented immigrants as number of years spent in the US increases. This suggests that the patterns found in the data do not support the selection effect hypothesis.[11]

Return-migrant effect. This explanation assumes that undocumented immigrants return to their home country following the period of illness or unemployment. The reason for higher return rate of undocumented immigrants than legal immigrants is the earlier group have no access to health care services when they are ill or social benefits when they are unemployed. The return of the sick undocumented immigrants will result in average better health outcomes for the undocumented population.[12]

This line of reasoning might suggest the health disparity is larger at older ages if the return-migrant effect prevails. I use data from the NHIS to test this conjecture. Table 4 shows estimates of logit model in Equation 1, stratifying by age. Each parameter in Table 4 presents the health disparity between undocumented and legal immigrants in that age group, holding the number of years spent in the US constant.

The results in Table 4 provide evidence that there are significant return-migrant

[11]This is consistent with the findings of Palloni and Arias (2004) that selection effects or assimilation does not explain the Hispanic adult mortality advantage (Hispanics in the US experience lower mortality rates in adulthood than do non-Hispanic whites).

[12]Ullmann et al. (2011) found that Mexican returned migrants from the US are healthier before they migrate, and less healthy when they return compared with non-migrants in Mexico.

effects. Moving from Column (1) to Column (4), I find a disparity in health that increases with age. In addition, the larger health disparity with the longer duration of stay found in Table 3 could reflect the attrition of unhealthier undocumented immigrants as the duration of stay increases and is consistent with return-migrant effect. These findings mean that the return-migrant effect may prove important in explaining the healthy undocumented immigrant effect.

5 Discussion and Conclusion

I first discuss two limitations of my article. First, I acknowledge that the residual method mistakenly impute undocumented status for college graduated immigrants, despite accuracy for low-skilled immigrants (Albert, 2019). Specifically, Borjas and Cassidy (2019) suspects the algorithms mistakenly classify the high-skilled immigrants who are in the US temporarily under H-1B visa as undocumented.[13] However, I am not worry about this since it will result in downwardly bias the health disparity towards zero.[14] Also, the estimation results are robust if high-skilled immigrants are excluded from the analysis (Table A.5).

Second, I reach the conclusion that the health advantage of undocumented immigrants is related to the return migration of those who are in poor health using indirect evidence. The direct test for the return-migrant hypothesis is to compare health outcomes of recent return undocumented migrants to the health outcomes of

[13]The stratification by education results (Table A.6) supports Borjas and Cassidy (2019)'s suspection. In particular, the results are mixed for high-skilled groups (some college and college graduate).

[14]Given that legal immigrants report worse health outcomes than undocumented immigrants found in this article.

undocumented migrants who remained in the US. Such a comparison is difficult, since there is no follow-up of undocumented migrants who return to their country. Future research with data on the return unauthorized immigrants could better explain the paradoxical finding here.

To conclude, this paper makes two contributions. First, I identify undocumented immigrants in the NHIS and document what I term the "Healthy Undocumented Immigrant Effect": undocumented immigrants are healthier than legal immigrants. Second, I test two hypotheses that may explain the health advantage of undocumented immigrants. I find suggestive evidence that the return-migrant effect may prove important in explaining the healthy undocumented immigrant effect.

References

Albert, C. (2019). The labor market impact of immigration: Job creation vs. job competition. Working paper, CEMFI.

Antecol, H. and Bedard, K. (2006). Unhealthy assimilation: Why do immigrants converge to american health status levels? *Demography*, 43(2):337–360.

Arbona, C., Olvera, N., Rodriguez, N., Hagan, J., Linares, A., and Wiesner, M. (2010). Acculturative stress among documented and undocumented latino immigrants in the united states. *Hispanic Journal of Behavioral Sciences*, 32(3):362–384.

Bachmeier, J. D., Van Hook, J., and Bean, F. D. (2014). Can we measure immigrants' legal status? lessons from two u.s. surveys. *International Migration Review*, 48(2):538–566.

Baker, M., Stabile, M., and Deri, C. (2004). What do self-reported, objective, measures of health measure? *The Journal of Human Resources*, 39(4):1067–1093.

Blewett, L. A., Rivera Drew, J. A., Griffin, R., King, M. L., and Williams, K. C. (2018). Ipums health surveys: National health interview survey, version 6.3 [dataset].

Borjas, G. J. (2016). The earnings of undocumented immigrants. Working Paper 23236, National Bureau of Economic Research.

Borjas, G. J. (2017). The labor supply of undocumented immigrants. *Labour Economics*, 46:1–13.

Borjas, G. J. and Cassidy, H. (2019). The wage penalty to undocumented immigration. *Labour Economics*, 61:101757.

Borjas, G. J. and Slusky, D. J. (2018). Health, employment, and disability: Implications from the undocumented population. Working Paper 24504, National Bureau of Economic Research.

Burström, B. and Fredlund, P. (2001). Self rated health: Is it as good a predictor of subsequent mortality among adults in lower as well as in higher social classes? *Journal of Epidemiology & Community Health*, 55(11):836–840.

Carter-Pokras, O. and Zambrana, R. E. (2006). Collection of legal status information: Caution! *American Journal of Public Health*, 96(3):399–399.

Case, A., Lubotsky, D., and Paxson, C. (2002). Economic status and health in childhood: The origins of the gradient. *American Economic Review*, 92(5):1308–1334.

Chavez, L. R. (2012). Undocumented immigrants and their use of medical services in orange county, california. *Social Science & Medicine*, 74(6):887 – 893. Part Special Issue: Migration, 'illegality', and health: Mapping embodied vulnerability and debating health-related deservingness.

Cohen, M. S. and Schpero, W. L. (2018). Household immigration status had differential impact on medicaid enrollment in expansion and nonexpansion states. *Health Affairs*, 37(3):394–402.

Currie, J. and Stabile, M. (2003). Socioeconomic status and child health: Why is the relationship stronger for older children? *American Economic Review*, 93(5):1813–1823.

Flood, S., King, M., Rodgers, R., Ruggles, S., and Warren, J. R. (2018). Integrated public use microdata series, current population survey: Version 6.0 [dataset].

Goldstein, M. S., Siegel, J. M., and Boyer, R. (1984). Predicting changes in perceived health status. *American Journal of Public Health*, 74(6):611–614.

Gunadi, C. (2019). On the association between undocumented immigration and crime in the United States. *Oxford Economic Papers*.

Idler, E. L. and Kasl, S. V. (1995). Self-Ratings of Health: Do they also Predict change in Functional Ability? *The Journals of Gerontology: Series B*, 50B(6):S344–S353.

Kennedy, S., Kidd, M. P., McDonald, J. T., and Biddle, N. (2015). The healthy immigrant effect: Patterns and evidence from four countries. *Journal of International Migration and Integration*, 16(2):317–332.

Krogstad, J. M., Passel, J. S., and Cohn, D. (2019). 5 facts about illegal immigration in the u.s. Technical report, D.C. Pew Hispanic Center.

Markides, K. S. and Coreil, J. (1986). The health of hispanics in the southwestern united states: An epidemiologic paradox. *Public Health Reports*, 101(3):253–265.

Martinez, O., Wu, E., Sandfort, T., Dodge, B., Carballo-Dieguez, A., Pinto, R., Rhodes, S., Moya, E., and Chavez-Baray, S. (2015). Evaluating the impact of immigration policies on health status among undocumented immigrants: A systematic review. *Journal of Immigrant and Minority Health*, 17(3):947–970.

Moullan, Y. and Jusot, F. (2014). Why is the healthy immigrant effect different between european countries? *European Journal of Public Health*, 24:80–86.

Neuman, S. (2014). Are immigrants healthier than native residents? *IZA World of Labor*, page 108.

Palloni, A. and Arias, E. (2004). Paradox lost: Explaining the hispanic adult mortality advantage. *Demography*, 41(3):385–415.

Passel, J. S. and Cohn, D. (2011). Unauthorized immigrant population: National and state trends, 2010. Technical report, D.C. Pew Hispanic Center.

Passel, J. S. and Cohn, D. (2014). Unauthorized immigrant totals rise in 7 states, fall in 14: Decline in those from mexico fuels most state decreases. Technical report, D.C. Pew Research Center's Hispanic Trends Project.

Poon, K. K., Dang, B. N., Davila, J. A., Hartman, C., and Giordano, T. P. (2013). Treatment outcomes in undocumented hispanic immigrants with hiv infection. *PLOS ONE*, 8(3):1–7.

Pourat, N., Wallace, S. P., Hadler, M. W., and Ponce, N. (2014). Assessing health care services used by californias undocumented immigrant population in 2010. *Health Affairs*, 33(5):840–847.

Solon, G., Haider, S. J., and Wooldridge, J. M. (2015). What are we weighting for? *Journal of Human Resources*, 50(2):301–316.

Ullmann, S. H., Goldman, N., and Massey, D. S. (2011). Healthier before they migrate, less healthy when they return? the health of returned migrants in mexico. *Social Science & Medicine*, 73(3):421 – 428.

Van Hook, J., Bachmeier, J. D., Coffman, D. L., and Harel, O. (2015). Can we spin straw into gold? an evaluation of immigrant legal status imputation approaches. *Demography*, 52(1):329–354.

Vang, Z. M., Sigouin, J., Flenon, A., and Gagnon, A. (2017). Are immigrants healthier than native-born canadians? a systematic review of the healthy immigrant effect in canada. *Ethnicity & Health*, 22(3):209–241.

Wang, J. S.-H. and Kaushal, N. (2018). Health and mental health effects of local immigration enforcement. Working Paper 24487, National Bureau of Economic Research.

Warren, R. and Passel, J. S. (1987). A count of the uncountable: Estimates of undocumented aliens counted in the 1980 united states census. *Demography*, 24(3):375–393.

Waters, M. C. and Jimnez, T. R. (2005). Assessing immigrant assimilation: New empirical and theoretical challenges. *Annual Review of Sociology*, 31(1):105–125.

Wu, S., Wang, R., Zhao, Y., Ma, X., Wu, M., Yan, X., and He, J. (2013). The relationship between self-rated health and objective health status: A population-based study. *BMC public health*, 13:320.

6 Tables

Table 1: Summary Statistics

	Natives	Legal	Undocumented
Percent of population	83.4	10.7	5.9
Male	48.7	47.8	55.6
Age	41.4	42.2	36.1
Self-reported overall health (Likert scale 1-5)	2.2	2.2	2.1
Poor health	**10.6**	**10.3**	**7.4**
Mental health (K6 score 0-24; higher values indicating higher levels of mental health distress)	2.6	2.1	1.8
Mental distress (K6 score ≥ 5)	**19.7**	**16.8**	**14.4**
Activity limit	**12.87**	**6.97**	**2.54**
Asthma	13.05	6.98	3.85
Cancer	5.49	2.61	0.82
Bronchitis	4.32	1.91	0.82
Diabetes	9.25	10.51	6.09
Heart condition	6.06	3.18	1.24
Hypertension	23.19	18.92	10.84
Liver condition	1.44	1.41	0.91
Obesity	31.32	22.91	22.59
Ulcer	6.85	4.60	2.98
High school drop out	7.55	19.40	39.13
High school graduate	29.67	24.05	23.42
Some college	33.47	24.17	13.76
College graduate	29.32	32.38	23.69
Percent employed	80.97	78.07	75.54
Percent married	56.37	71.02	58.50
Sample size	335,923	47,446	29,836

Notes: Weighted. The native's statistics are for reference only (not included in the regressions). Data are from 1998-2017 IPUMS NHIS. The sample includes persons aged 18-64. The values are in percentages.

Table 2: Logit Model, Health Outcomes by Immigration Status

	All immigrants (1)	Hispanic immigrants (2)	Non-hispanic immigrants (3)
Panel A: Poor health			
Undocumented	-0.023***	-0.023***	-0.023***
	(0.004)	(0.005)	(0.004)
Observations	79,750	44,927	34,823
Panel B: Mental distress			
Undocumented	-0.025***	-0.024***	-0.024***
	(0.006)	(0.007)	(0.005)
Observations	79,750	44,927	34,823
Panel C: Has any activity limitation			
Undocumented	-0.036***	-0.034***	-0.037***
	(0.003)	(0.006)	(0.009)
Observations	79,655	44,845	34,810
Demographics	X	X	X
Year and region fixed effects	X	X	X

Notes: Marginal effects are reported, with standard errors clustered at the region level in brackets. Each parameter is from a separate logit model regression of the outcome variable on a dummy variable equal to one if the person has the health condition listed in panels A - C. The reference category is legal immigrants. Data are from the 1998-2017 IPUMS NHIS. The sample includes legal and undocumented immigrants aged 18-64. All estimations include year fixed effects, region fixed effects, and demographic controls: age, age squared, sex, race, dummies for education, married dummy, dummy for health insurance coverage, number of persons in family, and number of years spent in the US. All results are estimated using sample weights. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table 3: Logit Model, Health Outcomes by Number of Years Spent in the US

	0-5 year (1)	5-10 year (2)	10-15 year (3)	> 15 year (4)
Panel A: Poor health				
Undocumented	-0.007	-0.018***	-0.015***	-0.031***
	(0.011)	(0.001)	(0.005)	(0.006)
Observations	11,326	12,414	12,939	42,995
Panel B: Mental distress				
Undocumented	0.000	-0.019***	-0.026***	-0.037***
	(0.005)	(0.006)	(0.004)	(0.010)
Observations	11,326	12,414	12,939	42,995
Panel C: Has any activity limitation				
Undocumented	-0.017***	-0.016***	-0.020***	-0.052***
	(0.006)	(0.006)	(0.006)	(0.005)
Observations	11,279	12,204	12,920	42,956
Demographics	X	X	X	X
Age fixed effects	X	X	X	X
Year and region fixed effects	X	X	X	X

Notes: Each parameter is from a separate logit model regression of the outcome variable on a dummy variable equal to one if the person has the health condition listed in panels A - C. The reference category is legal immigrants (including both naturalized citizens and legal residents). Marginal effects are reported, with standard errors clustered at the region level in brackets. Data are from the 1998-2017 IPUMS NHIS. The sample includes legal and undocumented immigrants aged 18-64. All estimations include age fixed effects, year fixed effects, region fixed effects, and demographic controls: age squared, sex, race, dummies for education, married dummy, dummy for health insurance coverage, number of persons in family, and number of years spent in the US. All results are estimated using sample weights. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table 4: Logit Model, Health Outcomes by Age Group

	18-24 (1)	25-36 (2)	37-49 (3)	50-64 (4)
Panel A: Poor health				
Undocumented	-0.007	-0.013*	-0.021***	-0.051***
	(0.006)	(0.007)	(0.003)	(0.010)
Observations	8,307	26,957	26,581	17,895
Panel B: Mental distress				
Undocumented	-0.004	-0.013*	-0.034***	-0.049***
	(0.012)	(0.007)	(0.005)	(0.008)
Observations	8,317	26,957	26,581	17,895
Panel C: Has any activity limitation				
Undocumented	-0.005	-0.015***	-0.037***	-0.082***
	(0.004)	(0.001)	(0.007)	(0.007)
Observations	8,287	26,928	26,531	17,878
Demographics	X	X	X	X
Years spent in the US fixed effects	X	X	X	X
Year and region fixed effects	X	X	X	X

Notes: Each parameter is from a separate logit model regression of the outcome variable on a dummy variable equal to one if the person has the health condition listed in panels A - C. The reference category is legal immigrants (including both naturalized citizens and legal residents). Marginal effects are reported, with standard errors clustered at the region level in brackets. Data are from the 1998-2017 IPUMS NHIS. The sample includes legal and undocumented immigrants aged 18-64. All estimations include years spent in the US fixed effects, year fixed effects, region fixed effects, and demographic controls: age, age squared, sex, race, dummies for education, married dummy, dummy for health insurance coverage, number of persons in family. All results are estimated using sample weights. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

A Appendix

A.1 Identifying Undocumented Immigration Status[15]

Micro survey data do not include documentation status. As a result, I used Borjas (2016)'s algorithm to impute immigration status in the 1998-2017 NHIS. This approach is similar to residual methodologies used by Pew Research Center and the Department of Homeland Security to estimate the size of the undocumented immigrant population.

The algorithm is as follows: a foreign-born survey respondent is first identified as a documented immigrant if any of the following criteria apply:

a. that person arrived before 1980;

b. that person is a citizen;

c. that person receives Social Security benefits, SSI, Medicaid, Medicare, or Military Insurance;

d. that person is a veteran, is currently in the Armed Forces;

e. that person works in the government sector;

f. that person resides in public housing or receives rental subsidies, or that person is a spouse of someone who resides in public housing or receives rental subsidies;

g. that person was born in Cuba (as practically all Cuban immigrants were granted refugee status);

h. that person's occupation requires some form of licensing (such as physicians, registered nurses, air traffic controllers, and lawyers);

i. that person's spouse is a legal immigrant or citizen.

Any remaining foreign-born individuals are then categorized as likely to have undocumented immigration status.

[15] Borjas G. The Earnings of Undocumented Immigrants. NBER Working Paper 23236, page 9

Table A.1: More Health Outcomes

	(1) Asthma	(2) Bronchitis	(3) Cancer
Undocumented Immigrants	-0.022***	-0.008***	-0.007***
	(0.004)	(0.002)	(0.001)
Observations	79,722	79,717	79,713
	(4) Diabetes	**(5) Heart condition**	**(6) Hypertension**
Undocumented Immigrants	-0.005***	-0.010***	-0.019***
	(0.002)	(0.002)	(0.007)
Observations	79,707	79,705	79,643
	(7) Liver condition	**(8) Obesity**	**(9) Ulcer**
Undocumented Immigrants	-0.003**	-0.010**	-0.010***
	(0.001)	(0.005)	(0.002)
Observations	79,699	79,750	79,672

Notes: Each parameter is from a separate logit model regression of the outcome variable on a dummy variable equal to one if the person has the health condition (ever diagnosed) listed in columns (1)-(9). The reference category is legal immigrants (including both naturalized citizens and legal residents). Marginal effects are reported, with standard errors clustered at the region level in brackets. Data are from the 1998-2017 IPUMS NHIS. The sample includes legal and undocumented immigrants aged 18-64. All estimations include year fixed effects, region fixed effects, and demographic controls: age, age squared, sex, race, dummies for education, married dummy, dummy for health insurance coverage, number of persons in family, and number of years spent in the US. All results are estimated using sample weights. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table A.2: Logit Model, Reference Group Is Naturalized Citizen

	All immigrants (1)	Hispanic immigrants (2)	Non-hispanic immigrants (3)
Panel A: Poor health			
Undocumented	-0.018***	-0.014**	-0.022***
	(0.004)	(0.006)	(0.004)
Observations	63,828	34,369	29,459
Panel B: Mental distress			
Undocumented	-0.017***	-0.010*	-0.022***
	(0.005)	(0.005)	(0.006)
Observations	63,828	34,369	29,459
Panel C: Has any activity limitation			
Undocumented	-0.030***	-0.027***	-0.029***
	(0.004)	(0.008)	(0.008)
Observations	63,744	34,297	29,447
Demographics	X	X	X
Year and region fixed effects	X	X	X

Notes: Each parameter is from a separate logit model regression of the outcome variable on a dummy variable equal to one if the person has the health condition listed in panels A - C. The reference category is naturalized citizen. Marginal effects are reported, with standard errors clustered at the region level in brackets. Data are from the 1998-2017 IPUMS NHIS. The sample includes legal and undocumented immigrants aged 18-64. All estimations include year fixed effects, region fixed effects, and demographic controls: age, age squared, sex, race, dummies for education, married dummy, dummy for health insurance coverage, number of persons in family, and number of years spent in the US. All results are estimated using sample weights. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.3: Logit Model, Reference Group Is Legal Resident

	All immigrants (1)	Hispanic immigrants (2)	Non-hispanic immigrants (3)
Panel A: Poor health			
Undocumented	-0.028***	-0.031***	-0.021***
	(0.007)	(0.008)	(0.006)
Observations	46,570	30,897	15,673
Panel B: Mental distress			
Undocumented	-0.032***	-0.038***	-0.027***
	(0.007)	(0.009)	(0.007)
Observations	46,570	30,897	15,673
Panel C: Has any activity limitation			
Undocumented	-0.033***	-0.032***	-0.034***
	(0.002)	(0.003)	(0.008)
Observations	46,502	30,834	15,668
Demographics	X	X	X
Year and region fixed effects	X	X	X

Notes: Each parameter is from a separate logit model regression of the outcome variable on a dummy variable equal to one if the person has the health condition listed in panels A - C. The reference category is naturalized citizen. Marginal effects are reported, with standard errors clustered at the region level in brackets. Data are from the 1998-2017 IPUMS NHIS. The sample includes legal and undocumented immigrants aged 18-64. All estimations include year fixed effects, region fixed effects, and demographic controls: age, age squared, sex, race, dummies for education, married dummy, dummy for health insurance coverage, number of persons in family, and number of years spent in the US. All results are estimated using sample weights. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.4: Logit Model, Reference Group Is Propensity Matched Legal Immigrants

	All immigrants (1)	Hispanic immigrants (2)	Non-hispanic immigrants (3)
Panel A: Poor health			
Undocumented	-0.023***	-0.023***	-0.023***
	(0.004)	(0.005)	(0.004)
Observations	79,750	44,927	34,823
Panel B: Mental distress			
Undocumented	-0.025***	-0.024***	-0.024***
	(0.006)	(0.007)	(0.005)
Observations	79,750	44,927	34,823
Panel C: Has any activity limitation			
Undocumented	-0.036***	-0.034***	-0.037***
	(0.003)	(0.006)	(0.009)
Observations	79,655	44,845	34,810
Demographics	X	X	X
Year and region fixed effects	X	X	X

Notes: Each parameter is from a separate logit model regression of the outcome variable on a dummy variable equal to one if the person has the health condition listed in panels A - C. The reference category is naturalized citizen. Marginal effects are reported, with standard errors clustered at the region level in brackets. Data are from the 1998-2017 IPUMS NHIS. The sample includes legal and undocumented immigrants aged 18-64. All estimations include year fixed effects, region fixed effects, and demographic controls: age, age squared, sex, race, dummies for education, married dummy, dummy for health insurance coverage, number of persons in family, and number of years spent in the US. All results are estimated using sample weights. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.5: Logit Model, Robustness to exclude high-skilled immigrants

	All immigrants (1)	Hispanic immigrants (2)	Non-hispanic immigrants (3)
Panel A: Poor health			
Undocumented	-0.028***	-0.023***	-0.036***
	(0.005)	(0.005)	(0.007)
Observations	58,838	40,565	18,273
Panel B: Mental distress			
Undocumented	-0.031***	-0.025***	-0.042***
	(0.007)	(0.009)	(0.008)
Observations	58,838	40,565	18,273
Panel C: Has any activity limitation			
Undocumented	-0.045***	-0.037***	-0.056***
	(0.004)	(0.006)	(0.012)
Observations	58,756	40,489	18,267
Demographics	X	X	X
Year and region fixed effects	X	X	X

Notes: Each parameter is from a separate logit model regression of the outcome variable on a dummy variable equal to one if the person has the health condition listed in panels A - C. High-skilled immigrants are defined as who have a college degree or more. Marginal effects are reported, with standard errors clustered at the region level in brackets. Data are from the 1998-2017 IPUMS NHIS. The sample includes legal and undocumented immigrants aged 18-64. All estimations include year fixed effects, region fixed effects, and demographic controls: age, age squared, sex, race, dummies for education, married dummy, dummy for health insurance coverage, number of persons in family, and number of years spent in the US. All results are estimated using sample weights. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.6: Logit Model, Health Outcomes by Education Group

	HS dropout (1)	HS graduate (2)	Some college (3)	College graduate (4)
Panel A: Poor health				
Undocumented	-0.036***	-0.029***	-0.013	-0.010***
	(0.009)	(0.003)	(0.008)	(0.003)
Observations	24,095	18,794	15,949	20,912
Panel B: Mental distress				
Undocumented	-0.038***	-0.038***	-0.010	-0.009
	(0.013)	(0.001)	(0.008)	(0.010)
Observations	24,095	18,794	15,949	20,912
Panel C: Has any activity limitation				
Undocumented	-0.053***	-0.041***	-0.033***	-0.010*
	(0.005)	(0.006)	(0.010)	(0.005)
Observations	24,041	18,776	15,939	20,899
Demographics	X	X	X	X
Year and region fixed effects	X	X	X	X

Notes: Each parameter is from a separate logit model regression of the outcome variable on a dummy variable equal to one if the person has the health condition listed in panels A - C. The reference category is legal immigrants (including both naturalized citizens and legal residents). Marginal effects are reported, with standard errors clustered at the region level in brackets. Data are from the 1998-2017 IPUMS NHIS. The sample includes legal and undocumented immigrants aged 18-64. All estimations include year fixed effects, region fixed effects, and demographic controls: age, age squared, sex, race, dummies for education, married dummy, dummy for health insurance coverage, number of persons in family, and number of years spent in the US. All results are estimated using sample weights. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.