LUND UNIVERSITY
School of Economics and Management

Master's Programme in Economic Growth, Population and Development

Disability risk disparity among the immigrant population in the United States, 2009-2018

by

Wooseong Kim

wooseong.kim.5248@student.lu.se

Abstract: This thesis attempts to examine the relationship between immigrant generation status and disability risk in the US. The difference in disability risk, resulting from a series of chronic health conditions, is one of the major indicators of health disparities among immigrants which contribute to persistent ethnic/racial stratification of health. Existing literature has primarily focused on the first generation immigrant's health advantage in terms of disability observed in cross-sectional surveys. However, researchers have paid scant attention to the disability risk of the descendants of the first generation immigrants to date. This thesis uses the data extracted from the Current Population Survey (CPS) to investigate the relationship between immigrant generation status and disability risk. Furthermore, by employing the longitudinal design of the BMS, this thesis examines the differences in short-term disability risk among the respondents. The results show that the first generation immigrants show a lower risk of having a disability. Meanwhile, the descendants of the first generation immigrants generally show a higher risk of disability. Moreover, regardless of generation status, socioeconomic status is negatively associated with short-term disability risk. The main findings indicate that existing theories concerning the immigrant's health inequalities, e.g. the Healthy Immigrant Effect and the Fundamental Causes Theory may predict disability risk of immigrants in the US context as well. Although the implication is limited due to the short observation period, the associations found in this research should be studied further with an advanced longitudinal approach.
Keywords: Disability, Health inequality, Immigration, Immigrant health disparity

EKHS02
Master's Thesis (15 credits ECTS)
June 2020
Supervisor: Kirk Scott
Examiner: Martin Dribe
Word Count: 16,989

Acknowledgements

I would like to express my greatest gratitude to my supervisor, Kirk Scott, for helpful and supporting comments throughout the writing process. He always welcomed me whenever I encountered challenging issues and needed his advice. Although the unprecedented outbreak of COVID-19 has fundamentally affected every aspect of daily life, he and I managed to have several online thesis supervision meetings, and he continued to help me to the end. I would also like to thank Anna Tegunimataka for helping me further my initial idea on this topic and find the relevant data source. Also, I am appreciative of supporting words by wonderful friends in Sweden and Korea. Last but not least, I would like to express my deepest love and thankfulness to my mother, Tae-sook Choi, and father, Jong-oh Kim. Your selfless love will always be with me.

Table of Contents

List of Tables

List of Figures

1 Introduction

The health trajectory of the immigrants has been a significant area in migration studies. In the US, the health differences between the immigrants and the natives have been a primary concern for two reasons. Firstly, the health disparities by nativity account for the persistent racial/ethnic health inequalities in the US society. After the passage of the 1965 Immigration Act, a massive flow of immigrants entered the US economy (Gans, 1992; Park & Myers, 2010). Compared with the previous immigration influx from Europe during the late 19th century, the new immigration flow has consisted of individuals with various ethnic and national backgrounds. Therefore, the health differences between natives and the new diverse immigrants have become a new source of racial/ethnic health stratification in US society.

Secondly, studying immigrant health trajectories contributes to the understanding of migration and acculturation in the US. Most post-1965 immigrants came from developing countries with inferior standards of living and healthcare systems compared to those of the US. Notwithstanding, they have been known to show a peculiar pattern regarding their health trajectories, which is relatively superior health status upon arrival combined with diminishing health advantages as immigrants spend more time in the hosting country. This phenomenon which is documented as the "Healthy Immigrant Effect (HIE)", and the link between the HIE and immigrant's selection and assimilation have received considerable scholarly attention for decades (Cho & Hummer, 2001; Kennedy, Kidd, McDonald & Biddle, 2015).

Indeed, a large volume of research has attempted to investigate the overall pattern of the US immigrant health differentials relative to the natives and explain the persistent HIE. Several researchers have found evidence supporting the HIE amongst Hispanic (Abraido-Lanza, Dohrenwend, Ng-Mak & Turner, 1999; Markides, Eschbach, Ray & Peek, 2007), Asian (Cho & Hummer, 2001; Frisbie, Cho & Hummer, 2001; Mutchler, Prakash & Burr, 2007), and African American first generations (Elo, Mehta & Huang, 2011). The high level of self-selection, along with selection by the hosting country and underreporting of previous health issues, account for immigrant's better health status at the time of arrival (Akresh & Frank, 2008). Meanwhile, acculturation to the hosting country, i.e. convergence in norms and lifestyle has been argued to be the main factor which deteriorates the health advantage of first generations. Besides, previous research suggests that the acculturation process, both within and between generation may differ by the ethnic/racial backgrounds and the country of origins (Cho & Hummer, 2001; Elo, Mehta & Huang, 2011).

1.1 Research Problem

Although previous studies on immigrant health in the US recognise the overall health trajectories of the first generation immigrants, they share two limitations: Firstly, research has yet to systematically investigate the health outcomes of *the second generation* immigrant, who is defined as a native with two foreign-born parents. Studying the second generation sheds new light on migration studies because its generation status reflects “the length of exposure to the host culture” (Allen, Elliott, Morales, Diamant, Hambarsoomian & Schuster, 2007, p.337) and therefore can be a reasonable proxy for acculturation. Despite their relevance in immigrant health research, the second and further generation immigrants have received scant attention from researchers for a couple of reasons. Firstly, previous literature primarily focuses on the elderly population and age-specific mortality rates, which excludes most existing descendants of the post-1965 immigrants who have not reached the age threshold. Moreover, few exceptions investigating the acculturation relevant to health among the second generations primarily focus on health-risk behaviours, e.g. smoking, dietary habits or physical exercise (Afable-Munsuz et al., 2010; Allen et al., 2007), which are only indirect indicators of later health outcomes. Secondly, most previous studies fail to employ data which has relevant information to separate second generations from a native with two native-born parents. In a similar vein, most studies do not distinguish the second generation and the *2.5 generation immigrant*, who is defined as a native-born individual with one native-born parent and one foreign-born parent, although previous literature implies that the two groups should be dealt as a separate group (Ramakrishnan, 2004).

Another common drawback of previous research is that it often employs a cross-sectional design. Although this approach explores associations between nativity/immigrant generation status and health outcomes, it fails to provide a reliable causal interpretation since it is vulnerable to reverse causality and omitted variable bias issues. Given that migration studies are prone to unobservable factors due to selection bias and imperfect information on the pre-immigration period, this issue may considerably limit our understanding of the problem.

1.2 Aim, Purpose, and Scope

The primary aim of this paper is to contribute to the understanding of how immigrants in US society integrate to the destination society in terms of their health, mainly focusing on the comparison among the natives, the first, second and 2.5 generation immigrants. This paper attempts to fill the current research gap by employing the Current Population Survey (CPS) dataset. The recent contribution by the Integrated Public Use Microdata Series (IPUMS) team has made it available to the CPS as a short-term longitudinal dataset which follows an individual up to 16 months (Rivera Drew, Flood & Warren, 2014). The CPS survey also contains information to separate the natives and individuals with foreign background, i.e. the first, second and the 2.5 generation immigrants. This paper uses self-reported disability status to measure health outcome. Disability has been used as one of the indicators for health especially among the elderlies in several previous studies for it is commonly received as a result of chronic health conditions that considerably limit individual's daily activities (Mutchler, Prakash & Burr, 2007). Moreover, this thesis aims to analyse data obtained from a recent period, which has not covered in the current literature. Thus, it investigates ten years of survey responses from 2009 to 2018.

To summarise, this study makes significant contributions to research on immigrant socioeconomic integration by demonstrating how health status and acculturation process amongst the first generation immigrants continue to affect the health of their descendant. Also, this paper makes use of the recent progress in the CPS dataset to utilise the longitudinal design of it. Since most previous research on immigrant health in the US could not employ ample information on socioeconomic status, demographic characteristics, and immigrant status in a longitudinal setting, analysing a panel dataset extracted from CPS to examine the short-term disability trajectories provide new insights into the old question.

1.3 Research Question

To achieve the research goal, this paper tries to answer the two research questions as below:

RQ 1 How is the disability risk difference between immigrants and natives in the US? Furthermore, how is it different by immigrant generation status and race/ethnicity?

RQ 2 Which socioeconomic factors affect an individual's disability risk in the US?

The first question is to link existing knowledge on immigrant's health and acculturation to a relatively understudied topic of disability risk. The second question also relates the main topic to a well-documented association between health and socioeconomic status, also known as the Fundamental Causes Theory (Link & Phelan, 1995)

1.4 Outline of the Thesis

The rest of the paper is organised as follows: The second section of this paper critically reviews the theoretical framework and previous. It will delve into the main explanations of HIE, the role of acculturation on immigrant health, the association between socioeconomic status and health, and how these theories explain disability risk differences. Section 3 describes the data extraction procedure and how each variable is defined. Section 4 explains how to construct a short-term longitudinal data using the CPS and analytic strategies employed by this paper for the statistical analysis. The following section reports the main findings of the empirical analysis, followed by Section 6, which is devoted to present a further discussion of the findings. Section 7 concludes this paper with a summary of the research and some suggestions for future research.

2 Theory and Previous Research

2.1 Immigrant Health Trajectories

2.1.1 The Healthy Immigrant Effect

A considerable amount of literature has documented overall better health outcomes among first generation immigrants, using various indicators of health, e.g. age-specific mortality, self-reported health, or functional/mental difficulties (Huh, Prause & Dooley, 2008; Hummer, Rodgers, Nam & LeClere, 1999). The HIE hypothesis demonstrates that recent immigrants are healthier than their native-born counterparts in the US, but this advantage diminishes as immigrants spend more time in the destination country (Gee, Kobayashi & Prus, 2004). After the HIE received notable academic attention, follow-up studies have attempted to test HIE among immigrants in other developed countries, and report similar findings across several Western countries (Johansson, Helgesson, Lundberg, Nordquist, Leijon, Lindberg & Vingård, 2012; Kennedy et al., 2015).

As mentioned above, the HIE can be divided into two parts: the initial health outcome differences and later convergence to their native counterparts. Three potential explanations account for the initial health advantage: (a) The selectivity of immigrant health, i.e. healthier immigrants self-select into the migration process, (b) The screening, i.e. the receiving countries also require a particular level of health status and individuals with severe health conditions are likely to be disqualified, and (c) The faulty data argument which insists that the overall health advantage of newly arrived immigrants is a mere combination of under-reporting of the existing health issues (Gee, Kobayashi & Prus, 2004; Jass & Massey, 2004).

The explanations for a levelled-off health advantage of the first generation immigrants are following the logic of the occurrence of the initial health gap. Provided that the health gap exists, the following argument is that immigrants are prone to be exposed to the factors negatively affecting individual’s health, e.g. less access to the healthcare, adverse socioeconomic status or stressful experience relevant to migration and discriminatory practices (Gee, Ro, Shariff-Marco & Chae 2009; Uretsky & Mathiesen, 2007). It can also be maintained that the convergence into the host country’s lifestyle results in changes in health-relevant behaviours and increased risk (Franzini & Fernandez-Esquer, 2004). On the other hand, if the effect is mere statistical artefacts, the convergence does not necessarily mean actual physical health deterioration but probably results from delayed diagnosis of the conditions immigrants already have (Riosmena, Kuhn & Jochem, 2017).

2.1.2 Intergenerational Transmission of Health among Immigrants

Examining the health disparities between the descendants of immigrants and their native counterparts enhances the understanding of immigrant integration process. Although generation status indirectly measures acculturation to the hosting country, it is a crucial and feasible measure associated with key determinants of health, e.g. healthcare access and utilisation, education, social capital and health-relevant behaviours (Allen et al., 2007). Previous literature expects that the health of the second generation will fall relative to the first generation for a couple of reasons. Firstly, it can be a necessary result of the HIE and regression towards the mean. According to Jass and Massey (2004), although a large volume of previous literature relates the decline in the health of the second generation immigrants to the factor relevant to acculturation process, the mechanism of *the regression towards mediocrity* (Galton, 1886) may account for the most of the tendency. Although this argument gives an account of the overall decline of the health outcomes of the descendants of the first generation immigrants, it lacks explanation that to what extent the regression towards the mean explains the decline. Moreover, it fails to explain the persistent differences within second generations belonging to different ethnic/racial groups or from different countries of origin.

Secondly, it is argued that understanding acculturation is essential for explaining the intergenerational transmission of health among the immigrant population. Berry (2003) defines acculturation as the internal, psychological, or sociocultural adaptations that materialise when an individual attempts to bridge two different cultural groups. He emphasises the multifaceted aspects of acculturation which implies that acculturation could be either individual/group level, could be postponed or even result in a reactive adaptation of traditional mode of life (Berry, 2003). Indeed, the literature focusing on the role of acculturation on the intergenerational health trajectories points out that different sub-groups of immigrants show heterogeneous patterns of the transmission of health-risk behaviours or acculturation stress associated with mental health conditions (Allen et al., 2007; Liddell, Nickerson, Sartor, Ivancic & Bryant 2016).

In a similar vein, different acculturation pathways related to the health-risk behaviour are predicted by the *Segmented Assimilation Hypothesis* (Portes & Zhou, 1993). In their original work, Portes and Zhou (1993) demonstrate that the assimilation process of the children of the post-1965 immigrants in the US is shaped by the obstacles elevating the risk of downwards assimilation and their utilisation of resources to confront these vulnerabilities. Later, the subsequent research identifies three pathways of acculturation resulting in different assimilation outcomes: consonant (i.e. the immigrant parents and children jointly learn the host country’s cultural values and language), selective (i.e., the accommodation of the host country’s cultural values coincides with the preservation of vital elements of the origin culture), and dissonant (i.e. adaptation of cultural values of the host society is in company with the rejection of those brought by and associated with their parents) acculturation (Haller, Portes & Fernández-Kelly, 2009). The authors suggest that the second generations experiencing selective acculturation are associated with positive outcomes because having a good command of both languages and appreciation of both culture functions as advantages in education and employment. Concerning health-risk behaviour transmission, the selective acculturation may be associated with better health outcomes of the second generations if cultural or ethnic-specific healthy lifestyles account for the first generation's health advantage. Meanwhile, given that the first generation's acculturation to the host country's lifestyle is related to the deterioration of health in their later lives, their children undergoing dissonant acculturation or consonant acculturation would report worsened health status relevant to their parents.

The final point relevant to acculturation is the potential difference between the second and 2.5 generation. The current segmented assimilation theory may not demonstrate how the 2.5 generation's experience differs from those with two native-born parents. They may expose to two distinctly different cultural backgrounds per se, implying that acculturation occurs at the household level. They may have an appearance similar to minority groups as well. Although an exceptional level of paucity remains in the existing literature regarding the 2.5 generation's acculturation experience, the evidence from existing literature indicates that the 2.5 generations show better educational attainment and higher average personal income compared with their second generation counterparts and even those with two native-born parents (Levels, Dronkers & Kraaykamp, 2008; Mateus, 2019; Ramakrishnan, 2004, p.5). Therefore, it is probable to expect that their health status is better than those groups, but the advantage is mainly due to their superior socioeconomic factors.

2.2 Social Determinants of Health and Disability

2.2.1 The Social Determinants of Health and the Fundamental Causes Theory

Evidence from several empirical studies has established that the socioeconomic condition of individuals plays a pivotal role in determining their health conditions. Early empirical studies on this topic find that the social inequality mortality between social groups persists and even widen, notwithstanding the overall decline in all-cause mortality (Mackenbach, Kunst, Cavelaars, Groenhof & Geurts, 1997; Marmot & Mcdowall, 1986). A significant association between income and life expectancy is also found in the US context as well (Chetty, Abraham, Lin, Scuderi, Bergeron, Cutler & Turner, 2016). Consequently, extensive empirical evidence concerning the association between socioeconomic conditions and health outcome motivated the official recognition of the social determinants of health by WHO (Marmot, 2005), as a critical source of health inequalities, defined as "the unfair and avoidable differences in health status seen within and between countries" (WHO, 2020).

As the association between socioeconomic status and health outcomes has gained scholarly attention, researchers have attempted to explain why the health disparities persist and whether the differences will converge or remain. Among them, the Fundamental Causes Theory (FCT) by Link and Phelan (1995) provides the most consistent explanation for the presence and persistence of social inequalities in health. In their influential paper, Link & Phelan (1995) demonstrate that "the fundamental social cause of disease involves resources that determine the extent to which people are able to avoid risks for morbidity and mortality" (Link & Phelan, 1995, p.88). This explanation indicates that the unequal distribution of critical resources such as knowledge, money, power, and social networks is linked to multiple disease outcome differences (Phelan, Link & Tehranifar, 2010). Also, this theory implies that belonging to a certain group associated with advantageous/adverse socioeconomic conditions is also related to health risk disparities.

Furthermore, they argue that the effects of fundamental social causes hold regardless of the changes in the profile of risk factors. According to this argument, the association between social causes and health outcomes is reproduced over time (Link & Phelan, 1995). In Phelan, Link and Tehranifar (2010), the authors elaborate on the original theory of fundamental causes by presenting recent empirical findings supporting their significant points. They re-emphasise their explanations that crucial resources can be employed regardless of what risk and protective factors are dominant in a given circumstance and the demonstrated association is reproduced over time through replacement of intervening mechanisms (Phelan, Link & Tehranifar, 2010).

Although the FCT provides new insight into the association between the socioeconomic condition and health outcomes, some recent findings concerning the historical pattern of social inequalities in health contradict one of the main arguments that the people with higher socioeconomic status have always had advantages in preventing or avoiding health risks. As a matter of fact, studies on the development of social inequalities in health demonstrate that the onset of widening social differences in health is a relatively modern phenomenon, mainly observed after the industrialisation ((Bengtsson & Dribe, 2011; Molitoris & Dribe, 2016). Due to the newly found historical evidence, the theory of fundamental cause is criticised for it does not fully explain how the social inequalities in health started to develop during the modernisation period (Bengtsson & van Poppel, 2011). Nevertheless, there exist little controversy that the theory adequately addressees the existing association between socioeconomic status and health. Therefore, given the scope of this paper, this thesis assumes that the main implications of this theory remain valid.

2.2.2 Disability and the Fundamental Causes Theory

The term disability commonly refers to a difficulty experienced in doing activities in any domain of life due to the health of physical problem (Verbrugge & Jette, 1994). According to the authors, disability often occurs as a consequence of non-fatal chronic disease and an individual can accumulate multiple chronic health conditions which may lead to disability (Minkler, Fuller-Thomson & Guralnik, 2006; Verbrugge & Jette, 1994). The literature on the recent conceptualisation of disability, known as the biopsychological model of disability, emphasises the multifaceted characteristics of disability and the understanding of it within the social context. This model synthesises the medical and social approaches to disablement (Bickenbach, Chatterji, Badley & Üstün, 1999). The revision of the initial International Classification of Impairments, Disabilities and Handicaps (ICIDH), named as ICIDH-2 embodied the new model and succeeded by the current International Classification of Functioning, Disability and Health (ICF) (WHO, 2020). To date, the assessment and classification of disability in most countries are based on ICF.

Understanding disability as a consequence of several chronic and non-fatal health or physical conditions has a crucial implication in applying the FCT to predicting disability risk. Throughout a lifetime, lack of key resources would lead to a higher risk of chronic health conditions, and it results in a higher probability of having a disability as well. Since several risk factors determine the overall risk of having various types of disability, it is challenging to identify each pathway to a specific disability over a lifetime. However, according to the theory, clarifying a specific mechanism is irrelevant to predicting the social gradient in disability because individuals with better socioeconomic status can employ their vital resources to protect themselves from various potential risk factors (Phelan, Link & Tehranifar, 2010). In summary, disparities in disability risk are predicted to be observed across socioeconomic strata. Also, the observed disability risk by nativity or immigrant generation status may partly be explained by mere socioeconomic differences, indicating that these factors should be included in the empirical model to avoid omitted variable bias.

2.3 Disability Pattern among US Immigrants

2.3.1 Existing Literature on Disability Pattern among US Immigrants

A number of empirical studies have documented the disability risk differences amongst different immigrant groups in the US relevant to their nativity, socioeconomic and demographic characteristics. Similar to other health outcomes such as adult/perinatal mortality and self-reported health, the evidence supporting the HIE was found by several researchers. For instance, Cho and Hummer (2001) show that the HIE is supported among the most Asian and Pacific Islander American (API) groups, i.e. the foreign-born individuals are on average have a lower disability risk compared with their native-born counterparts. However, this health advantage diminishes as the first generation spend more time in US society. Fuller-Thomson, Brennenstuhl and Hurd (2011) also report that API groups show lower prevalence rates in terms of four different categories of disability, i.e. functional limitations, limitations in activities of daily living, cognitive problems, and visual or hearing difficulties. One significant difference between the two studies is that the latter shows that the first generation API groups are compared with non-Hispanic White natives and still show a lower disability risk. In the same vein, previous studies on African Americans (Elo, Mehta & Huang, 2011) and Hispanic (Huh, Prause & Dooley, 2008; Mutchler, Prakash & Burr, 2007) also state the same pattern of a lower disability risk of the first generation immigrants compared with their native counterparts and the negative association between their health advantage and the years since migration. Although Huh, Prause and Dooley (2008) do not directly compare the disability risk by nativity but compare self-reported health and chronic disease morbidity risk, the main finding is relevant because chronic health conditions are closely linked to disability in later life (Minkler, Fuller-Thomson & Guralnik, 2006). To summarise, concerning the HIE, the evidence found in the US reaches consensus.

Also, most of the previous studies point out the within-group heterogeneity in terms of disability risks. For API Americans, the disparity in disability risk is explicitly shown between the Japanese Americans and the Southeast Asian Americans, e.g. Vietnamese and

Filipino Americans (Cho & Hummer, 2001; Fuller-Thomson, Brennenstuhl & Hurd, 2011). African American immigrants from Africa, the Caribbean and Europe/Canada also show differences in risk of physical activity limitations, although their overall advantages over the native-born African Americans remain (Elo, Mehta & Huang, 2011). Furthermore, the significant within-group differences by the country of origin among Hispanic first generations (Markides et al., 2007) corroborate the existing evidence.

While most studies agree on emphasising that ignoring the role of country of origin within the same racial/ethnic immigrant group may attenuate the HIE, they suggest a couple of distinct explanations for the discrepancies. On the one hand, some researchers point out the different levels of self-selection among immigrants ((Cho & Hummer, 2001; Markides et al., 2007). Indeed, the level of self-selection may be different by type of migration, such as economic migrants, refugees, or tied-movers. According to Jass and Massey (2004), the labour migrants are likely to be most positively selected in terms of their health conditions as their expected benefit of migration and their ability is both positively correlated with health. On the other hand, other scholars demonstrate that socioeconomic status in the destination countries may account for most of the story since the observed disability risk disparities significantly weaken after controlling for individual's SES, e.g. educational attainment, marital status or income (Cho & Hummer, 2001; Elo, Mehta & Huang, 2011; Mutchler, Prakash & Burr, 2007). This line of argument is following the main implication of the theory of fundamental causes (Link & Phelan, 1995). However, these two explanations are not mutually exclusive for a migrant with primary health advantages due to higher level of selectivity may achieve better outcome since migration as well. Jass and Massey (2004) also recognise that an individual's characteristics appreciated in the labour market often comes with excellent health conditions. Therefore, the main implication from the discussion concerning the within-group heterogeneity among post-1965 immigrants in the US is that it is considered necessary to take the country/region of origin and socioeconomic characteristics of immigrants into consideration in the empirical analysis.

2.3.2 The Paucity of Previous Research on the US Immigrant Disability Risk

Although previous studies have recognised the relationship between nativity and socioeconomic status and the disability risk among the first generation immigrants in the US, research has yet to systematically investigate the second generation's disability risk differences compared with the first or third-generation counterparts. The second and the 2.5 generation's disability is predicted to be more determined by the acculturation process and socioeconomic background they have been exposed to, rather than health selection and life events relevant to migration itself. In other words, the mechanisms which account for the HIE become less relevant, and the identified key resources emphasised by the theory of fundamental causes play vital roles. To date, relatively little research has been carried out on the second generation immigrant's disability pattern in the US and even less on the 2.5 generations. Although studies comparing the foreign-born and the native-born counterparts belonging to same racial/ethnic groups shed some light on this topic (Cho & Hummer, 2001; Elo, Mehta & Huang, 2011), it fails to recognise individual's generation status adequately.

For instance, A native-born African American could be a second-generation immigrant born between two foreign-born parents or a descendant of African American family who has been living in the US for centuries. Although the migration history of API and the Hispanic population may be shorter relevant to African American group, the same difficulty in identifying the generation status of *native-born* minority group remains.

Another research gap, also linked to the first one, is that researchers mostly examine the disability risk differences amongst the elderly with few exceptions (Mutchler, Prakash & Burr, 2007). Since the prevalence rates of common disability categories are significantly higher in the aged population, the majority of scientific studies regarding disability focus on them. This general trend becomes problematic when studying the second generation immigrants, especially relevant to post-1965 immigrants, in that there exist few observations of individual who is a second-generation and over certain age threshold to be included in studies. As a consequence, the second and further generation immigrants have received scant attention by scholars of the late 1990s and early 2000s periods. This gap raises concern not only because it unintentionally excludes the recent second generation immigrants, but also the evidence examining a more comprehensive range of age offers contradictory findings of the strength of the HIE (Mutchler, Prakash & Burr, 2007) compared with literature only including the aged group. Therefore, expanding observation's age range may contribute to clarifying the mixed evidence as well.

The final common drawback of previous studies is that most of these studies have suffered from a lack of causal interpretation due to their cross-sectional approaches. Since a cross-sectional survey does not follow individuals over time, most results concerning disability risk differences are limited to present the differences in prevalence rates. These results only suggest the association between nativity, generation status or SES and disability. Surprisingly, no published research investigates the change in disability rates amongst immigrants or compares disability risk trajectories between immigrants and natives by using the longitudinal dataset in the US context, even though this gap has been repeatedly discussed by scholars. Consequently, utilising longitudinal data will offer new insight into the topic.

2.4 Research Hypotheses

Following the discussion of the existing theories and previous findings, several predictions can be stated. Disability status can be employed as a health outcome variable. According to the HIE and the acculturation theory concerning the intergenerational adjustment of lifestyle to the destination countries, the disability risk among natives, the first and the second generation immigrants may be different. The first generations are expected to show the lowest risk, while the second generations will show a higher risk relevant to the former generation. Besides, according to the segmented assimilation hypothesis, the acculturation process will show heterogeneous effects across different racial/ethnic groups. Thus, the interaction between immigrant status and race is predicted as well. Notwithstanding the insufficient evidence, the 2.5 generation may show a lower disability prevalence and risk compared with the second generation. Moreover, the observed differences in the prevalence and risk of disability will diminish when the socioeconomic and demographic characteristics are considered. Given these predictions and the aforementioned research questions, the three hypotheses are proposed:

2.4.1 Hypothesis 1: The Healthy Immigrant Effect on Disability

The first central hypothesis is to test whether the HIE exists among the sample of immigrants collected recently. If the HIE holds, the first generation shows a lower disability risk while this advantage is negatively associated with the time spent in the destination country, i.e. the years since migration.

H 1-1 The first generation immigrant will show a lower prevalence and risk of disability compared with the native population without an immigrant background.

H 1-2 The years since migration will be positively correlated with disability prevalence and risk among the first generation immigrants.

2.4.2 Hypothesis 2: The Intergenerational Transmission of Disability Risk

The second main hypothesis is to explore how disability risk among the descendants of the first generation is different from their former generation and native-born counterparts. Notwithstanding the lack of existing empirical finding, the literature on the relationship between acculturation and health outcome offers some a priori predictions. Also, the heterogeneity between the second and the 2.5 generation is predicted to be linked to disability risk differences among them. Moreover, as some previous scholarly works argue, the heterogeneity in acculturation process among different racial/ethnic immigrant group in the US (Allen et al., 2007; Elo, Mehta & Huang, 2011) motivates further testing of the interaction between individuals immigrant generation status and race.

H 2-1 Compared with the first generation, the second generation immigrants will show higher prevalence and risk of disability

H 2-2 Compared with the second generation, the 2.5 generation immigrants will show a lower prevalence and risk of disability

H 2-3 The interaction between immigration status and the race will be statistically significant.

2.4.3 Hypothesis 3: The Social Determinants of Health

The final hypothesis is to confirm that the main implications of the theory of fundamental causes hold in explaining the differences in short term disability risk. Education, household income and marital status are chosen to be significant indicators of SES since they have been received as significant determinants of health and relatively easy to obtain and measure.

H 3-1 The education will be negatively correlated with disability prevalence and risk regardless of their immigration and generation status.

H 3-2 The household income will be negatively correlated with disability prevalence and risk regardless of their immigration and generation status.

H 3-2 The marital status will be negatively correlated with disability prevalence and risk regardless of their immigration and generation status.

3 Data

3.1 Data

3.1.1 Background on the Integrated Public Use Microdata Series Current Population Survey (IPUMS CPS)

The data used in order to test the aforementioned hypotheses are extracted from a subsample of IPUMS CPS. The CPS is a monthly US household survey conducted jointly by the US Census Bureau and the Bureau of Labour Statistics (BLS) (IPUMS CPS, 2020). These surveys gather information on education, labour force participation status, demographic characteristics, and other aspects of the nationally-representative samples of the US population. Due to its sample size, high response rates and extensive subject coverage, the CPS has been one of the most widely employed data resources in social science and economic research (Rivera Drew, Flood & Warren, 2014).

Despite their relevance to the scholarly community, the initial CPS is not without several compatibility issues, e.g. inconsistent naming of variables and coding changes. These challenges motivated the construction of an integrated set of data from the Current Population Survey (CPS) from 1962 forward, named as IPUMS CPS. IPUMS CPS provides personal- and household-level information, and it consists of the harmonised original variables from the CPS, i.e. the variables are identically coded to solve compatibility issue. This revision implies that researchers can create tabulations and multivariate analyses tailored to their research questions, exploiting their desired set of variables (IPUMS CPS, 2020). Also, researchers can access and download the IPUMS CPS data free of charge upon completion of the registration form.

There exists another remarkable improvement made with CPS data. Notwithstanding the embedded longitudinal design of the CPS, technical difficulties in linking surveys across different months prohibited researchers from exploiting the advantageous feature of the data. However, a pivotal contribution by Rivera Drew, Flood and Warren (2014) enables researchers can use IPUMS CPS data as longitudinal microdata which follows an individual up to 16 months after the first enumeration. According to the authors, their new linking methods are based on previously suggested algorithms by a group of scholars such as Madrian and Lefgren (2000) and Feng, (2001), but with more massive scale. They create a reliable household- and a personal-level identifier for the CPS BMS from 1989 and onwards to materialise several new options of longitudinal data-based research designs. The underlying mechanism of following a person is as follows: When a new observation comes into a BMS, the individual is followed up to 4 consecutive months. Then, the person disappears for 8 months and reappears for another 4 consecutive months. After the second observation period, they permanently disappear from the survey. Therefore, one individual from a particular household appears up to 8 times within sixteen months. This rotating design is also called the CPS 4-8-4 rotation panel design (Rivera Drew, Flood & Warren, 2014, p.123).

To make this pattern more recognisable, each respondent is given *Month-In-Sample (MIS)* number, which is 1 in the first month of enumeration and 8 in the last month. In summary, each person may have MIS numbered from 1 to 8. When linking two surveys in the subsequent year, researchers can use this number. For instance, if they wish to collect a one-year panel data by linking two BMSs conducted in January, they can include individuals whose MIS is 1 in the first year's January, and they should reappear in the next January BMS with MIS5. According to Rivera Drew, Flood and Warren (2014), in 2009 and 2010 survey, approximately 90% of the January MIS1 sample responded to every subsequent survey through April 2009, while circa 68% of them responded to all eight BMS through April 2010 (Rivera Drew, Flood & Warren, 2014, p.135). With little change in the sampling methods, researchers expect a similar level of attrition rates for IPUMS CPS data in recent surveys.

One application of this sampling method is the National Bureau of Economic Research (NBER) extracts of the CPS annual earning file, known as Merged Outgoing Rotation Groups (MORG). By including individuals whose MIS is either 4 or 8, a researcher can make sure an individual appears only once even though they collect monthly or yearly data from multiple years of CPS original data (NBER, 2020). This method can be used If they are interested in calculating a proportion such as a prevalence rate. This thesis also employs a sample extraction method based on MORG to overview the association between disability risk and explanatory other control variables by presenting the sample prevalence rates of disability categories and odds ratios.

3.1.2 Data Collection and Sample

The IPUMS CPS data is made accessible from 1962 forward. However, it is after the first BMS in 2009 that it provides harmonised disability variables with 6 different categories and one aggregated variable. Therefore, this paper extracted March BMS from 2009 to 2018. March was chosen due to the potential need for using it with Annual Social and Economic Supplement (ASEC), often referred to as the Annual Demographic File, although the MORG and liking identifier methods can be applied to any month of a year. As mentioned above, the raw data was download upon completion of the registration form on Feb 25, 2020. The data format of extraction was a DAT file, which is compatible with the STATA programme.

This thesis uses two types of samples of the same raw dataset for each analytic step. The first sample (Sample 1) is for a research design using IPUMS CPS as a cross-sectional survey. At this stage, this paper aims to compare prevalence rates of each disability category by immigrant generation status and to conduct multiple logistic regression analyses to examine the association between disability risk and the independent/control variables. The second sample (Sample 2) is following a research design involving a longitudinal or panel data design which follows an individual for one year to see the differences in short-term disability trajectory pattern. The detailed analytical strategy for this step is explicated in Section 4.

Figure 3.1 presents the selection procedure of Sample 1. It initially consists of 1,984,729 observations, which are selected from the raw data by the MORG methods, i.e. every individual whose MIS is 4 is selected from the March BMS between 2009 and 2018. Firstly, 390,801 respondents aged under 15 were excluded because the universe of disability variable is a person aged 15 or more. Secondly, 1,623 observations without nativity information were dropped. Thirdly, 14,640 individuals whose residency is unknown were also excluded. Fourthly, 24,687 individuals without household income information were excluded as well. Therefore, the final Sample 1 consists of 1,552,972 different individuals.

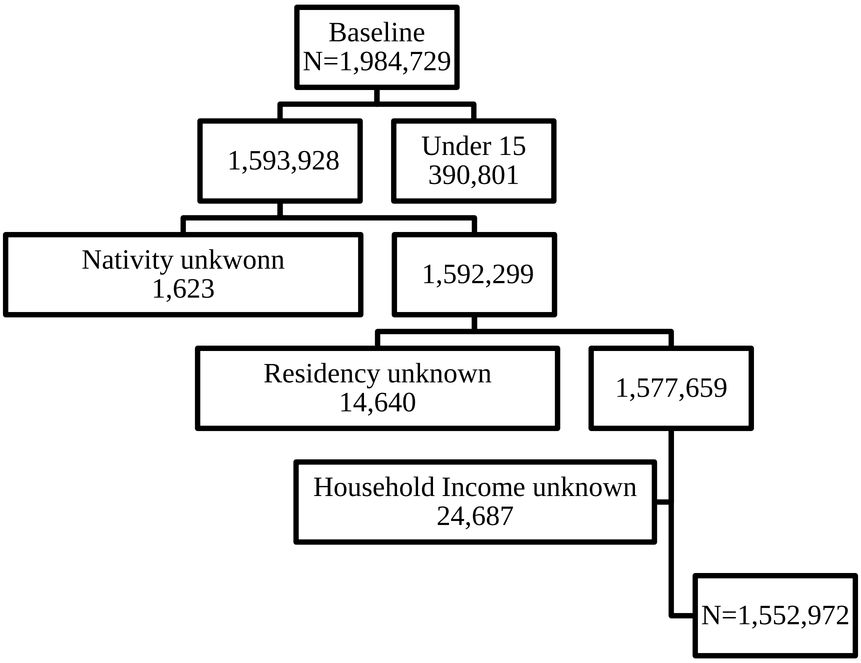

Figure 3.1 Sample 1 Selection Flow Chart

Likewise, Figure 3.2 describes the sample selection procedure of Sample 2. The baseline sample size is different because Sample 2 is not selected based on the MORG, but individuals who appeared in two March BMS from two consecutive years were included. For instance, individuals whose MIS is between 1 and 4 is initially selected, and they are only included when they remerged as respondents with MIS5 to MIS8 in March 2010.

This pre-screening procedure results in having 1,300,696 observations. The next few steps are identical to those of Sample 1, which gives 1,015,544 observations. Then, to fulfil the research design, which follows every individual without any difficulty in Year 1 for one year, 65,227 persons who answered they have at least one disability were excluded. This step is critical because every respondent included in Sample 2 should be disability-free in Year 1 by the research design. Another crucial issue is that the data should be constructed as balanced panel data. Therefore, about one-third of the remaining observations were ruled out, and 607,608 observations remained. The final step was to identify those who were mismatched by IPUMS CPS liking mechanism and excluding them from the sample. IPUMS CPS explicitly warns researchers that the matching does not guarantee perfect matching, and it suggests that researchers can qualify the reliability of matching by examining bey demographic characteristics such as sex, race, and nativity. This paper thoroughly followed the guideline and found 12,658 cases of mismatched observations. To summarise, after selection, the final version of Sample 2 consists of 594,920 observations or 297,460 individuals.

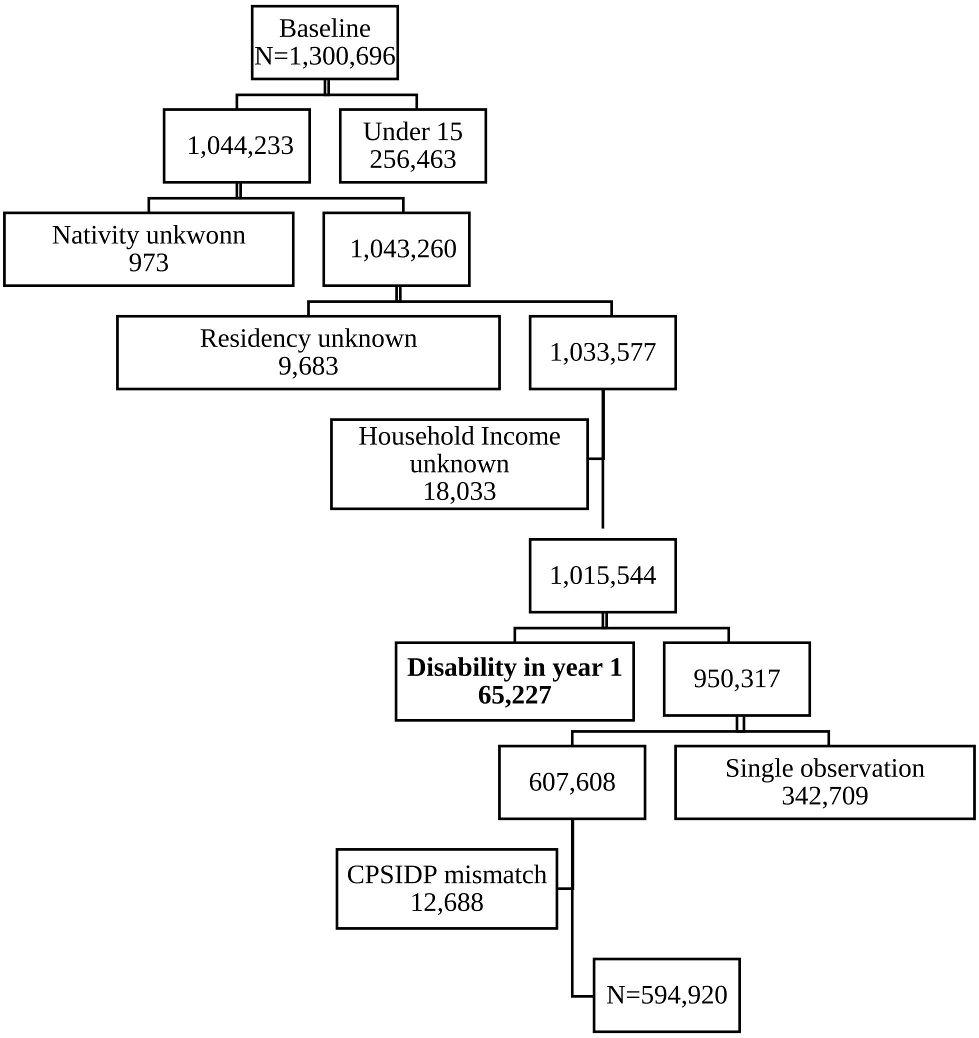

Figure 3.2 Sample 2 Selection Flow Chart

3.2 Variable

3.2.1 Dependent Variable

The dependant variable of the hypotheses mentioned above is a disability status. IPUMS CPS has six questions for each type of difficulty, i.e. DIFFHEAR (hearing), DIFFEYE (visual), DIFFREM (memory), DIFFPHYS (physical), DIFFMOB (mobility), DIFFCARE (self-caring difficulties), and one for aggregated status (DIFFANY). Following the description on the IPUMS CPS website, each dependent variable is defined follows:

(1) *Diffany* indicates "whether the respondent has any physical or cognitive difficulty, as measured by an affirmative response to at least one of the CPS' six cognitive and physical difficulties" (IPUMS CPS, 2020).

(2) *Hearing* indicates "whether the respondent is deaf or has serious difficulty hearing" (IPUMS CPS, 2020).

(3) *Visual* indicates "whether the respondent is blind or has serious difficulty seeing even with corrective lenses" (IPUMS CPS, 2020).

(4) *Memory* indicates "whether the respondent has cognitive difficulties, such as remembering, concentrating, or making decisions, because of a physical, mental, or emotional condition" (IPUMS CPS, 2020).

(5) *Physical* indicates "whether the respondent has serious difficulty walking or climbing stairs" (IPUMS CPS, 2020).

(6) *Mobility* indicates "whether the respondent has any physical, mental, or emotional condition lasting six months or more that makes it difficult or impossible to perform basic activities outside the home" by themselves (IPUMS CPS, 2020).

(7) *Selfcare* indicates "whether respondents have any physical or mental health condition that has lasted at least 6 months and makes it difficult for them to take care of their own personal needs, such as bathing, dressing, or getting around inside the home" (IPUMS CPS, 2020).

It should be noted that Mobility and Selfcare do not include "temporary health conditions, such as broken bones or pregnancies" (IPUMS CPS,2020). All disability variables are dummy variables and coded as 1 if a respondent answered *Yes* to each category. There was no missing value in disability status.

This thesis makes use of all the disability sub-categories throughout the empirical analysis. On the one hand, categories such as *Physical* and *Mobility* is predicted to be more associated with chronic health conditions and disadvantageous SES as previous research suggests (Minkler, Fuller-Thomson & Guralnik, 2006). Therefore, it will yield meaningful findings to analyse the prevalence and risk of them respectably. On the other hand, given the importance of considering the relevance of comorbidity in disease (Huh, Prause & Dooley, 2008), it may also be significant to distinguish individual without any difficulty with those who have at least one condition as well. Moreover, particularly for the second step of the analysis, the number of people whose disability status changed within a year for each category may be minimal—in this case, having the aggregated variable as the dependent variable would be the most desirable way. Therefore, the aggregated variable, *Diffany*, will be treated as a key dependent variable.

3.2.2 Independent Variable

The key explanatory variable is the immigrant's generation status. Using existing Nativity variable containing information on the place of birth of each respondent and parents', A set of 3 dummy variables were created: (1) The first generation refers to an immigrant born outside the US. Since the event of migration per se is considered as a major difference between the foreign-born and native-born group, this paper does not separate the first generation and 1.5 generation, i.e. the first generation who came to the destination country before they start schooling or becoming a teenager; (2) The second generation refers to an individual born in the US and has two foreign-born parents; (3) The 2.5 generation refers to an individual born in the US and have one native-born and one foreign-born parent. For both second and the 2.5 generations, a parent refers to their biological parents; (4) The reference group, hereafter referred as to *Native,* is a respondent born in the US with two native-born parents. Although this group may consist of a third-generation immigrant, which is defined as a native-born at least one of whose grandparent is a foreign-born, they are treated as a homogeneous reference group mainly due to lack of information on the nativity of grandparents and the fact that it is beyond this thesis's scope. Table 3.1 shows the proportion of each group in Sample 1 and 2.

Table 3.1 Frequency table by nativity/immigrant generation status

Immigrant generation status	N	Per cent	N	Per cent
	Sample 1		Sample 2	
Native	1,204,798	77.58	230,051	77.34
2.5 generation	57,078	3.68	10,928	3.67
Second generation	66,810	4.3	12,576	4.23
First generation	224,286	14.44	43,905	14.76

3.2.3 Other Covariates

The empirical analyses of this thesis involve several covariates to control the impact of demographic and socioeconomic characteristics on an individual's disability risk. The included covariates are described as below:

Age: *Age* is a continuous variable from 0 to 85. The respondents aged over 85 are coded as 85 as well. Because the universe of disability variables are individuals aged 15 or over, the actual range of age variable is 15 to 85. The Age-squared variable, *age2,* is also included to capture the potential non-linear effect of age on the probability of having a disability.

Sex: *Sex* is a dummy variable which is coded as 1 for female and 0 for male.

Race: Based on the original race variable of the CPS, it has been re-constructed as a group of dummy variables: Non-Hispanic White, Hispanic White, African American, Indigenous, Asian American, Pacific Islander, and *Other*. Indigenous category refers to American Indian, Aleut and Eskimo populations (IPUMS CPS, 2020). *Other* refers to those who reported that their race is mixed. The term *race* is used as the most published literature regarding US context uses this term, while *ethnicity* can be an alternative way to describe this feature in different contexts. The exact proportion is presented in Table 3.2.

Table 3.2 Frequency table by race

Race	N	Per cent	N	Per cent
	Sample 1		Sample 2	
White, non-Hispanic (Reference)	1,090,672	70.23	213704	71.84
White, Hispanic	175,907	11.33	32933	11.07
African American	159,971	10.3	27329	9.19
Indigenous	17,536	1.13	2754	0.93
Asian American	76,765	4.94	15227	5.12
Pacific Islander	6,572	0.42	1115	0.37
Other	25,549	1.65	4398	1.48

Education: Educational attainment is coded as a group of dummy variables. Instead of years of schooling, the CPS provides the highest years of school or degree completed. It has 4 categories: Less than high school diploma, which is the reference group, high school diploma, some college, and college degree or more. Table 3.3 presents an overview of the variable.

Table 3.3 Frequency table by education

Educational attainment	Number	Per cent	Number	Per cent
	Sample 1		Sample 2	
Less than high school	244,388	15.74	43,763	14.71
High school diploma	448,643	28.89	81,592	27.43
Some college	276,284	17.79	51,991	17.48
College degree	583,657	37.58	120,114	40.38

Marital status: The original marital status variable has several categories. This variable is adjusted to be a dummy variable which is coded as one if a respondent is married and living with their partner and 0 otherwise.

Residence: Residence variable is a dummy variable which indicates that a respondent lives in a metropolitan area or not, regardless of the detailed location (inside/outside the centre city), it is coded as one if a respondent answered Yes and 0 otherwise.

Household Income: This variable reports the total income for the respondent's family. Amounts are expressed as they were reported to the interviewer, i.e., in the survey year's dollar amounts. Therefore, IPUMS CPS states that it may be possible for researchers to adjust the original value using price index such as the CPI. However, it is not feasible to adjust it since it is given as a categorical variable. Therefore, notwithstanding the potential limitation, this paper re-constructed this variable without adjustment. For parsimony, it was revised as a set of 4 dummy variables dividing the sample into five categories, with each category including approximately 20 to 25 per cent of the sample, except for the highest income group, which accounts for 9.8% of the respondents in sample 1. In Sample 2, the highest income group is over-represented, while the lowest group is under-represented compared with Sample 1. Dropping individuals with current disability condition might affect the distribution. Table 3.4 presents the detail.

Table 3.4 Frequency by household income

Household income	Number	Per cent	Number	Per cent
	Sample 1		Sample 2	
Under $25,000	325,664	20.97	46,615	15.67
$25,000 - 49,999	386,987	24.92	71,086	23.9
$50,000 - 74,999	299,518	19.29	61,792	20.77
$75,000 - 149,999	387,913	24.98	84,648	28.46
$150,000 or more	152,890	9.84	33,319	11.2

Veteran Status: Veteran Status is a dummy variable identifying veterans, that is defined as an individual who served in the military forces of the US, i.e. Army, Navy, Air Force, Marine Corps, or Coast Guard in time of war or peace, but who were dismissed at the time of the survey (IPUMS CPS, 2020).

State: Although the geographical variable is beyond this thesis's scope, State variable referring to different States in the US is included to estimate the State-fixed effect.

Year: There may exist unobserved macroeconomic trend associated with individual's short term disability trajectory which distorts the estimation. Therefore, this thesis employs *Year* variable to calculate the Year-fixed effect estimator.

Years Since Migration (YSM): This variable means how long a first generation immigrant has resided in the US since their immigration. It is constructed as a set of dummy variables, i.e. less than 10 years, 20 years, 30 years, and 30 years or more. The BMS has a severe compatibility issue regarding this information. The coding method has been changed several times, resulting in each category, including different periods, e.g. one, two, three or five years. Therefore, this variable should be interpreted as a proxy of YSM with potential measurement errors. More information on YSM variable is presented in Table 3.5 and excluded in the full summary statistics table since it is only relevant to the first generations.

Table 3.5 Frequency table of YSM, Sample 2, Year 1

Years Since Migration	Number	Per cent
Less than 10 years	13,623	31.04
10-19 years	7,611	17.34
20-29 years	9,285	21.16
30 years +	13,367	30.46
Number of Observation	43,886	100

The summary statistics of Sample 1 is presented in Table 3.6. The summary statistics of Sample 2 is provided in Appendix A for comparison.

Table 3.6 Summary statistics of Sample 1

VARIABLES	(1) mean	(2) s.d.	(3) min	(4) max
Year	2,014	2.830	2,009	2,018
State (FIPS code)	28.03	16.03	1	56
Age	46.27	18.72	15	85
Age squared	2,491	1819	225	7,225
Hearing	0.0382	0.192	0	1
Visual	0.0192	0.137	0	1
Memory	0.0385	0.192	0	1
Physical	0.0745	0.263	0	1
Mobility	0.0415	0.199	0	1
Selfcare	0.0208	0.143	0	1
Diffany	0.128	0.334	0	1
2.5 generation	0.0368	0.188	0	1
Second generation	0.0430	0.203	0	1
First generation	0.144	0.352	0	1
High school	0.289	0.453	0	1
Some college	0.178	0.382	0	1
College degree or more	0.376	0.484	0	1
Metro	0.798	0.401	0	1
Marital status	0.521	0.500	0	1
Hispanic White	0.113	0.317	0	1
African American	0.103	0.304	0	1
Indigenous	0.0113	0.106	0	1
Asian	0.0494	0.217	0	1
Pacific islander	0.00423	0.0649	0	1
Other (race)	0.0165	0.127	0	1
$25,000 - 49,999	0.249	0.433	0	1
$50,000 - 74,999	0.193	0.395	0	1
$75,000 - 149,999	0.250	0.433	0	1
$150,000 or more	0.0984	0.298	0	1
Veteran status	0.0856	0.280	0	1
Sex	0.522	0.500	0	1
Number of Observation: 1,552,972				

4 Methods

The research hypotheses mentioned above state that the probability of having a disability is the outcome variable of the analysis. This probability is identified with the dependent variable described in Section 3.2. Therefore, the dependent variable is a typical case of Limited Dependent Variable (LDV), which is defined as a dependent variable whose range of values is restricted (Wooldridge, 2016). To test these hypotheses, two models are proposed: Model 1 uses multiple regression analysis to investigate the association between the dependent variable and independent variable with a vector of demographic and socioeconomic characteristics included. Sample 1, extracted from the raw data by the MORG method, will be used for this analytic step. Next, Model 2 uses a similar multiple regression analysis in a different setting to make it as similar to panel regression as possible. It uses Sample 2, which only consists of individuals without any disability in Year 1. Then, it regresses the likelihood of having a disability after a Year (Year2) on the independent variable and covariates measured at Year 1.

The rest of this section will explain the theoretical foundation of the methods will be elaborated, and it is followed by the specification of two models with a detailed explanation of the reasoning of Model 2 design. All the statistical analyses were carried out using STATA, version 16.1.

4.1 Logistic Regression Model

A simple regression method of predicting the maximum likelihood estimate in LDV setting is a Linear Probability Model (LPM). However, LPM has two crucial drawbacks: Firstly, the predicted value of probability could be outside the [0, 1] interval, which violates the fundamental axiom of probability. Another issue is that LPM assumes a constant marginal effect of the independent variable. Given the fact that the primary explanatory variable is a set of dummy variables in this study, this problem may not be severe. Notwithstanding, the first issue usually motivates researchers to find non-linear alternatives such as probit or logit model, which assume the probabilistic distribution of the outcome variable (Wooldridge, 2016). Following most previous studies concerning similar analyses, this thesis also uses the logit model for the statistical analyses.

In a simple LDV of a binary response model, one is particularly interested in the response probability.

$$P(y = 1|X) = P(y = 1|x_1, x_2, x_3, \dots, x_k), \text{ [4.1.1]}$$

where X is a vector consisting of the full set of explanatory variables. In this study, X contains the primary explanatory variable, the immigrant generation status and other demographic and socioeconomic variables as described in Section 3.2. The right-hand-side of equation [4.1.1] can be written as a function G, which is a function of X and whose value is strictly limited between zero and one: $0 < G(z) < 1$, for all real number z.

$$P(y = 1|X) = G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) = G(\beta_0 + X\beta), \text{[4.1.2]}$$

where $X\beta = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$, and G suffice the above restriction. For parsimony, $X\beta$ can also be denoted as a real number w. In the logit model, G is the logistic function as follows:

$$G(z) = \frac{\exp(z)}{1+\exp(z)}, \text{[4.1.3]}$$

which is between 0 and 1 for all ream numbers z. This function is the cumulative distribution function for a standard logistic random variable (Wooldridge, 2016). Therefore, the logit model of interest can be written as follows:

$$G(w) = \frac{\exp(w)}{1+\exp(w)} = \frac{\exp(X\beta)}{1+\exp(X\beta)}, \text{[4.1.4]}$$

Also, if $P(y = 1|x_i)$ is defined as p_i, the log odds ratio of an event can be expressed as follows:

$$\log\left(\frac{p_i}{1-p_i}\right) = x_i'\beta, \text{[4.1.5]}$$

4.2 Model Specification

4.2.1 Model 1

The first model consists of multiple logistics regressions using the following specifications. Firstly, the baseline model follows this specification:

$$y_i = \beta_0 + \beta_1 First_i + \beta_2 Second_i + \beta_3 Secondhalf_i + \sigma Year_i + \tau State_i + \varepsilon_i, \text{[4.2.1]}$$

Where y_i is the likelihood of having a disability of the respondent i at the time of the interview, $First_i, Second_i, and\ Secondhalf_i$ the independent variables, i.e. the immigrant generation status of the respondent i, and ε_i the error term. Besides, $Year_i$ and $State_i$ variables allow the model to control for unobserved regional-level effects and time trend. As discussed in the previous section, this base model, i.e. the short model, is suspected of generating a biased and inconsistent estimator because of the omitted variable bias since it does not include key demographic and socioeconomic variables which are probably associated with the outcome variable. Therefore, the extended version of the equation 4.2.1 includes such covariates. The extension has two steps: the first step follows this specification:

$$y_i = \beta_0 + \beta_1 First_i + \beta_2 Second_i + \beta_3 Secondhalf_i + D_i'\beta + \sigma Year_i + \tau State_i + u_i, \text{[4.2.2]}$$

Where $D_i{}'$ refers to the demographic characteristic variables such as age, age squared, sex, and race of the respondent i, and u_i the error term for the revised model.

$$y_i = \beta_0 + \beta_1 First_i + \beta_2 Second_i + \beta_3 Secondhalf_i + D_i'\beta + S_i'\gamma + \sigma Year_i + \tau State_i + \mathrm{v_i}, \quad [4.2.3]$$

The third alternative, namely the second extended model, also includes marital status, household income, residential area, and veteran status of the individual i, which is referred as to vector $S_i{}'$, and v_i the error term. The reason behind separating covariates included in the second and the third model is to see which group of variables is more involved with change in the coefficient of the independent variable.

$$y_i = \beta_0 + I_i{}'\theta + \delta I_i{}' * R_i{}' + D'\beta + S'\gamma + \sigma Year_i + \tau State_i + w_i, \quad [4.2.4]$$

The final version of Model 1 includes a vector of the independent variables ($I_i{}'$), a vector of *race* variable ($R_i{}'$), an interaction term between the vector of the independent variable and race variable ($I_i{}' * R_i{}'$), and w_i the error term for equation 4.2.4. This regression equation is to test the H 2-3 regarding Model 1 directly.

4.2.2 Model 2

The main objective of the second model is to overcome the common limitation of the previous research presenting the mere association and find rigorous evidence for a causal interpretation. As described in Section 3.1, the recent update of IPUMS CPS enables researchers to use it for a longitudinal microdata analysis setting. Therefore, this thesis constructed Sample 2 as a dataset which observes an individual twice: Once in March, Year 1 and once more in March, Year2. It is identical to an annual penal data with two waves.

The fact that it only has two waves is one of the major limitations of this study. However, when other longitudinal microdata with more extended periods of observation is not available, this method is a reasonable alternative. For instance, the Medical Expenditure Panel Survey in the US, also provided by IPUMS, could have been an alternative only if it contained information on detailed information to identify individual's nativity and immigrant generation status. A couple of other candidates also fail to provide sufficient information to test the research hypotheses as well. Besides, existing research such as Jass and Massey (2004) also investigate short-term health trajectories of the immigrant population during a year, faced with similar data availability issue. Therefore, although limited, this approach is a feasible alternative.

Provided that observing short-term disability risk trajectory is a reasonable approach, another concern is which penal regression methods to choose for the analysis. The most critical issue here is the endogeneity issue between unobserved individual characteristics and the independent variable. The Random effect (RE) model assumes that these unobserved characteristics are exogenous. Consequently, the model is prone to the endogeneity problem. This problem poses a severe problem since RE estimator with endogeneity problem yields inconsistent estimates. By employing the *one-way error component model*, the Fixed-effect model (FE) partly solves this problem. FE method controls for both observed and unobserved time-invariant factors, which implies that it may give a consistent estimator without the assumption needed in the RE model. Still, the zero conditional mean assumption between unobserved time-variant characteristics and the independent must be fulfilled for the FE model to generate consistent estimates, which is relatively realistic assumption compared with the one of the RE model (Angrist & Pischke, 2009).

Nonetheless, a standard FE model cannot be used in this analysis because immigrant generation status is per se a time-invariant variable. At the same time, the result from Hausman testing suggests that the assumption for the RE model is likely to be violated in Sample 2. Another possibility is a revision of the RE model proposed by Mundlak (1978), also known as the *Mundlak correction*. This method adds group means of the independent variables into the regression equation to let groups means capture the time-invariant factors, and the rest part deals with only time-variant factors, similar to the mechanism of FE model. Still, researchers can get a coefficient of time-consistent variables such as sex or nativity. Therefore, this regression strategy is often used as a compromise between traditional RE and FE regression (Greene, 2012). However, this method does not allow an independent variable to be a dummy or factor variables. In summary, both traditional FE and Mundlak correction regression is not feasible for the empirical analysis of this thesis.

Another possible option to circumvent the reverse causality issue found in most cross-sectional studies and to make use of the longitudinal design of Sample 2 is to use lagged-independent and control variables. It means that likelihood of having a disability after one year is regressed on demographic and socioeconomic status at the first observation. Since the primary explanatory variable and most covariates are time-invariant, this setting is feasible. The essential requirement for this setting is to exclude every respondent who already had at least one disability at the time of the first survey, and as shown in Section 3, this exclusion does not seriously reduce the sample size. Furthermore, by comparing results from Model 1, which include all the respondent and the second model, whether this exclusion is involved with a particular selection, can be examined as well.

The model specifications are identical to those of Model 1 except that all the right-hand-side variables are lagged, and this feature is reflected on its time subscript, $t-1$. At this stage of analysis, two the full-length models which include both demographic and socioeconomic variables with and without the interaction between the independent variable and race are compared with each other:

$$y_i = \beta_0 + \beta_1 First_{it-1} + \beta_2 Second_{it-1} + \beta_3 Secondhalf_{it-1} + D'_{it-1}\beta + S'_{it-1}\gamma + \sigma Year_{it-1} + \tau State_{it-1} + \mathrm{v_i}, \text{ [4.2.5]}$$

where y_{it} indicates a likelihood of having a disability of individual i at time t, and other variables with $t-1$ notation that those variables are lagged by one year. The model with interaction terms to test the H 2-3 follows this specification:

$$y_i = \beta_0 + I'_{it-1}\theta + \delta I'_{it-1} * R'_{it-1} + D'_{it-1}\beta + S'_{it-1}\gamma + \sigma Year_{it-1} + \tau State_{it-1} + w_i, \text{ [4.2.6]}$$

Furthermore, this thesis uses a weight variable for the BMS users named as WTFIN, provided by IPUMS CPS. Although there exist a couple perceived compatibility issues with this weight, they are irrelevant to the samples used in this thesis.

4.3 Robustness Testing

To further investigate the validity of the main results, two sensitivity tests are proposed: The first testing is restricting the age range of the original sample to see if the magnitude or significance of the coefficient changes. With the same model specifications, the dependent variable is regressed with an age restriction of over 30, 45, and 65. According to Kenny et al. (2008), age 45 is a meaningful threshold where the risk of functional limitation or disability starts to elevate at a faster pace. Therefore, this cut-off may give a relevant sample restriction if the impact of generation status is heterogenous between the younger and the older groups. Likewise, 65 is pointed out as another notable turning point by previous research. For instance, Gee, Kobayashi and Prus (2004) argue that the HIE disappears for those over 65 in their sample. Although the authors fail to elaborate on the exact mechanism of the phenomenon, the argument can be tested as a part of the sensitivity testing.

The second testing is to divide the sample by sex and perform the regression respectably. There remains a possibility that the impact of the independent variable and other control variable is different between male and female respondents. The systematic review concerning the difference in disability risk by sex states that after controlling the relevant covariates, the net differences is insignificant (Rodrigues, Facchini, Thumé & Maia, 2009). However, research in the field has paid attention to gender differences in health amongst the immigrant population since there might exist a peculiar mechanism which leads to differences in self-selection or adjustment to the destination society (e.g. De Souza & Fuller-Thomson, 2013; Warner & Brown, 2011). As a consequent, this thesis attempts to see if the results change when the total sample is divided into male and female groups.

5 Results

5.1 Descriptive Results

The weighted prevalence rates of each disability category by nativity/immigration generation status can be found in Table 5.1. Firstly, compared with the total sample average and Native group, the first generations show lower prevalence rates in all six categories of disability and aggerated disability. The second generation group also shows lower prevalence rates compared with Native, but higher compared with the First generation group. Meanwhile, the 2.5 generation group shows a similar level of prevalence rates as the Native group. This group shows slightly lower prevalence rates in *Diffany*, *Memory*, *Mobility* and *Selfcare* variables while it shows approximately the same or even higher risk in *Hearing* and *Visual* variables.

In summary, the descriptive results seem to follow the previous findings, which may suggest the superior health status of the First generation (HIE) and diminished advantage amongst the Second generation. The notable point is that the 2.5 generation shows approximately identical patterns compared to the reference group in Sample 1. However, these results are without conditioning on any control variable or year/state fixed effects.

Table 5.1 Disability prevalence rates by nativity/immigrant generation status in Sample 1

	Native	2.5 generation	Second generation	First generation	Total
Diffany	0.132	0.130	0.092	0.076	0.120
Hearing	0.038	0.041	0.031	0.019	0.034
Visual	0.019	0.019	0.015	0.014	0.018
Memory	0.041	0.038	0.029	0.022	0.037
Physical	0.077	0.074	0.051	0.046	0.070
Mobility	0.043	0.041	0.037	0.031	0.040
Selfcare	0.021	0.020	0.018	0.016	0.020

As described in Section 2, the racial difference in disability risk is also a vital issue itself and in relevance to immigration. The weighted prevalence rates of each disability category by race can be seen in Table 5.1. The non-Hispanic White group is ranked middle among various racial groups and a similar level of prevalence compared with average rates in Sample 1. The lowest rates can be seen amongst the Asian group, followed by Pacific Islanders and white Hispanic groups. Meanwhile, African American and Indigenous group show increased risk of each category of disability. For *Diffany* and *Hearing* categories, non-Hispanic White group shows higher rates while the African American group shows the opposite results in the rest 5 categories. The indigenous group always shows similar or higher prevalence rates relevant to other groups. Although this result does not explain the origin of racial differences without controlling for any covariate, it seems to be in accordance with previous findings on racial disparity in disability risk in the US.

Table 5.2 Disability prevalence rates by race in Sample 1

	Non-Hispanic White	Hispanic White	African American	Indigenous	Asian	Pacific Islander	Other	Total
Diffany	0.131	0.079	0.129	0.161	0.057	0.076	0.144	0.120
Hearing	0.042	0.018	0.019	0.042	0.017	0.020	0.033	0.034
Visual	0.019	0.015	0.023	0.032	0.009	0.013	0.023	0.018
Memory	0.039	0.026	0.045	0.058	0.017	0.019	0.064	0.037
Physical	0.076	0.044	0.083	0.097	0.032	0.046	0.075	0.070
Mobility	0.042	0.029	0.049	0.060	0.025	0.027	0.046	0.040
Selfcare	0.021	0.016	0.026	0.033	0.012	0.015	0.022	0.020

In analyses concerning Sample 2, variations in disability status are critical since an insufficient number of changes threatens the validity of the research. The number of changes in each disability category by different age groups is presented in Table 5.3. As the observation period is one year, those who experience any change in disability status accounts for approximately 5.1% of the total number of observations in Sample 2. The table also shows which category of disability is prone to change within a short period, i.e. physical, hearing, memory and mobility difficulty are relatively changeable while self-care and visual difficult show fewer number of changes.

Table 5.3 Number of changes in disability status by age group, Sample 2

Variable	Diffany	Hearing	Visual	Memory	Physical	Mobility	Selfcare
Total	15,223	4,018	1,784	3,774	8,186	3,649	1,662
30+	14,172	3,896	1,636	3,165	7,943	3,376	1,579
(%)	93	97	92	84	97	93	95
45+	12,538	3,603	1,425	2,534	7,263	2,975	1,394
(%)	82	90	80	67	89	82	84
65+	7,003	2,448	763	1,209	4,060	1,816	772
(%)	46	61	43	32	50	50	46

Dividing the whole sample into different age groups make it possible to see which disability condition occurs relatively at younger ages and how many cases of changes the additional models for the sensitivity testing lose by introducing the age threshold. According to Table 5.3, using 30+ age limit leads to losing approximately 7% of the respondents who already had a status change regarding *Diffany*, but at the same time, this restriction results in losing 16% of the respondents who reported memory difficulty under 30. Indeed, memory difficulty occurs more rapidly compared with other categories and hearing difficulty appears to be the slowest one. Similarly, introducing age 45+ and 65+ limitation leads to excluding nearly 20% and 54% of the respondents in Sample 2.

5.2 Model 1

The multiple logistic regression results of Model 1 regarding *Diffany* is presented in Table 5.4. As described in Section 4.2.1, three regression results are presented from the baseline model to the full model, including a vector of demographic and socioeconomic covariates, along with state/year fixed effects. In Column (1), Table 5.4, odds ratios of the *Second generation* and the *First generation* are all less than one and statistically significant at 1 % level (0.718 and 0.583 respectably). Meanwhile, the *2.5 generation* shows an odds ration slightly higher than 1 (1.034), which is significant at 5 % level. This result is predictable from the simple comparison of prevalence rates amongst those groups. After controlling for *Race, Age,* and *Sex*, in Column (2), the pattern shows minor changes. The most notable change is that the coefficient of the 2.5 generation group becomes lower than one, and it is significant at 1% level. The other two group's odds ratio is still significantly lower than 1 and significant.

Concerning demographic characteristics, three points are worth paying attention. After controlling for *Age* and *Sex*, the Hispanic White shows a higher risk of disability compared with the non-Hispanic White group. Also, Pacific Islander's coefficient is significantly higher than 1. These two results contradict the result presented in Table 5.2. The odds ratios of Asian, African American, and Indigenous groups indicate similar results relevant to Table 5.2, i.e. Asian group's odds ratio is significantly less than 1, and African American and Indigenous groups' odds ratios are higher than 1. Also, both *Age* and *Age-squared* variable show significant results, which implies that the assumption of the non-linear relationship between age and disability risk may be valid. Furthermore, compared with Male, Female shows smaller odds ratio (0.982), and this result is significant at 1 % level. This finding may contradict the previous findings which support female has at least equal or higher risk of having a non-fatal disability risk. The fact that other socioeconomic variables are not controlled, and the sample includes a broader age range might affect the result.

The third model in Column (3) presents the regression result, including all the covariates. The coefficient of the 2.5 generation becomes 1.016 and not significant at the 10% level. However, the odds ratios of the second and first generation groups show little change in magnitude or significance. The pattern of racial differences changes, as well. After controlling for socioeconomic covariates, Hispanic White and African American group shows odds ratios significantly less than 1. The other result regarding the race variable is similarly relevant to

Model 1-2. Compared with the reference group, the married group and *Metropolitan* group show significantly lower odds ratios. Concerning education level and household income, distinct gradients, which indicates that those two variables are negatively associated with disability risk, are observed. Moreover, as the models include more control variables, the pseudo-R-squared increases from 0.001 to 0.194, as shown in Table 5.4.

Table 5.4 Multiple logistic regression results of Model 1

Dependent variable=*Diffany*	(1) Model 1-1	(2) Model 1-2	(3) Model 1-3
2.5 generation	1.034**	0.944***	1.016
	(0.015)	(0.015)	(0.017)
Second generation	0.718***	0.833***	0.868***
	(0.011)	(0.014)	(0.015)
First generation	0.583***	0.631***	0.590***
	(0.005)	(0.007)	(0.008)
Hispanic white		1.198***	0.784***
		(0.015)	(0.011)
African American		1.411***	0.928***
		(0.014)	(0.010)
Indigenous		2.089***	1.328***
		(0.060)	(0.039)
Asian		0.754***	0.794***
		(0.015)	(0.017)
Pacific Islander		1.174***	0.972
		(0.073)	(0.062)
Other		2.072***	1.648***
		(0.048)	(0.039)
Age		0.986***	1.038***
		(0.001)	(0.001)
Age squared		1.001***	1.000***
		(0.000)	(0.000)
Sex		0.982***	0.959***
		(0.006)	(0.006)
Marital status			0.603***
			(0.004)
Metropolitan			0.947***
			(0.008)
High school			0.655***
			(0.006)
Some college			0.603***
			(0.006)
College degree			0.456***
			(0.005)
$25,000 - 49,999			0.543***
			(0.004)

$50,000 - 74,999			0.414***
			(0.004)
$75,000 - 149,999			0.314***
			(0.003)
$150,000 and over			0.238***
			(0.004)
Veteran status			1.392***
			(0.014)
Constant	0.185***	0.057***	0.071***
	(0.004)	(0.002)	(0.002)
Year FE	YES	YES	YES
State FE	YES	YES	YES
Pseudo R-squared	0.00966	0.139	0.194

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Next, a summary of the regression analysis considering the interaction between immigrant generation status and race variable is presented in Table 5.5. The interaction effects in terms of odds ratios are directly calculated based on the full table in Appendix B. It shows the differences in disability risk within each racial group by the respondent's generation status, i.e. the Natives in each group becomes the reference group for their race. Generally, the 2.5 generation shows the most elevated risk, followed by the second and first generations. However, the differences between the 2.5 and the second generation are more pronounced within Asian and Hispanic White groups, whereas the difference is minimal among Non-Hispanic White, with other groups showing modest variations. Nonetheless, reading this table needs additional caution since not every interaction term is statistically significant. The relevant information can also be found in Appendix B.

Table 5.5 Summary of the regression results including interaction terms, Model 1

Dependent variable=diffany	Native	2.5 generation	Second generation	First generation
Non-Hispanic White	1	0.972	0.965	0.748
Hispanic White	1	1.110	0.740	0.460
African American	1	1.095	0.811	0.516
Indigenous	1	0.918	0.684	0.379
Asian	1	1.209	0.820	0.728
Pacific Islander	1	1.198	1.027	0.914
Other	1	0.798	0.428	0.406

Finally, the summary of the results regarding the sub-categories of disability variable is presented in Table 5.6. For all categories, the first generation group shows consistent patterns of showing odds ratios significantly less than 1. The coefficients of the 2.5 generation are significant in 3 categories, i.e. memory, mobility, and self-care difficulty, while those of the second generation are significant in 4 categories, i.e. memory, physical, mobility, and self-care difficulty. However, the overall pattern is following that of *Diffany*. The full regression results of the six disability categories, based on Model 1-3 are presented in Appendix C.

Table 5.6 Summary of Multiple regression results of each disability category, without interaction

	Hearing	*Visual*	*Memory*	*Physical*	*Mobility*	*Selfcare*
2.5 generation	1.008	0.981	0.938**	1.007	0.934***	0.919**
	(0.026)	(0.036)	(0.025)	(0.021)	(0.024)	(0.033)
Second generation	1.009	0.946	0.774***	0.929***	0.921***	0.921**
	(0.027)	(0.035)	(0.022)	(0.021)	(0.024)	(0.033)
First generation	0.691***	0.745***	0.548***	0.614***	0.718***	0.752***
	(0.016)	(0.021)	(0.012)	(0.010)	(0.014)	(0.020)

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

In summary, the main results of Model 1 imply a significantly negative association between being the first and the second generation and having a disability. In contrast, the association between being the 2.5 generation and having a disability condition is insignificant regarding *Diffany* variable. The coefficients of covariates are also in line with existing literature, except for that female group shows a lower risk of having a disability after controlling for other demographic and socioeconomic characteristics. However, these results only suggest associations between the independent and the dependent variables, and therefore, it is insufficient to draw any causal interpretation out of them.

5.3 Model 2

The logistic regression results of Model 2 regarding *Diffany* variable is presented in Table 5.7. As described in Section 4, the major difference between Model 1 and 2 is that Model 1 only present the probability of having a disability conditioning on covariates at the time of the survey, while Model 2 only includes respondents having no disability condition in Year 1 and estimates the risk of having a disability status change within a year, observed in Year 2, conditioning on covariates observed in Year 1. In Column (1), Table 5.7, the result of the baseline model is presented. Compared with the reference group, all the other group show odds ratio less than 1, but the coefficient of the *2.5 generation* is insignificant. However, from the second model in Column (2), the 2.5 generation group consistently shows odds ratios less than 1, 0.833 and 0.882 respectably, which is significant at 1% level. The result from the full model in Column (3) indicates that the difference between the 2.5 and the second generation group is minimal, while the first generation group shows the most substantial differences. Regarding the race variable, the coefficient of the *Asian* is not different from the reference group, which is the most notable change from Model 1.

Other socioeconomic variables show similar results relevant to Model 1, except that the coefficient of residence variable is not significant in Model 2. Apparent disability risk disparity by education level and household income is shown in Column (3). The coefficient of the sex variable also indicates that lower risk of having any disability for the female group is significant, although it is not shown in the model including only demographic characteristics in Column (2). Similar to Model 1, as regression equations include more control variables, the pseudo-R-squared increases from 0.005 to 0.135.

Table 5.7 Multiple regression results of Model 2

Dependent variable=*Diffany*	(1) Model 2-1	(2) Model 2-2	(3) Model 2-3
2.5 generation	0.925	0.833***	0.882**
	(0.047)	(0.044)	(0.048)
Second generation	0.747***	0.862***	0.881**
	(0.038)	(0.047)	(0.050)
First generation	0.725***	0.774***	0.715***
	(0.022)	(0.029)	(0.029)
Hispanic white		1.246***	0.870***
		(0.054)	(0.041)
African American		1.524***	1.108***
		(0.051)	(0.038)
Indigenous		1.822***	1.272**
		(0.182)	(0.128)
Asian		0.926	0.977
		(0.058)	(0.063)
Pacific Islander		1.191	1.051
		(0.254)	(0.224)
Other		1.318***	1.105
		(0.127)	(0.107)
Age		0.982***	1.027***
		(0.003)	(0.003)
Age squared		1.001***	1.000***
		(0.000)	(0.000)
Sex		0.980	0.944***
		(0.019)	(0.021)
Marital status			0.718***
			(0.016)
Metropolitan			0.966
			(0.026)
High school			0.782***
			(0.024)
Some college			0.666***
			(0.024)
College degree			0.531***
			(0.018)
$25,000 - 49,999			0.610***
			(0.016)
$50,000 - 74,999			0.499***
			(0.016)
$75,000 - 149,999			0.391***
			(0.013)
$150,000 and over			0.296***
			(0.016)

Veteran status			1.186***
			(0.039)
Constant	0.063***	0.021***	0.025***
	(0.004)	(0.002)	(0.003)
Year FE	YES	YES	YES
State FE	YES	YES	YES
Pseudo R-squared	0.00514	0.105	0.135

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

As in Model 1, the logistic regression, including the interaction between immigrant generation status and the race is also conducted as a part of Model 2. The key result is presented in Table 5.8. The table is constructed in the same way as Table 5.5 is constructed. For Non-Hispanic White, Indigenous and Asian groups, the 2.5 generation now shows smaller odds ratios compared to the second generation while the results are opposite among Hispanic White, African American, Pacific Islander and Other groups. However, most of the interaction terms suffer from a lack of statistical significance. Also, the dramatic results observed among Indigenous, Pacific Islander and Other groups is likely due to the small observation numbers in Sample 2. Consequently, it needs extra caution to interpret the figures shown in Table 5.8 as well. The full version of the regression table is in Appendix D.

Table 5.8 Summary of the regression results including interaction terms, Model 2

Dependent variable=diffany	Native	2.5 generation	Second generation	First generation
Non-Hispanic White	1	0.871	0.952	0.806
Hispanic White	1	1.002	0.946	0.680
African American	1	0.780	0.488	0.570
Indigenous	1	0.505	0.530	0.210
Asian	1	0.714	0.782	0.981
Pacific Islander	1	1.984	0.401	0.458
Other	1	1.106	0.284	0.835

Furthermore, the summary of the logistic regression results of different categories of disability conditions is shown in Table 5.9. The differences in patterns are pronounced. For instance, for *Visual*, *Memory*, and *Selfcare* categories, the odds ratios of the first generation groups are not significantly different from 1. However, as seen in 5.1, the majority of disability status change occurs among *Hearing*, *Memory* and *Physical* categories, and the coefficients of the first generation group are significantly less than 1. Therefore, it is reasonable to interpret that the results in these major categories shape the result concerning *Diffany* variable as well. For the second generation group, *Memory* and *Physical* category show significantly lower coefficient relevant to the reference group, while the odds ratios of the 2.5 generation group are significantly less than 1 in *Physical*, *Mobility*, and *Selfcare* category. In summary, compared with the results in Table 5.7, the disability risk pattern by the independent variable is more heterogeneous in Model 2 analysis, which indicates the importance of investigating each category, along with focusing on the aggregated variable as well. The full regression results of the six disability categories are presented in Appendix E.

Table 5.9 Summary of multiple logistic regression results of Model 2, with subcategories

	(1)	(2)	(3)	(4)	(5)	(6)
	Hearing	Visual	Memory	Physical	Mobility	Selfcare
2.5 generation	0.944	0.861	1.017	0.809***	0.827*	0.760*
	(0.088)	(0.133)	(0.099)	(0.060)	(0.088)	(0.121)
Second generation	1.034	0.911	0.679***	0.853**	0.879	0.912
	(0.101)	(0.140)	(0.076)	(0.067)	(0.090)	(0.135)
First generation	0.733***	0.933	0.624***	0.763***	0.964	1.072
	(0.060)	(0.094)	(0.049)	(0.040)	(0.072)	(0.114)

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Finally, additional logistic regression on each disability variables only with the first generation controlling for years since migration variable is performed to test the first RH regarding the HIE adequately. Table 5.10 presents the central part of the results. The number of observations based on Sample 2 is reduced to 43,886 as it only includes the first generation immigrant with available years since migration information. The full results are presented in Appendix F.

The coefficients of years since migration variable generally indicate that the health advantage of the first generation immigrants diminishes as they spend more time in the US. Compared with those who spent less than 10 years, those who spent 10-19 years, 20-29 years and 30 + years show higher odds ratio, although only 20-29 and 30+ group's odds ratios are statistically significant. Regarding 6 sub-categories, the evidence is mixed when comparing the reference group and 10-19 years group. In contrast, among those who spent more than 20 years, odds ratios which are significantly higher than 1 is observed in most cases.

Table 5.10 The summary of the multiple regression results of Model 2, including YSM

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Diffany	Diffany	Visual	Memory	Physical	Mobility	Selfcare
YSM 10-19	1.073	1.189	0.700*	1.278*	1.111	0.972	0.883
	(0.088)	(0.207)	(0.142)	(0.172)	(0.121)	(0.120)	(0.138)
YSM 20-29	1.221***	1.040	0.902	1.390***	1.519***	1.159	1.060
	(0.090)	(0.167)	(0.148)	(0.173)	(0.138)	(0.123)	(0.140)
YSM 30+	1.348***	1.539***	1.112	1.287**	1.386***	1.023	0.801*
	(0.092)	(0.217)	(0.167)	(0.155)	(0.117)	(0.101)	(0.099)
Constant	0.008***	0.003***	0.001***	0.008***	0.000***	0.002***	0.003***
	(0.004)	(0.003)	(0.001)	(0.006)	(0.000)	(0.002)	(0.003)
Year FE	YES	YES	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES	YES	YES
Pseudo R-squared	0.229	0.228	0.144	0.163	0.243	0.256	0.221

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

5.4 Sensitivity Testing

5.4.1 Regression using Different Age Thresholds

Appendix G presents the results of the multiple logistic regressions of *Diffany* variable employed in Model 2 by different age categories, i.e. over 30, 45, and 65. Firstly, the coefficient of the first generation group is significantly less than one in the models with age 30 and 45 thresholds, but it loses its significance and approximates to 1 with the age group over 65. Secondly, the second generation variable's odds ratios also approximate to one and statistically insignificant. Meanwhile, the odds ratios of the 2.5 generation become significantly less than 1 throughout the whole results in Appendix G. Thirdly, each category of *race* variables shows a similar pattern seen in Table 5.7, except that it becomes generally irrelevant amongst respondents over 65. Likewise, the coefficient of *Sex* variable remains significant among over 30 and over 45 groups, although it also becomes insignificant with over 65 group. Fourthly, the association between *Education level*, *Household income*, *Marital status*, and *Veteran status* and the dependent variable holds regardless of the age groups. One peculiar finding is that the coefficient of *age* variable is either insignificant amongst over 30 groups in Column (1). However, when the same regressions are conducted without the *Age-squared* variable, the coefficient becomes significant again for all regressions (1.052, 1.503, and 1.068 respectably).

5.4.2 Comparing Female and Male groups

Appendix H shows the results of the multiple logistic regressions on *Diffany* variable with total, female, and male respondents respectably. Regarding immigrant generation status, there are minor differences between male and female groups compared with each other or the total sample. The odds ratios of the second and first generation male are smaller than those of females, and especially the odds ratio of the second generation among female is not different from 1. If the argument that foreign-born female immigrant could be less self-selected (Markides et al., 2007) holds for most immigrant groups, it may partly explain this finding. Meanwhile, for the 2.5 generations, only female group shows significance result, but the odds ratios of the male and female group are almost identical.

Also, regarding *race* variable, African American and Indigenous group shows heterogeneous results by biological sex. African American male's odds ratio is not significantly different from the non-Hispanic White male, conditioning on all other variables included in the model. In contrast, African American female still experiences significantly higher risk relevant to the reference group. The indigenous group shows the opposite pattern, where male groups have a significantly higher risk, and the female group's odds ratio is higher than one, although insignificant. It should be considered that the significance of the results is likely to be affected by the subsample size as Indigenous people only accounts for 1% of the total sample.

For other demographic and socioeconomic variables, no considerable differences between male and female are observed in the results. Therefore, there does not exist evidence supporting that the associations between these variables and disability risk are heterogeneous between sex. To summarise, there is little difference observed between regressions on *Diffany*. The testing results of each disability category is illustrated in Appendix I.

6 Discussion

6.1 The Healthy Immigrant Effect and Disability Risk

The first research hypothesis predicts that the HIE will hold regarding the disability risk of the first generation immigrant. According to the results from Model 1, there exists evidence supporting a significant association between being the first generation immigrant and having a lower probability of having a disability. However, the series of regression results from Model 1 is not sufficient to make any causal interpretation. The model's design is particularly vulnerable to the reverse causality issue since the independent variable, covariates and dependent variable are observed at the same time, and multiple logistic regression cannot rule out the reverse causality issue in this setting.

In contrast, the results from Model 2 attempt to circumvent this issue by using lagged independent and covariates and regress the dependent variable on those lagged variables. Although it tracks each respondent for a short period, it removes any possibility of the outcome variable's affecting the explanatory variable and other covariates. According to the main findings from Table 5.7, the first generation immigrants experience a lower probability of having a disability within a year. Also, further analysis of the first generation group and considering the YSM variable (Table 5.10) implies that the health advantage diminishes as they spend more time in the destination country. Therefore, both H 1-1 and H 1-2 are supported, albeit there exist minor variations in results from different categories of disability status.

6.2 The intergenerational transmission of Disability Risk

The second research hypothesis consists of three sub-statements: firstly, H 2-1 predicts that the second generation shows a higher risk of having a disability compared with the first generation group. Also, H 2-2 expects that the 2.5 generation shows a lower risk of disability compared with the second generation group. Lastly, H 2-3 predicts the interaction between race and immigrant generation status is statistically significant. The main findings from Model 1 and 2 indicate that these hypotheses are only partially supported. The second generation's odds ratio is significantly lower than 1 in Model 1, which means they have a lower risk compared to their native counterparts, but their odds ratios are higher than the first generation in most cases. Meanwhile, when demographic and socioeconomic variables are controlled, the 2.5 generation does not differ from the reference group concerning disability risk. Besides, the results from Model 2 suggest that the 2.5 and second generation are both negatively associated with disability risk, and there exist minimal differences between the two groups. However, when the age threshold of over 30 is introduced, a notable change occurs; the 2.5. generation shows significantly lower disability risk while the second generation's risk differences are statistically insignificant. Since the evidence from Model 1 may be inconsistent due to methodological threats, the findings from regressions based on Model 2 may be more convincing between the two.

One possible explanation to this change is that when limiting the age from 15 to 30, the most attrition happens with *Memory*, where the 2.5 shows the highest odds ratio and the second generation shows the lowest one. Indeed, based on the outcome of Table 5.9, without *Memory*, the total sampling, including age over 15, may also show similar results to Appendix G. Another implication of the age limitation is that the disability risk differences amongst older respondent are more closely associated with chronic health conditions and underlying fundamental socioeconomic causes, rather than inherited or genetic risks. Consequently, the respondent's socioeconomic background, education and labour market outcomes may play a more vital role in determining disability. Due to insufficient observation period and lack of information regarding the individual's acculturation, making a decisive argument on this point is beyond the main implication of this research and remains as one of the limitations.

Another part of the second hypothesis is about the interaction between race and immigrant generation status—the summary results in Table 5.5 and 5.8 present different patterns of disability risk differences across racial groups. Concerning Table 5.8, it suggests one implication that the main results of Model 2-3 presented in Table 5.7 might be the result of the advantage of Non-Hispanic White and Asian group 2.5 generations is cancelled out by the relative disadvantage of the African American and Hispanic White counterparts. However, the statistical significance of the interaction terms is insufficient to support H 2-3. Also, extreme results found in some racial groups are probably affected by the sample size. Due to the sample restriction of Model 2, the observations from Indigenous, Pacific Islander and Other groups are minimal. Besides, the coefficients of the indigenous group need caution to interpret because they account for a tiny portion of the whole sample, and the implication of the result is ambiguous, for they reported their race is indigenous and experienced migration at the same time. To summarise, H 2-1 is supported, H 2-2 is partially supported, and H 2-3 is not supported by the results presented in Section 5.

6.3 The Fundamental Causes Theory and Disability

The final research hypothesis is concerning the relationship between a couple of socioeconomic variables and disability risk, i.e. marital status, household income, and education level. Including these variables has two crucial implications: firstly, without controlling those variables, the estimates of the independent variable probably suffer from the omitted variable bias; secondly, investigating the potential effect of those variable is relevant to examine whether the theory of fundamental causes applies to short-term disability trajectories amongst individuals with foreign background in the US. The main findings from Model 1 and 2 show apparent evidence that advantageous socioeconomic status captured by education and household income may have a protective effect regarding disability risks.

These results are consistent regardless of the different settings of the regression model presented in the main results and the sensitivity testing sections. It is also following the main implication of the previous studies regarding immigrant health or disability issue, e.g. Cho and Hummer (2001); Elo, Mehta and Huang (2011); Minkler, Fuller-Thomson and Guralnik (2006).

Furthermore, according to the results, being married may also provide a significant protective function regarding disability. Identifying the causal relationship between marriage or marital status and health is typically prone to reverse causality because health advantage could result from the positive effect of marriage or a healthier individual are more likely to get married if partner's health increases the expected utility of marriage. Again, the research design of Model 2 can solve this issue since the disability risk is regressed on last year's marital status. Besides, according to the sensitivity testing results with male and female-only group, the gender differences in the coefficient of marital status is not remarkably large. Consequently, this result may provide evidence that the protective effects of marriage are universal.

In summary, education level, household income, and marital status turn out to be negatively associated with short-term disability risk among the respondents. Therefore, H 3-1, H 3-2, and H 3-3 are supported by the main results. What remains regarding this hypothesis is that it does not track the long-term effects of those socioeconomic conditions, which will be addressed as another limitation of this research.

6.4 Internal and External Validity of the Results

6.4.1 Internal Validity

There are a couple of issues regarding the internal validity of the results. Firstly, the dependant variable, disability status, is self-measured. Previous studies question the reliability of self-reported health measure in migration studies in several ways: A specific group of the immigrant population, such as Asian Americans may more negatively perceive and report their health status (Huh, Prause & Dooley, 2008); men may underreport their health conditions while female overreport their health issues (Merrill, Kasl, & Berkman, 1997); and compared with objective criteria, self-reported health is more vulnerable to measurement error (Mutchler, Prakash & Burr, 2007). Taking these critiques into consideration, however, it can be argued that the female group and Asian group's significantly lower odds ratio might be the upper bound of the true ones. In other words, the main conclusion from the analyses may not be attenuated regarding *sex* and *race* variables. On the other hand, measurement errors regarding self-reporting remain, regardless of the characteristics of respondents. Without feasible objective disability measure or the possibility of introducing an instrument variable from IPUMS CPS, this study shares this limitation with other existing literature.

Secondly, one may question whether education level and household income, two key socioeconomic status variables, measures intended SES or degree of acculturation. This question is relevant for these variables could be proxies for acculturation in specific contexts.

For the first generation, years since migration or citizenship is a commonly used measure of acculturation. However, this study aims to investigate the second and 2.5 generations as well, which means standard acculturation measures are not feasible. Moreover, another common way to capture acculturation, spoken language at home, was not available for this study as well. Against this backdrop, if one adheres on linear or one-way of the acculturation process, i.e. ones pursuing the mainstream of the hosting country, having at least average level of education and income may be interpreted as outcomes of the successful acculturation process. However, if one follows the implications of segmented assimilation theory, mere education or income level does not represent one's acculturation. Instead, which types of cultural and social norms an individual follow and how one perceives the destination and origin country's culture gains relevance. This thesis does not include variables to measure the diverse acculturation, mostly focusing on how the second generation develops their social networks and identity as Haller, Portes and Fernández-Kelly (2009), Porte, Aparicio, Haller & Vickstrom, (2010) or Portes and Zhou (1993) do. Instead, it relies on the more rough way of examining intergenerational acculturation, following what Allen et al. (2007) suggest. Therefore, this thesis interprets the association between education or household income and disability risk as evidence of the FCT, and different immigrant generation status, the primary explanatory variable captures possible acculturation impact, albeit it may be insufficient to make a decisive conclusion.

6.4.2 External Validity

Regarding the external validity of the results, two issues could be addressed: Does the implication hold in other industrialised countries? Furthermore, does it hold for longer observation period? The answer to the first question probably depends on other pivotal factors influencing individual's health and immigrant integration, e.g. the public healthcare system, labour market institution, and most crucially, how immigrants self-select upon migration. Phelan, Link and Tehranifar (2010) also suggest that the effect of social conditions can be mitigated by interventions changing the individual's dependency on essential resources in utilising education and healthcare system. However, as the evidence of HIE observed in other receiving countries, and the implications of the FCT is applicable in most societies, the main findings can be replicated in other contexts as well.

The second question is more challenging to answer. Although evidence given in short-term observation may signal the long-term effect, it needs caution to link it to long-term results directly. Therefore, further research on immigrant's disability risk should aim for making use of longitudinal microdata with extended observation period and adequate analytic strategies to address this issue. Nonetheless, the research designed applied to Model 2 moves one step further to the longitudinal design compared with cross-sectional survey-based regression methods used by most previous literature on this topic.

7 Conclusion

7.1 Research Summary

This thesis attempted to investigate the disability risk disparities among the first and the second generation immigrants in the US, by using datasets extracted from the IPUMS CPS Basic Monthly Survey. This paper constructed two samples based on BMS: the first sample to see the overall association between the immigrant generation status as the independent variable and the disability status as the dependent variable, controlling a vector of demographic and socioeconomic covariates; the second sample to further the analysis by including respondents having no disability condition in Year 1, and using lagged independent variables and covariates, in order to examine the causal relationship between variables. Furthermore, a couple of additional robustness tests were conducted.

As discussed in the previous section, the results provide evidence corroborating the HIE among the first generation immigrants included in the samples. Also, covariates to measure SES such as education, household income, and marital status turned out to be significantly associated with having a disability condition, which indicates that the main argument of the FCT applies to immigrant's disability risk differences as well. In contrast, findings regarding the second and the 2.5 generation immigrant did not suggest conclusive evidence. On the one hand, the descendants of the first generation immigrant lose the former generation's health advantage to a degree in terms of disability risk. On the other hand, whether the second and the 2.5 generation group face dissimilar risk may depend on age groups, sex, or disability categories. It may imply that two groups are indeed exposed to a similar level of the health risk conditioning on their SES. However, it needs caution to draw such a conclusion because this thesis did not include enough variables to measure the acculturation process in detail. Moreover, at least for a short-term disability trajectory, the interaction between the race and immigrant's generation status was not statistically significant overall. Lastly, the results of sensitivity testing suggest that the main results are robust.

7.2 Research Implications

This thesis has three significant research implications. The first contribution to existing knowledge on immigrant health is to suggest that the established pattern of social gradient and differences by nativity/immigrant generation status in health can be shown by using disability as a measure of health. This thesis identified disability status as the primary outcome. However, disability status, in turn, can be an important explanatory factor for other aspects of behaviour and social cost e.g. extended period of sick leave, early retirements, disability pension, and additional healthcare expenditure. Therefore, a future study focusing on these phenomena should take the mechanism that immigrant's self-selection, integration, and other well-discussed fundamental conditions affect an individual's decision concomitant physical/mental disability into account. Besides, although the results do not provide evidence that individuals with foreign-background are associated with a higher risk of disability, it does not necessarily mean that their risks are lower in reality as well. If they are associated with insufficient financial resource or lower education level, the observed risks can be elevated. The main reason for elaborating on the results based on the FCT is to emphasise this point.

Secondly, this thesis is one of the initial attempts to investigate the relationship between disability risk and immigrant generation status involving the second and 2.5 generation immigrants. It is worth mentioning that according to the results, differences in disability risks even appear at a relatively earlier stage of life. Furthermore, as the second and 2.5 generation from the post-1965 immigrants get old, policymakers and researchers may pay more attention to second generation immigrant health issues. The main implication of this paper may serve as a point of departure for future discussion regarding the topic.

The final contribution of this thesis is that it presented one way to make use of the longitudinal design of IPUMS CPS data in migration studies and recognised some technical issues. As Rivera Drew, Flood and Warren (2014) state, the recent revision on the dataset motivates massive amount of potential research. With limited access to longitudinal microdata, including relevant information to answer the proposed research questions, this dataset provided an excellent alternative to materialise a research design to circumvent common drawbacks of a cross-sectional survey. However, throughout the data collection, this thesis recognises a couple of compatibility issues amongst variables and mismatched personal identifier problem. Even though IPUMS CPS explicitly warns users of the possible errors, researchers aiming for similar research design to this paper should mind this issue. Also, the CPS team could improve the perceived problems during the upcoming revision.

7.3 Limitations and Future Research

Notwithstanding its contribution, this thesis is not without limitations. Firstly, this thesis used self-reported disability status variables to measure disability risk. The overall reliability and explicit definition of each disability category probably prevent a severe measurement error. However, alternative measurement based on expert's diagnosis or a measure to quantify the burden of disability or mortality such as Disability-Adjusted Life Year (DALY) may solve this issue in future research. Therefore, it is also crucial to find a source which includes the objective measurement of disability.

Secondly, this study follows each respondent for a year, which is not sufficient to claim the long-term association between the independent/control variables and the dependent variable. Research on disability trajectory would benefit from an extended period of tracking because disability risk is linked to accumulated and multiple chronic health conditions. Thus, future research may aim to make use of panel dataset related to an individual's health or extract one from the administrative data. Provided that these alternatives are available, different methods, such as survival analysis, can be employed to estimate the differences in long-term risks.

Finally, the source of the observed disability risk differences between the first and second generation immigrant is not explicitly identified by this research. The evidence is insufficient to conclude whether acculturation or regression towards mean plays a more pivotal role. This ambiguity of the role of acculturation amongst the second or further generation immigrants may be solved by considering variables measuring the quantity and quality of the social network, individual's specific health risk-relevant behaviours or other measures of assimilation to the destination country. Another suggestion is to employ data which can match the first generation and their children, which allow researchers to control for parent's health status. Using data extracted from the administrative data or panel with household- and individual-level observations can be a desirable alternative.

References

Abraido-Lanza, A. F., Dohrenwend, B. P., Ng-Mak, D. S. & Turner, J. B. (1999). The Latino Mortality Paradox: A Test of the" Salmon Bias" and Healthy Migrant Hypotheses., *American journal of public health*, vol. 89, no. 10, pp.1543–1548, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Afable-Munsuz, A., Ponce, N. A., Rodriguez, M. & Perez-Stable, E. J. (2010). Immigrant Generation and Physical Activity among Mexican, Chinese & Filipino Adults in the U.S., *Social Science & Medicine*, [e-journal] vol. 70, no. 12, pp.1997–2005, Available Online: http://www.sciencedirect.com/science/article/pii/S0277953610001887 [Accessed 26 May 2020]

Akresh, I. R. & Frank, R. (2008). Health Selection among New Immigrants, *American journal of public health*, vol. 98, no. 11, pp.2058–2064, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Allen, M. L., Elliott, M. N., Morales, L. S., Diamant, A. L., Hambarsoomian, K. & Schuster, M. A. (2007). Adolescent Participation in Preventive Health Behaviors, Physical Activity, and Nutrition: Differences Across Immigrant Generations for Asians and Latinos Compared With Whites, *American Journal of Public Health*, [e-journal] vol. 97, no. 2, pp.337–343, Available Online: https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2005.076810 [Accessed 24 March 2020]

Angrist, J. D. & Pischke, J.-S. (2009). Mostly Harmless Econometrics: An Empiricist's Companion, Princeton: Princeton University Press.

Bengtsson, T. & Dribe, M. (2011). The Late Emergence of Socioeconomic Mortality Differentials: A Micro-Level Study of Adult Mortality in Southern Sweden 1815-1968, *Explorations in Economic History*, vol. 48, no. 3, pp.389–400, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Bengtsson, T. & van Poppel, F. (2011). Socioeconomic Inequalities in Death from Past to Present: An Introduction, *Explorations in Economic History*, [e-journal] vol. 48, no. 3, pp.343–356, Available Online: https://linkinghub.elsevier.com/retrieve/pii/S0014498311000180 [Accessed 28 October 2019]

Berry, J. W. (2003). Conceptual Approaches to Acculturation, in K. M. Chun, P. Balls Organista, & G. Marín (eds), *Acculturation: Advances in Theory, Measurement, and Applied Research.*, [e-book] Washington, DC: American Psychological Association, pp.17–37, Available Online: http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2002-18425-004&site=eds-live&scope=site [Accessed 7 April 2020]

Bickenbach, J. E., Chatterji, S., Badley, E. M. & Üstün, T. B. (1999). Models of Disablement, Universalism and the International Classification of Impairments, Disabilities and Handicaps, *Social Science & Medicine*, [e-journal] vol. 48, no. 9, pp.1173–1187, Available Online: http://www.sciencedirect.com/science/article/pii/S0277953698004419 [Accessed 25 March 2020]

Chetty, R., Abraham, S., Lin, S., Scuderi, B., Bergeron, A., Cutler, D. & Turner, N. (2016). The Association between Income and Life Expectancy in the United States, 2001-

2014, *JAMA - Journal of the American Medical Association*, vol. 315, no. 16, pp.1750–1766, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Cho, Y. & Hummer, R. A. (2001). Disability Status Differentials across Fifteen Asian and Pacific Islander Groups and the Effect of Nativity and Duration of Residence in the U.S., *Social Biology*, [e-journal] vol. 48, no. 3–4, pp.171–195, Available Online: https://doi.org/10.1080/19485565.2001.9989034 [Accessed 24 March 2020]

De Souza, L. R. & Fuller-Thomson, E. (2013). Acculturation and Disability Rates among Filipino-Americans, *Journal of immigrant and minority health*, vol. 15, no. 3, pp.462–471, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Elo, I. T., Mehta, N. K. & Huang, C. (2011). Disability Among Native-Born and Foreign-Born Blacks in the United States, *Demography*, [e-journal] vol. 48, no. 1, pp.241–265, Available Online: https://doi.org/10.1007/s13524-010-0008-x [Accessed 24 February 2020]

Feng, S. (2001). The Longitudinal Matching of Current Population Surveys: A Proposed Algorithm, *Journal of Economic and Social Measurement*, vol. 27, no. 1–2, pp.71–91, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Flood, S., King, M., Rodgers, R., Ruggles, S. &Warren, J.R. (2018). Integrated Public Use Microdata Series, Current Population Survey: Version 6.0 [dataset]. Minneapolis, MN: IPUMS, Available through: https://doi.org/10.18128/D030.V6.0

Franzini, L. & Fernandez-Esquer, M. E. (2004). Socioeconomic, Cultural, and Personal Influences on Health Outcomes in Low Income Mexican-Origin Individuals in Texas, *Social science & medicine*, vol. 59, no. 8, pp.1629–1646, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Frisbie, P. W., Cho, Y. & Hummer, R. A. (2001). Immigration and the Health of Asian and Pacific Islander Adults in the United States, *American journal of epidemiology*, vol. 153, no. 4, pp.372–380, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Fuller-Thomson, E., Brennenstuhl, S. & Hurd, M. (2011). Comparison of Disability Rates among Older Adults in Aggregated and Separate Asian American/Pacific Islander Subpopulations, *American Journal of Public Health*, vol. 101, no. 1, pp.94–100, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Galton, F. (1886). Regression Towards Mediocrity in Hereditary Stature, *The Journal of the Anthropological Institute of Great Britain and Ireland*, [e-journal] vol. 15, p.246, Available Through: edsjsr http://ludwig.lub.lu.se/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsjsr&AN=edsjsr.10.2307.2841583&site=eds-live&scope=site

Gans, H. J. (1992). Second-generation Decline: Scenarios for the Economic and Ethnic Futures of the Post-1965 American Immigrants, *Ethnic and racial studies*, vol. 15, no. 2, pp.173–192, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Gee, E. M. T., Kobayashi, K. M. & Prus, S. G. (2004). Examining the Healthy Immigrant Effect in Mid- To Later Life: Findings from the Canadian Community Health Survey, *Canadian Journal on Aging / La Revue canadienne du vieillissement*, [e-journal] vol. 23, no. 5, pp.S55–S63, Available Online: https://muse.jhu.edu/article/182895 [Accessed 24 March 2020].

Gee, G. C., Ro, A., Shariff-Marco, S. & Chae, D. (2009). Racial Discrimination and Health among Asian Americans: Evidence, Assessment, and Directions for Future Research, *Epidemiologic reviews*, vol. 31, no. 1, pp.130–151, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Greene, W. H. (2012). Econometric Analysis, 7. ed., Boston ; London: Pearson.

Haller, W., Portes, A. & Fernández-Kelly, P. (2009). The Adaptation of the Immigrant Second Generation in America: A Theoretical Overview and Recent Evidence, *Journal of Ethnic and Migration Studies*, [e-journal] vol. 35, no. 7, pp.1077–1104, Available Online: http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edselc&AN=edselc.2-52.0-70449502808&site=eds-live&scope=site [Accessed 24 March 2020]

Huh, J., Prause, J. A. & Dooley, C. D. (2008). The Impact of Nativity on Chronic Diseases, Self-Rated Health and Comorbidity Status of Asian and Hispanic Immigrants, *Journal of Immigrant and Minority Health*, [e-journal] vol. 10, no. 2, pp.103–118, Available Online: https://doi.org/10.1007/s10903-007-9065-7 [Accessed 23 April 2020]

Hummer, R. A., Rogers, R. G., Nam, C. B. & LeClere, F. B. (1999). Race/Ethnicity, Nativity, and US Adult Mortality, *Social Science Quarterly*, pp.136–153, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

IPUMS CPS. (2020). Available Online: https://cps.ipums.org/cps/intro.shtml [Accessed 30 April 2020]

Jass, G. & Massey, D. S. (2004). Immigrant Health: Selectivity and Acculturation, Working Paper, 04/23, IFS Working Papers, Available Online: https://www.econstor.eu/handle/10419/71469 [Accessed 24 March 2020]

Johansson, B., Helgesson, M., Lundberg, I., Nordquist, T., Leijon, O., Lindberg, P. & Vingård, E. (2012). Work and Health among Immigrants and Native Swedes 1990–2008: A Register-Based Study on Hospitalization for Common Potentially Work-Related Disorders, Disability Pension and Mortality, *BMC Public Health*, vol. 12, no. 1, p.845. Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Kennedy, S., Kidd, M. P., McDonald, J. T. & Biddle, N. (2015). The Healthy Immigrant Effect: Patterns and Evidence from Four Countries, *Journal of International Migration and Integration*, [e-journal] vol. 16, no. 2, pp.317–332, Available Online: https://doi.org/10.1007/s12134-014-0340-x [Accessed 24 March 2020]

Kenny, G. P., Yardley, J. E., Martineau, L. & Jay, O. (2008). Physical Work Capacity in Older Adults: Implications for the Aging Worker, *American journal of industrial medicine*, vol. 51, no. 8, pp.610–625, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Levels, M., Dronkers, J. & Kraaykamp, G. (2008). Immigrant Children’s Educational Achievement in Western Countries: Origin, Destination, and Community Effects on Mathematical Performance, *American Sociological Review*, [e-journal] vol. 73, no. 5, pp.835–853, Available Online: https://doi.org/10.1177/000312240807300507 [Accessed 27 May 2020]

Liddell, B. J., Nickerson, A., Sartor, L., Ivancic, L. & Bryant, R. A. (2016). The Generational Gap: Mental Disorder Prevalence and Disability amongst First and Second Generation Immigrants in Australia, *Journal of psychiatric research*, vol. 83, pp.103–111, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Link, B. & Phelan, J. (1995). Social Conditions as Fundamental Causes of Disease, *JOURNAL OF HEALTH AND SOCIAL BEHAVIOR*, vol. 35, pp.80–94, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Mackenbach, J. p., Kunst, A. e., Cavelaars, A. e. j. m., Groenhof, F. & Geurts, J. j. m. (1997). Socioeconomic Inequalities in Morbidity and Mortality in Western Europe, *Lancet*, [e-journal] vol. 349, no. 9066, pp.1655–1659, Available Online: http://ludwig.lub.lu.se/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edselc&AN=edselc.2-52.0-0030911641&site=eds-live&scope=site [Accessed 3 October 2019].\

Madrian, B. C. & Lefgren, L. J. (2000). An Approach to Longitudinally Matching Current Population Survey (CPS) Respondents, *Journal of Economic and Social Measurement*, vol. 26, no. 1, pp.31–62, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Markides, K. S., Eschbach, K., Ray, L. A. & Peek, M. K. (2007). Census Disability Rates among Older People by Race/Ethnicity and Type of Hispanic Origin, in *The Health of Aging Hispanics*, Springer, pp.26–39, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Marmot, M. (2005). Social Determinants of Health Inequalities, *The Lancet*, [e-journal] vol. 365, no. 9464, pp.1099–1104, Available Online: http://www.sciencedirect.com/science/article/pii/S0140673605711466 [Accessed 3 November 2019]

Marmot, M. G. & Mcdowall, M. E. (1986). Mortality Decline and Widening Social Inequalities, *The Lancet*, vol. 328, no. 8501, pp.274–276, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Mateus, S. (2019). Advantage in Diversity: 9th Grade Pupils of Mixed Origin in Portugal, *Portuguese Journal of Social Science*, [e-journal] vol. 18, no. 1, pp.57–72, Available Online: http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=137236563&site=eds-live&scope=site [Accessed 16 March 2020]

Merrill, S. S., Seeman, T. E., Kasl, S. V. & Berkman, L. F. (1997). Gender Differences in the Comparison of Self-Reported Disability and Performance Measures, *The Journals of Gerontology: Series A*, [e-journal] vol. 52A, no. 1, pp.M19–M26, Available Online: https://academic.oup.com/biomedgerontology/article/52A/1/M19/550219 [Accessed 21 March 2020]

Minkler, M., Fuller-Thomson, E. & Guralnik, J. M. (2006). Gradient of Disability across the Socioeconomic Spectrum in the United States, *New England Journal of Medicine*, vol. 355, no. 7, pp.695–703, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Molitoris, J. & Dribe, M. (2016). Industrialization and Inequality Revisited: Mortality Differentials and Vulnerability to Economic Stress in Stockholm, 1878-1926, *European Review of Economic History*, vol. 20, no. 2, pp.176–197, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data, *Econometrica: journal of the Econometric Society*, pp.69–85, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Mutchler, J. E., Prakash, A. & Burr, J. A. (2007). The Demography of Disability and the Effects of Immigrant History: Older Asians in the United States, *Demography*, [e-journal] vol. 44, no. 2, pp.251–263, Available Online: https://doi.org/10.1353/dem.2007.0015 [Accessed 24 February 2020]

NBER. (2020). CPS Merged Outgoing Rotation Groups, *The National Bureau of Economic Research*, Available Online: https://data.nber.org/data/morg.html [Accessed 4 May 2020]

Park, J. (1) & Myers, D. (2). (2010). Intergenerational Mobility in the Post-1965 Immigration Era: Estimates by an Immigrant Generation Cohort Method, *Demography*, [e-journal] vol. 47, no. 2, pp.369–392, Available Online: http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edselc&AN=edselc.2-52.0-77952926561&site=eds-live&scope=site [Accessed 24 November 2019]

Phelan, J. C., Link, B. G. & Tehranifar, P. (2010). Social Conditions as Fundamental Causes of Health Inequalities: Theory, Evidence, and Policy Implications, *Journal of Health and Social Behavior*, [e-journal] vol. 51, no. 1_suppl, pp.S28–S40, Available Online: https://doi.org/10.1177/0022146510383498 [Accessed 24 March 2020]

Portes, A., Aparicio, R., Haller, W. & Vickstrom, E. (2010). Moving Ahead in Madrid: Aspirations and Expectations in the Spanish Second Generation, *International Migration Review*, vol. 44, no. 4, pp.767–801, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Portes, A. & Zhou, M. (1993). The New Second Generation: Segmented Assimilation and Its Variants, *The annals of the American academy of political and social science*, vol. 530, no. 1, pp.74–96, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Ramakrishnan, S. K. (2004). Second-Generation Immigrants? The “2.5 Generation” in the United States*, *Social Science Quarterly*, [e-journal] vol. 85, no. 2, pp.380–399, Available Online: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0038-4941.2004.08502013.x [Accessed 24 March 2020]

Riosmena, F., Kuhn, R. & Jochem, W. C. (2017). Explaining the Immigrant Health Advantage: Self-Selection and Protection in Health-Related Factors Among Five Major National-Origin Immigrant Groups in the United States, *Demography*, [e-journal] vol. 54, no. 1, pp.175–200, Available Online: https://doi.org/10.1007/s13524-016-0542-2 [Accessed 24 March 2020]

Rivera Drew, J. A., Flood, S. & Warren, J. R. (2014). Making Full Use of the Longitudinal Design of the Current Population Survey: Methods for Linking Records across 16 Months\m {1}, *Journal of economic and social measurement*, vol. 39, no. 3, pp.121–144, Available through: LUSEM Library website http://www.lusem.lu.se/library [Accessed 26 May 2020]

Rodrigues, M. A. P., Facchini, L. A., Thumé, E. & Maia, F. (2009). Gender and Incidence of Functional Disability in the Elderly: A Systematic Review, *Cadernos de Saúde Pública*, [e-journal] vol. 25, pp.S464–S476, Available Online: https://www.scielosp.org/article/csp/2009.v25suppl3/S464-S476/ [Accessed 24 March 2020]

Uretsky, M. C. & Mathiesen, S. G. (2007). The Effects of Years Lived in the United States on the General Health Status of California's Foreign-Born Populations, *Journal of Immigrant and Minority Health*, [e-journal] vol. 9, no. 2, pp.125–136, Available Online: https://doi.org/10.1007/s10903-006-9017-7 [Accessed 27 May 2020]

Verbrugge, L. M. & Jette, A. M. (1994). The Disablement Process, *Social Science & Medicine*, [e-journal] vol. 38, no. 1, pp.1–14, Available Online: https://linkinghub.elsevier.com/retrieve/pii/0277953694902941 [Accessed 25 March 2020]

Warner, D. F. & Brown, T. H. (2011). Understanding How Race/Ethnicity and Gender Define Age-Trajectories of Disability: An Intersectionality Approach, *Social Science & Medicine*, [e-journal] vol. 72, no. 8, pp.1236–1248, Available Online: http://www.sciencedirect.com/science/article/pii/S0277953611001353 [Accessed 21 March 2020]

WHO. (2020). Available Online: http://www.who.int/social_determinants/sdh_definition/en/ [Accessed 22 April 2020]

WHO. (2020). Available Online: http://www.who.int/classifications/icf/en/ [Accessed 25 March 2020]

Wooldridge, J. M. (2016). Introductory Econometrics: A Modern Approach, 6th edn., Boston: Cengage Learning.

Appendix A

Table A.1 Summary statistics of Sample 2

VARIABLES	(1) mean	(2) sd	(3) min	(4) max
Age	45.95	17.62	15	85
Age squared	2,422	1,671	225	7,225
Sex	0.523	0.499	0	1
State (FIPS code)	27.81	15.96	1	56
Metropolitan	0.808	0.394	0	1
Marital status	0.581	0.493	0	1
Veteran status	0.0815	0.274	0	1
2.5 generation	0.0367	0.188	0	1
Second generation	0.0423	0.201	0	1
First generation	0.148	0.355	0	1
High school	0.274	0.446	0	1
Some college	0.175	0.380	0	1
College degree	0.404	0.491	0	1
Hispanic White	0.111	0.314	0	1
African American	0.0919	0.289	0	1
Indigenous	0.00926	0.0958	0	1
Asian	0.0512	0.220	0	1
Pacific Islander	0.00375	0.0611	0	1
Other (race)	0.0148	0.121	0	1
$25,000 - 49,999	0.239	0.426	0	1
$50,000 - 74,999	0.208	0.406	0	1
$75,000 - 149,999	0.285	0.451	0	1
$150,000 or more	0.112	0.315	0	1
Year	2,013	2.572	2,009	2,017
Diffany	0.0512	0.220	0	1
Hearing	0.0135	0.115	0	1
Visual	0.00600	0.0772	0	1
Memory	0.0127	0.112	0	1
Physical	0.0275	0.164	0	1
Mobility	0.0123	0.110	0	1
Selfcare	0.00559	0.0745	0	1
Number of Observation: 297,460				

Notes: The dependent variables are measured in Year 2, and all the other variables are measured in Year 1.

Appendix B

Table B.1 The multiple regression results of Model 1 with the interaction term

VARIABLES	Model 1-4
2.5 generation	0.972
	(0.018)
Second generation	0.965
	(0.023)
First generation	0.748***
	(0.013)
Hispanic White	0.916***
	(0.018)
African American	0.948***
	(0.010)
Indigenous	1.438***
	(0.046)
Asian	0.684***
	(0.042)
Pacific Islander	0.863
	(0.081)
Other	1.802***
	(0.047)
2.5 generation * Hispanic White	1.142***
	(0.052)
2.5 generation * African American	1.126
	(0.099)
2.5 generation * Indigenous	0.944
	(0.157)
2.5 generation * Asian	1.243*
	(0.149)
2.5 generation * Pacific Islander	1.232
	(0.297)
2.5 generation * Other	0.821**
	(0.075)
Second generation * Hispanic White	0.767***
	(0.031)
Second generation * African American	0.841**
	(0.071)
Second generation * Indigenous	0.709*
	(0.131)
Second generation * Asian	0.849*
	(0.071)
Second generation * Pacific Islander	1.122

	(0.245)
Second * Other	0.443***
	(0.079)
First generation * Hispanic White	0.615***
	(0.018)
First generation * African American	0.690***
	(0.027)
First generation * Indigenous	0.506***
	(0.057)
First generation * Asian	0.974
	(0.065)
First generation * Pacific Islander	0.965
	(0.131)
First generation * Other	0.543***
	(0.049)
Age	1.040***
	(0.001)
Age squared	1.000***
	(0.000)
Sex	0.958***
	(0.006)
Marital status	0.604***
	(0.004)
Metropolitan	0.947***
	(0.008)
High school	0.647***
	(0.006)
Some college	0.596***
	(0.006)
College degree	0.450***
	(0.005)
$25,000 - 49,999	0.544***
	(0.004)
$50,000 - 74,999	0.415***
	(0.004)
$75,000 - 149,999	0.315***
	(0.003)
$150,000 and over	0.238***
	(0.004)
Veteran status	1.394***
	(0.014)
Constant	0.070***
	(0.002)
Year FE	YES
State FE	YES
Pseudo R-squared	0.194

Appendix C

Table C.1 The multiple regression results of each disability category, Model 1

VARIABLES	(1) diffhear	(2) diffeye	(3) diffrem	(4) diffphys	(5) diffmob	(6) diffcare
2.5 generation	0.966	0.951	0.950*	0.991	0.939**	0.931**
	(0.024)	(0.034)	(0.025)	(0.020)	(0.024)	(0.033)
Second generation	0.928***	0.901***	0.778***	0.907***	0.929***	0.937*
	(0.024)	(0.033)	(0.022)	(0.020)	(0.024)	(0.033)
First generation	0.631***	0.712***	0.548***	0.598***	0.718***	0.758***
	(0.014)	(0.020)	(0.012)	(0.009)	(0.014)	(0.020)
Hispanic white	0.778***	1.093***	0.727***	0.877***	0.856***	0.999
	(0.019)	(0.031)	(0.016)	(0.015)	(0.018)	(0.029)
African American	0.541***	1.157***	0.832***	1.089***	1.045***	1.150***
	(0.011)	(0.024)	(0.013)	(0.013)	(0.016)	(0.023)
Indigenous	1.388***	1.799***	1.177***	1.524***	1.519***	1.708***
	(0.071)	(0.102)	(0.050)	(0.055)	(0.066)	(0.100)
Asian	0.863***	0.888**	0.859***	0.830***	1.008	0.987
	(0.030)	(0.042)	(0.030)	(0.022)	(0.031)	(0.043)
Pacific Islander	1.056	1.205	0.623***	1.258***	1.099	1.282*
	(0.122)	(0.177)	(0.070)	(0.098)	(0.110)	(0.173)
Other	1.469***	1.767***	1.717***	1.689***	1.540***	1.525***
	(0.062)	(0.087)	(0.055)	(0.051)	(0.058)	(0.079)
Age	1.012***	1.045***	1.027***	1.124***	1.017***	1.051***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Age squared	1.000***	1.000	1.000***	0.999***	1.000***	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sex	0.602***	0.962**	0.941***	1.199***	1.148***	1.037**
	(0.007)	(0.015)	(0.010)	(0.010)	(0.013)	(0.016)
Marital status	0.834***	0.606***	0.458***	0.612***	0.494***	0.524***
	(0.009)	(0.009)	(0.005)	(0.005)	(0.005)	(0.008)
Metropolitan	0.834***	0.895***	1.051***	0.955***	1.086***	1.086***
	(0.010)	(0.015)	(0.013)	(0.009)	(0.013)	(0.018)
High school	0.748***	0.671***	0.629***	0.688***	0.668***	0.705***
	(0.011)	(0.013)	(0.008)	(0.008)	(0.009)	(0.013)
Some college	0.742***	0.658***	0.506***	0.679***	0.522***	0.611***
	(0.013)	(0.015)	(0.008)	(0.009)	(0.009)	(0.014)
College degree	0.598***	0.553***	0.374***	0.502***	0.405***	0.497***
	(0.010)	(0.012)	(0.006)	(0.006)	(0.006)	(0.011)
$25,000 - 49,999	0.765***	0.584***	0.518***	0.545***	0.562***	0.580***
	(0.010)	(0.010)	(0.006)	(0.005)	(0.007)	(0.009)
$50,000 - 74,999	0.669***	0.436***	0.394***	0.394***	0.451***	0.441***
	(0.011)	(0.010)	(0.006)	(0.005)	(0.007)	(0.010)
$75,000 - 149,999	0.557***	0.346***	0.295***	0.285***	0.349***	0.339***
	(0.010)	(0.009)	(0.005)	(0.004)	(0.006)	(0.009)
$150,000 and over	0.443***	0.240***	0.228***	0.201***	0.288***	0.259***
	(0.012)	(0.010)	(0.007)	(0.005)	(0.008)	(0.011)

Veteran status	1.553***	1.103***	1.300***	1.239***	0.988	1.032
	(0.023)	(0.025)	(0.022)	(0.016)	(0.017)	(0.023)
Constant	0.012***	0.006***	0.058***	0.003***	0.024***	0.004***
	(0.001)	(0.000)	(0.002)	(0.000)	(0.001)	(0.000)
Year FE	YES	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES	YES
Pseudo R-squared	0.201	0.122	0.103	0.219	0.168	0.145

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Appendix D

Table D.1 The multiple regression results of each disability category, Model 2

VARIABLES	Model 2-4
2.5 generation	0.871**
	(0.053)
Second generation	0.952
	(0.070)
First generation	0.806***
	(0.045)
Hispanic White	0.884*
	(0.064)
African American	1.151***
	(0.042)
Indigenous	1.479***
	(0.158)
Asian	0.766
	(0.164)
Pacific Islander	1.217
	(0.357)
Other	1.078
	(0.116)
2.5 generation * Hispanic White	1.150
	(0.192)
2.5 generation * African American	0.896
	(0.279)
2.5 generation * Indigenous	0.580
	(0.375)
2.5 generation * Asian	0.819
	(0.334)
2.5 generation * Pacific Islander	2.278
	(1.980)
2.5 generation * Other	1.270
	(0.425)
Second generation * Hispanic White	0.995
	(0.134)
Second generation * African American	0.512*
	(0.183)
Second generation * Indigenous	0.557
	(0.331)
Second generation * Asian	0.822
	(0.238)
Second generation * Pacific Islander	0.421

	(0.346)
Second * Other	0.298
	(0.304)
First generation * Hispanic White	0.843*
	(0.083)
First generation * African American	0.707***
	(0.086)
First generation * Indigenous	0.260**
	(0.137)
First generation * Asian	1.217
	(0.275)
First generation * Pacific Islander	0.567
	(0.255)
First generation * Other	1.036
	(0.323)
Age	1.027***
	(0.003)
Age squared	1.000***
	(0.000)
Sex	0.943***
	(0.021)
Marital status	0.719***
	(0.016)
Metropolitan	0.966
	(0.026)
High school	0.777***
	(0.024)
Some college	0.662***
	(0.024)
College degree	0.527***
	(0.018)
$25,000 - 49,999	0.612***
	(0.016)
$50,000 - 74,999	0.500***
	(0.016)
$75,000 - 149,999	0.392***
	(0.013)
$150,000 and over	0.297***
	(0.016)
Veteran status	1.187***
	(0.039)
Constant	0.025***
	(0.003)
Year FE	YES
State FE	YES
Pseudo R-squared	0.136

Appendix E

Table E.1 The multiple regression results of each disability category, Model 2

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	diffhear	diffeye	diffrem	diffphys	diffmob	diffcare
2.5 generation	0.944	0.861	1.017	0.809***	0.827*	0.760*
	(0.088)	(0.133)	(0.099)	(0.060)	(0.088)	(0.121)
Second generation	1.034	0.911	0.679***	0.853**	0.879	0.912
	(0.101)	(0.140)	(0.076)	(0.067)	(0.090)	(0.135)
First generation	0.733***	0.933	0.624***	0.763***	0.964	1.072
	(0.060)	(0.094)	(0.049)	(0.040)	(0.072)	(0.114)
Hispanic white	0.821**	1.087	0.912	0.902	0.935	1.049
	(0.078)	(0.124)	(0.080)	(0.057)	(0.083)	(0.132)
African American	0.600***	1.287***	1.086	1.286***	1.362***	1.341***
	(0.053)	(0.113)	(0.068)	(0.055)	(0.084)	(0.120)
Indigenous	1.354*	1.508*	1.224	1.139	1.755***	1.019
	(0.247)	(0.343)	(0.223)	(0.161)	(0.315)	(0.319)
Asian	0.819	0.898	1.030	1.065	1.101	1.025
	(0.108)	(0.150)	(0.129)	(0.091)	(0.125)	(0.176)
Pacific Islander	0.299**	1.829	1.739	0.926	1.210	1.336
	(0.160)	(0.688)	(0.590)	(0.297)	(0.484)	(0.609)
Other	0.962	0.861	1.051	1.116	1.529**	1.270
	(0.200)	(0.251)	(0.176)	(0.150)	(0.261)	(0.325)
Age	1.009	1.024**	1.006	1.094***	0.993	1.037***
	(0.008)	(0.009)	(0.006)	(0.006)	(0.006)	(0.010)
Age squared	1.001***	1.000*	1.000***	1.000***	1.001***	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sex	0.581***	1.001	0.863***	1.175***	1.138***	1.036
	(0.026)	(0.060)	(0.036)	(0.036)	(0.051)	(0.066)
Marital status	0.940	0.739***	0.530***	0.745***	0.589***	0.601***
	(0.040)	(0.045)	(0.023)	(0.021)	(0.025)	(0.037)
Metropolitan	0.876***	1.023	1.067	1.003	1.135**	1.038
	(0.044)	(0.075)	(0.056)	(0.036)	(0.063)	(0.082)

High school	0.876**	0.861*	0.753***	0.798***	0.772***	0.886
	(0.052)	(0.069)	(0.042)	(0.032)	(0.042)	(0.074)
Some college	0.761***	0.716***	0.631***	0.700***	0.586***	0.869
	(0.054)	(0.070)	(0.040)	(0.033)	(0.039)	(0.085)
College degree	0.665***	0.658***	0.455***	0.538***	0.474***	0.640***
	(0.044)	(0.060)	(0.030)	(0.024)	(0.031)	(0.060)
$25,000 - 49,999	0.836***	0.560***	0.547***	0.614***	0.622***	0.646***
	(0.043)	(0.039)	(0.026)	(0.021)	(0.030)	(0.046)
$50,000 - 74,999	0.756***	0.462***	0.444***	0.472***	0.495***	0.522***
	(0.046)	(0.041)	(0.026)	(0.020)	(0.031)	(0.047)
$75,000 - 149,999	0.644***	0.381***	0.310***	0.377***	0.404***	0.470***
	(0.042)	(0.036)	(0.021)	(0.017)	(0.028)	(0.047)
$150,000 and over	0.534***	0.290***	0.237***	0.262***	0.321***	0.276***
	(0.051)	(0.043)	(0.025)	(0.020)	(0.035)	(0.046)
Veteran status	1.320***	1.178*	1.195***	1.115**	0.955	1.178*
	(0.070)	(0.104)	(0.078)	(0.049)	(0.066)	(0.108)
Constant	0.002***	0.003***	0.023***	0.002***	0.011***	0.001***
	(0.001)	(0.001)	(0.004)	(0.000)	(0.002)	(0.000)
Year FE	YES	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES	YES
Pseudo R-squared	0.162	0.0808	0.0774	0.152	0.133	0.101

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Appendix F

Table F.1 The multiple regression results of each disability category, including YSM, Model 2

VARIABLES	(1) Diffany	(2) Diffany	(3) Visual	(4) Memory	(5) Physical	(6) Mobility	(7) Selfcare
Hispanic white	0.868**	0.713***	1.135	0.852	0.933	0.865	0.780**
	(0.053)	(0.079)	(0.150)	(0.093)	(0.069)	(0.078)	(0.089)
African American	0.768***	0.554***	0.808	0.705**	0.852	0.960	0.897
	(0.071)	(0.114)	(0.170)	(0.112)	(0.092)	(0.123)	(0.145)
Indigenous	0.669	0.175*	0.499	1.017	0.549	1.320	0.678
	(0.193)	(0.185)	(0.391)	(0.440)	(0.242)	(0.482)	(0.440)
Asian	0.877**	0.838	0.913	1.024	0.945	0.997	0.853
	(0.055)	(0.092)	(0.131)	(0.111)	(0.072)	(0.089)	(0.096)
Pacific Islander	1.121	0.861	1.494	0.631	1.340	1.053	0.776
	(0.267)	(0.373)	(0.623)	(0.274)	(0.397)	(0.392)	(0.444)
Other	1.170	1.554	1.129	1.426	1.041	0.926	0.249**
	(0.252)	(0.585)	(0.511)	(0.439)	(0.329)	(0.333)	(0.152)
YSM 10-19	1.073	1.189	0.700*	1.278*	1.111	0.972	0.883
	(0.088)	(0.207)	(0.142)	(0.172)	(0.121)	(0.120)	(0.138)
YSM 20-29	1.221***	1.040	0.902	1.390***	1.519***	1.159	1.060
	(0.090)	(0.167)	(0.148)	(0.173)	(0.138)	(0.123)	(0.140)
YSM 30+	1.348***	1.539***	1.112	1.287**	1.386***	1.023	0.801*
	(0.092)	(0.217)	(0.167)	(0.155)	(0.117)	(0.101)	(0.099)
Age	1.025***	0.985	1.059***	0.992	1.106***	1.006	1.022
	(0.009)	(0.018)	(0.023)	(0.013)	(0.014)	(0.012)	(0.016)
Age squared	1.000***	1.001***	1.000	1.001***	1.000**	1.001***	1.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sex	1.047	0.685***	1.162	1.018	1.200***	1.199***	1.005
	(0.046)	(0.056)	(0.115)	(0.079)	(0.065)	(0.079)	(0.085)
Marital status	0.564***	0.832**	0.718***	0.389***	0.570***	0.471***	0.528***
	(0.025)	(0.070)	(0.074)	(0.031)	(0.030)	(0.030)	(0.044)
Metropolitan	1.015	0.807	1.007	1.030	0.953	1.126	0.804
	(0.106)	(0.137)	(0.230)	(0.182)	(0.124)	(0.182)	(0.162)
High school	0.758***	0.636***	0.708***	0.723***	0.751***	0.730***	0.841*
	(0.043)	(0.067)	(0.087)	(0.069)	(0.050)	(0.057)	(0.085)
Some college	0.759***	0.777*	0.715*	0.597***	0.840*	0.632***	0.814
	(0.060)	(0.115)	(0.129)	(0.083)	(0.080)	(0.076)	(0.125)
College degree	0.597***	0.685***	0.587***	0.520***	0.571***	0.501***	0.615***
	(0.037)	(0.077)	(0.082)	(0.059)	(0.043)	(0.045)	(0.072)
$25,000 - 49,999	0.519***	0.770***	0.575***	0.590***	0.468***	0.534***	0.468***
	(0.027)	(0.077)	(0.069)	(0.054)	(0.030)	(0.041)	(0.046)
$50,000 - 74,999	0.479***	0.686***	0.514***	0.533***	0.456***	0.495***	0.392***

	(0.031)	(0.084)	(0.080)	(0.061)	(0.036)	(0.048)	(0.051)
$75,000 - 149,999	0.403***	0.617***	0.525***	0.439***	0.353***	0.454***	0.354***
	(0.028)	(0.078)	(0.084)	(0.056)	(0.031)	(0.046)	(0.049)
$150,000 and over	0.275***	0.419***	0.296***	0.345***	0.281***	0.375***	0.256***
	(0.030)	(0.085)	(0.080)	(0.068)	(0.039)	(0.060)	(0.057)
Veteran status	1.276**	1.444**	1.452	1.480*	1.090	0.927	0.752
	(0.155)	(0.251)	(0.368)	(0.301)	(0.166)	(0.190)	(0.190)
Constant	0.008***	0.003***	0.001***	0.008***	0.000***	0.002***	0.003***
	(0.004)	(0.003)	(0.001)	(0.006)	(0.000)	(0.002)	(0.003)
Year FE	YES	YES	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES	YES	YES
Pseudo R-squared	0.229	0.228	0.144	0.163	0.243	0.256	0.221

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Appendix G

Table G.1 The multiple regression results by different age category, Model 2

VARIABLES	(1) 30+	(2) 45+	(3) 65+
2.5 generation	0.870**	0.876**	0.860**
	(0.049)	(0.052)	(0.062)
Second generation	0.939	0.938	1.010
	(0.058)	(0.062)	(0.076)
First generation	0.735***	0.798***	0.983
	(0.031)	(0.035)	(0.057)
Hispanic white	0.856***	0.878**	0.965
	(0.043)	(0.047)	(0.072)
African American	1.115***	1.115***	1.093
	(0.040)	(0.043)	(0.061)
Indigenous	1.192*	1.243*	1.562**
	(0.126)	(0.147)	(0.273)
Asian	0.973	0.997	0.986
	(0.065)	(0.071)	(0.094)
Pacific Islander	1.172	1.186	0.750
	(0.258)	(0.288)	(0.280)
Other	1.158	1.246*	0.987
	(0.123)	(0.148)	(0.187)
Age	1.004	0.928***	0.814***
	(0.006)	(0.010)	(0.013)
Age squared	1.000***	1.001***	1.002***
	(0.000)	(0.000)	(0.000)
Sex	0.954**	0.959*	0.947
	(0.022)	(0.024)	(0.035)
Marital status	0.717***	0.746***	0.809***
	(0.016)	(0.018)	(0.026)
Metropolitan	0.955*	0.961	0.959
	(0.027)	(0.029)	(0.038)
High school	0.759***	0.750***	0.769***
	(0.025)	(0.026)	(0.034)
Some college	0.669***	0.659***	0.641***
	(0.026)	(0.027)	(0.035)
College degree	0.527***	0.546***	0.582***
	(0.019)	(0.021)	(0.029)
$25,000 - 49,999	0.618***	0.640***	0.749***
	(0.017)	(0.019)	(0.028)
$50,000 - 74,999	0.505***	0.524***	0.672***
	(0.017)	(0.018)	(0.031)

$75,000 - 149,999	0.393***	0.407***	0.606***
	(0.014)	(0.015)	(0.031)
$150,000 and over	0.292***	0.303***	0.481***
	(0.016)	(0.018)	(0.040)
Veteran status	1.183***	1.179***	1.095**
	(0.039)	(0.040)	(0.049)
Constant	0.049***	0.613	47.071***
	(0.009)	(0.215)	(27.793)
Observations	238,947	164,992	53,041
Year FE	YES	YES	YES
State FE	YES	YES	YES
Pseudo R-squared	0.127	0.105	0.0666

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Appendix H

Table H.1 The multiple regression results of Model 2 by sex

VARIABLES	(1) Total	(2) Male	(3) Female
2.5 generation	0.882**	0.892	0.875*
	(0.048)	(0.071)	(0.065)
Second generation	0.881**	0.844**	0.912
	(0.050)	(0.071)	(0.071)
First generation	0.715***	0.619***	0.799***
	(0.029)	(0.038)	(0.042)
Hispanic white	0.870***	0.874*	0.876**
	(0.041)	(0.061)	(0.055)
African American	1.108***	1.069	1.141***
	(0.038)	(0.056)	(0.052)
Indigenous	1.272**	1.338**	1.209
	(0.128)	(0.186)	(0.178)
Asian	0.977	0.978	0.968
	(0.063)	(0.096)	(0.082)
Pacific Islander	1.051	1.185	0.917
	(0.224)	(0.345)	(0.289)
Other	1.105	1.108	1.096
	(0.107)	(0.157)	(0.147)
Age	1.027***	1.021***	1.031***
	(0.003)	(0.005)	(0.005)
Age squared	1.000***	1.000***	1.000***
	(0.000)	(0.000)	(0.000)
Marital status	0.718***	0.737***	0.705***
	(0.016)	(0.024)	(0.021)
Metropolitan	0.966	0.980	0.953
	(0.026)	(0.039)	(0.036)
High school	0.782***	0.814***	0.762***
	(0.024)	(0.037)	(0.032)
Some college	0.666***	0.691***	0.652***
	(0.024)	(0.036)	(0.032)
College Degree	0.531***	0.526***	0.540***
	(0.018)	(0.026)	(0.025)
$25,000 - 49,999	0.610***	0.582***	0.632***
	(0.016)	(0.023)	(0.022)
$50,000 - 74,999	0.499***	0.481***	0.510***
	(0.016)	(0.022)	(0.022)
$75,000 - 149,999	0.391***	0.359***	0.422***
	(0.013)	(0.018)	(0.020)

$150,000 and over	0.296***	0.278***	0.313***
	(0.016)	(0.020)	(0.023)
Veteran status	1.186***	1.162***	1.121
	(0.039)	(0.043)	(0.139)
Constant	0.025***	0.028***	0.022***
	(0.003)	(0.005)	(0.003)
Observations	297,460	141,907	155,553
Year FE	YES	YES	YES
State FE	YES	YES	YES
Pseudo R-squared	0.135	0.130	0.142

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Appendix I

Table I.1 The results of logistic regression of Hearing, divided by sex, Model 2

VARIABLES	(1) Total	(2) Male	(3) Female
2.5 generation	0.944	0.989	0.899
	(0.088)	(0.120)	(0.132)
Second generation	1.034	1.104	0.950
	(0.101)	(0.141)	(0.143)
First generation	0.733***	0.714***	0.756**
	(0.060)	(0.079)	(0.092)
Constant	0.002***	0.001***	0.003***
	(0.001)	(0.000)	(0.001)
Pseudo R-squared	0.162	0.165	0.158

Notes: Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table I.2 The results of logistic regression of Visual, divided by sex, Model 2

VARIABLES	(1) Total	(2) Male	(3) Female
2.5 generation	0.861	0.995	0.740
	(0.133)	(0.208)	(0.170)
Second generation	0.911	0.867	0.938
	(0.140)	(0.201)	(0.191)
First generation	0.933	0.758*	1.085
	(0.094)	(0.119)	(0.141)
Constant	0.003***	0.002***	0.005***
	(0.001)	(0.001)	(0.002)
Pseudo R-squared	0.0808	0.0881	0.0827

Table I.3 The results of logistic regression of Memory, divided by sex, Model 2

VARIABLES	(1) Total	(2) Male	(3) Female
2.5 generation	1.017	0.965	1.071
	(0.099)	(0.135)	(0.147)
Second generation	0.679***	0.593***	0.773*
	(0.076)	(0.098)	(0.118)
First generation	0.624***	0.471***	0.794**
	(0.049)	(0.058)	(0.081)
Constant	0.023***	0.027***	0.016***
	(0.004)	(0.007)	(0.004)
Pseudo R-squared	0.0774	0.0798	0.0811

Table I.4 The results of logistic regression of Physical, divided by sex, Model 2

VARIABLES	(1) Total	(2) Male	(3) Female
2.5 generation	0.809***	0.850	0.783**
	(0.060)	(0.098)	(0.076)
Second generation	0.853**	0.864	0.844
	(0.067)	(0.106)	(0.087)
First generation	0.763***	0.727***	0.786***
	(0.040)	(0.063)	(0.052)
Constant	0.002***	0.002***	0.002***
	(0.000)	(0.001)	(0.001)
Pseudo R-squared	0.152	0.148	0.152

Table I.5 The results of logistic regression of Mobility, divided by sex, Model 2

VARIABLES	(1) Total	(2) Male	(3) Female
2.5 generation	0.827*	0.711*	0.908
	(0.088)	(0.125)	(0.122)
Second generation	0.879	0.702**	1.007
	(0.090)	(0.125)	(0.126)
First generation	0.964	0.786*	1.093
	(0.072)	(0.100)	(0.099)
Constant	0.011***	0.013***	0.010***
	(0.002)	(0.004)	(0.003)
Pseudo R-squared	0.133	0.119	0.144

Table I.6 The results of logistic regression of Selfcare, divided by sex, Model 2

VARIABLES	(1) Total	(2) Male	(3) Female
2.5 generation	0.760*	0.775	0.740
	(0.121)	(0.179)	(0.164)
Second generation	0.912	0.819	0.986
	(0.135)	(0.185)	(0.193)
First generation	1.072	1.001	1.109
	(0.114)	(0.170)	(0.151)
Constant	0.001***	0.001***	0.001***
	(0.000)	(0.001)	(0.001)
Pseudo R-squared	0.101	0.107	0.102