DISCUSSION PAPER SERIES

Language Barriers and Immigrant Health Production

Andrew Clarke
Ingo E. Isphording

February 2015

I Z A

Language Barriers and Immigrant Health Production

Andrew Clarke
University of Melbourne

Ingo E. Isphording
IZA

ABSTRACT

Language Barriers and Immigrant Health Production[*]

We study the impact of language deficiency on the health production of childhood migrants to Australia. Our identification strategy relies on a quasi-experiment comparing immigrants arriving at different ages and from different linguistic origins by utilising a measure of differences along a continuous range of linguistic distances. Our main results indicate a large negative effect of English deficiency on physical health that is robust to a range of different specifications. In the presence of considerable non-classical measurement error in self-reported language proficiency, our results provide lower and upper bounds for the true effect of English deficiency on health of one half and a full standard deviation in the health score respectively. The empirical analysis is framed in terms of a Grossman model which indicates a twofold role of language skills in health production: language deficiency directly affects the efficiency of health production and indirectly affects access to health inputs. We provide some suggestive evidence on the relative importance of these distinct roles.

Corresponding author:

Ingo E. Isphording
IZA
Schaumburg-Lippe-Str. 5-9
53113 Bonn
Germany
E-mail: isphording@iza.org

1 Introduction

At the individual level, good health outcomes for the foreign born are both a requisite and an outcome of successful integration into their host country society and provide an important indicator for the well-being of the immigrant population. While at a more aggregate level, the health status of the immigrant population has large fiscal implications, shaping the net impact of immigration.

A crucial channel through which the health of immigrants is affected during their residence in the host country is the acquisition of the destination language. Although there is a large epidemiological literature providing empirical evidence for the association of low literacy with inferior health outcomes, there is only limited evidence on the causal mechanisms.

These causal pathways may be summarised in terms of two main effects. First, language proficiency affects the access to inputs into health production such as access to jobs that pay higher wages and involve greater safety. Language skills are also associated with improvements in the efficiency of health production either by raising the marginal productivity of the inputs to health production (productive efficiency) or by allowing a more efficient combination of the inputs to health production (allocative efficiency). These efficiency gains can involve a greater ability to interact with the health system and a greater ability to process health information.

Despite a solid theoretical foundation for a causal effect of language skills on health, empirical studies will find it difficult to quantity the magnitude of the causal effect because language skills are endogenous. First, confounding variables, such as immigrant's access to social networks, which are correlated with both language proficiency and health, generate an omitted variable bias in the estimated return to language skills on health. Second, immigrant incentives to accumulate language skills will depend, at least partially, upon the expected payoff in terms of labour market earnings to language proficiency. This itself will likely depend upon an individual's latent health stock, producing causality from better health to language proficiency. Third, when language proficiency is self-reported, measurement error in language skills might be

considerable and will lead to a downward bias in the estimated return to measured language skills on health (Dustmann and van Soest, 2001).

In order to identify the causal effect of language proficiency on health, we construct a quasi-experiment based on a systematic decrease in the ability to learn new languages during childhood. The linguistically based 'critical period hypothesis' implies there is no age-dependent relationship in language outcomes if the age at first exposure to a new language is before the onset of a distinct threshold. If the age of exposure to a new language occurs after this threshold there is a decay in the capacity to learn the language. Moreover, the steepness of this decay depends on the dissimilarity between the home and host country language. For example, relative to immigrants whose first language is linguistically distant to the destination language, linguistically close immigrants face lower costs of acquisition (in terms of effort) from a later exposure to the destination country language with an age at immigration profile in language acquisition that is less pronounced (Isphording, 2014). The identification strategy exploits these differences, across linguistic backgrounds, in the systematic decay in the ability to learn new languages during childhood.

This strategy follows in spirit the strategy of Bleakley and Chin (2004), who use a control group design comparing immigrants from English- and non-English speaking countries at different ages at arrival. Compared to their approach, we offer an important contribution. Instead of using a simple control and treatment group design by distinguishing between English and non-English-speaking countries of birth, we apply a measure of subtle differences along a continuous range of linguistic distances. Pooling together immigrants from non-English-speaking countries of birth masks a large amount of heterogeneity in the costs (effort) of language acquisition. Obviously, European immigrants, who exhibit a closer proximity in linguistic background to English speakers than Asian immigrants, face lower costs of language acquisition. Relying on a continuous measure of linguistic distances derived from comparative linguistic research, our identification strategy combines information on both linguistic background and the timing of migration during childhood to derive an instrumen-

tal variable that resembles a difference-in-differences estimator by age at arrival and linguistic background.

To ensure identification, we condition our regressions on a large set of background characteristics to control for potential confounding variables, including a full set of country-of-birth indicators. One threat to identification lies in a potential relationship between initial health status and both the timing of migration (a later age at migration implies a longer exposure to home country health conditions) and linguistic distance (linguistically distant Asian countries also share lower quality health systems). If childhood health differs by age at immigration and this is systematically related to linguistic origin then our instrument is not valid. We address this issue by including indicators for initial health status as controls in all specifications. In addition, childhood health provides the basis for a placebo test for our identification strategy.

There are several compelling reasons that motivate the study of the relationship between language skills and health within Australia. First, Australia is a large immigrant receiving country, in 2013 almost 28% of the Australian population was born overseas. The health status of this considerable immigrant population has large fiscal implications, shaping the net impact of immigration. Australia also operates a a binding and tightly controlled immigrant selection policy, with all applicants for permanent residence, including the main applicant, partner and any dependents, required to satisfy a minimum health requirement. Second, Australia has a public health system that provides subsidised payments for medical services and for a high proportion of prescription medications, available to all permanent residents. Consequently, differences in health status associated with differences in language proficiency are more likely to reflect differences in the utilisation of health services rather than access to health services. Third, in light of the identification strategy discussed above, and mostly as a result of historical immigrant flows, Australia has a substantial stock of immigrants with an English mother tongue who act as a large control group for identifying the causal effects of language proficiency on health.

This paper presents evidence for a large detrimental causal effect of language deficiency on the (physical) health status of immigrants.[1] Our main results indicate a distinctive negative effect of English deficiency that is robust to a range of different specifications. While OLS estimates indicate a health disadvantage of about half a standard deviation in physical health scores for deficient speakers, IV estimates are approximately twice as large. The mean health stock of childhood immigrants who are proficient in English is larger by approximately one standard deviation, relative to immigrants who are classified as deficient in English, conditional upon a full set of covariates. This main effect is quite large. For example, the unconditional mean difference in health between native-born tertiary educated individuals and native-born high school dropouts is slightly less than one half of a standard deviation in the health score.

The empirical analysis is framed in terms of a Grossman model of health production in order to provide a link to the existing literature and to provide guidance about potential underlying mechanisms. Against the background of a Grossman-style health production model, we discuss potential mechanisms and pathways behind this causal effect. As we cannot find even suggestive evidence for a mediating role of health inputs, we argue that language deficiency most likely plays an important role in shaping efficiency in health production. Language-deficient speakers are less able to utilize the available inputs and health information, even if the access to these health inputs is in principle unaffected.

As measurement error in subjective information on language proficiency is likely affecting our OLS and IV results in different directions, inducing a downward attenu-

---

[1] Guven and Islam (2014) use a related identification strategy for a large number of labour market and socioeconomic outcomes, including health as one of many indicators. Based upon a very specific sample of established immigrants, they provide quite large estimates (almost two standard deviations in the health score) for the effect of English proficiency on the health of female immigrants and a large yet imprecise estimate for male immigrants, suggesting lower health for male English proficient immigrants. Compared to this study, they do not interpret their results within a theoretical framework of health production, they do not take into account linguistic heterogeneity among non-native speakers, and do not include controls for initial health status which potentially threatens identification. Based upon the sample used in this paper, we can find no evidence for a differential main effect by gender.

ation bias in the OLS results and potentially inflating our IV results by non-classical measurement error, our results provide lower and upper bounds for the effect of language deficiency on the health scores. We use an auxiliary dataset containing both subjective and objective information on language skills to provide secondary evidence for considerable non-classical misclassification error.

The results, provided in the paper, contribute to at least four different strands of the existing literature. First, the strength of the causal estimates we present in this study highlights the importance of non-market returns to language skills, that have been often overlooked in the literature (see Bleakley and Chin (2010) as a notable exception). Recent immigrants may not take into account these long-term non-market returns to investments into their language acquisition, which could lead to socially undesirable under-investments, imposing additional pressure on strained universal health care systems and a decreased fiscal net effect of immigration.

Second, the evidence for a large detrimental effect of language deficiency on health status provides support for the role of health literacy in health production (Nutbeam, 2008; Anderson et al., 2003; Jacobs et al., 2006). Third, with our focus on child migrants, our results contribute to the literature on age at arrival effects on socioeconomic outcomes and health in later life (Gonzalez, 2003; Myers et al., 2009; van den Berg et al., 2014). Language barriers introduce heterogeneity in the ability to adapt to a receiving country's society and might affect child migrants in important stages of their physical and mental development, with considerable long-run effects. Fourth, our results offer a potential mechanism for the 'healthy immigrant' effect (Antecol and Bedard, 2006)) where the health of immigrants declines with years of residence. The evidence presented in this paper indicates that persistent language barriers might represent a potential driving factor behind this relationship.

The paper proceeds as follows. Section 2 provides a theoretical link between language deficiency and the health outcomes and discusses how we achieve the identification of the causal effect of English deficiency. Section 3 explains our data setup. The results are presented in Section 4, Section 5 discusses potential mechanisms and

pathways. Section 6 concludes.

2 Identification strategy

To establish a link between language skills of immigrants and their health status, it is useful to depart from a standard Grossman-style health production model (Grossman, 1972). In this model, individual health is a depreciating capital stock dependent on an initial endowment that can be restocked by a health production process taking into account health investments, expenditures and constraints. Against this theoretical background of health production, the influence of language proficiency might follow the same reasoning as the well-analyzed effect of education on health (for overviews, see Grossman and Kaestner (1997) and Cutler and Lleras-Muney (2006)).

Language proficiency plays a twofold role: First, it has an indirect role by affecting the vector of inputs into the health production function. This indirect effect incorporates the large influence of language skills on the social and economic integration of immigrants, in terms of better employment probabilities, better access to safer jobs, better wages, but also in terms of access to social capital, the support provided by friends, co-workers and the broader community through social networks. Being deficient in the destination language might impose binding restrictions in immigrant's access these common inputs of the health production.

A more direct effect is given by the role of language proficiency in improving the efficiency of health production. This efficiency gain might be achieved by increased productive efficiency, raising the marginal productivity of health inputs. Alternatively, the efficiency gain could accrue through an increased allocative efficiency, allowing a more efficient combination of the inputs to health production. For a language-proficient immigrant, each given unit of input into the health production will yield higher returns to health than for a language-deficient immigrant, due to their higher efficiency in health production. These efficiency gains can act through a greater ability to interact with the health care system, including direct communication with health care personnel, but also access to and the processing of health information provided in

magazines, booklets or public information campaigns. A dearth in both the quantity and quality of health information might then affect the health behavior of language deficient immigrants.

Both these established direct and indirect theoretical links between language deficiency and health imply an unambiguously negative effect of language deficiency on health outcomes. We start our analysis with a simple reduced form model where adult health (H_i) is a linear function of English deficiency ($EDEF_i$, defined as not being able to communicate in English "very well"). Due to our specific sample of child migrants, we allow health to differ by the age at arrival (AAI_i), the linguistic distance between English and the linguistic background of an immigrant (LD_i), and other economic and demographic characteristics ($\mathbf{X}_i$):

$$H_i = \beta_0 + \beta_1 \, EDEF_i + \beta_2 \, AAI_i + \beta_3 \, LD_i + \mathbf{X}_i^T \gamma + \varepsilon_i \tag{1}$$

Since it is likely that English deficiency is correlated with the unobservable determinants of health this reduced form will not allow us to identify a causal effect of English deficiency on the health outcomes. First, it is likely that unobserved variables that are correlated with both health and language skills generate omitted variable bias. Second, since lower health decreases expected wages and lowers the incentives to invest into host-country specific human capital, we might expect causality from health to language skills. Third, in the absence of objective test scores for language fluency, we are restricted to a self-reported measure of English deficiency. The proclivity of such a self-reported measure to measurement error is well-documented (Charette and Meng, 1994; Dustmann and van Soest, 2001; de Coulon and Wolff, 2007). A robust result from previous papers, where labour market earnings is the outcome, is that the downward bias associated with measurement error exceeds the potential upward bias associated with unobserved factors, leading to an under-estimation of the true effect of language skills in naive models estimated by OLS.[2]

We deal with the measurement error in our English deficiency indicator by using

---

[2]Dustmann and van Soest (2001), Bleakley and Chin (2004), Chiswick and Miller (1995)

a twofold identification strategy following Dustmann and van Soest (2002). First, we take advantage of the longitudinal dimension in our data and replace the binary English deficiency indicator $\widetilde{\mathrm{EDEF}}_i$ with predictions from a fixed effects model on a non-linear time trend to address measurement errors in language deficiency that are time-varying. By using these predictions instead of the actual values, we should be able to reduce the amount of time-variant noise, analogously to using moving averages to override noisy measurements. Further, we develop a valid instrument variable based on age at immigration and linguistic background to correct for time-invariant measurement error and omitted variables.

By construction, measurement error in a binary variable is negatively correlated with the 'true' value of the binary variable. Individuals reporting to have English deficiency can only deviate towards English non-deficiency while individuals reporting not to have difficulties can only deviate towards English deficiency. In this case, the OLS estimates will be downward biased and take the same sign as the 'true' effect provided there are not too many 'false positives' (individuals with good language skills reporting language deficiencies). This form of non-classical measurement error will also lead to first stage estimates in the IV regressions that are biased towards zero and subsequently inflated reduced form IV estimates, provided there are also not too many 'false positives', as shown by Brachet (2008). Consequently, in the presence of non-classical measurement error, the OLS estimates provide a lower bound on the effect of language skills on health while the IV estimates provide an upper bound. We use auxiliary information from the Australian Adult Literacy and Lifeskills survey (ALLS) which contains both subjective and objective measures of language skills to assess the extent of this non-classical measurement error.

To derive a valid instrument for English deficiency, we follow the approach by Bleakley and Chin (2004) in using the systematic decrease in the ability to acquire new languages during childhood. Neuro-biologic and linguistic research indicates a certain 'critical period' up until which young children are able to acquire a native level of proficiency of new languages almost effortlessly (Newport, 2002). After this

critical period, the efficiency in learning new languages is drastically decreasing. As this critical period in language acquisition might coincide with further critical periods in youth development (van den Berg et al., 2014, see, e.g.), we cannot simply use this structural break as identifying variation. Importantly, the effect of the age at arrival on language acquisition is not homogenous, but differs by linguistic origin. Isphording (2014) shows that the effect is increasing with linguistic dissimilarity between the host and home country language. These sources of heterogeneity in language acquisition allow us to construct a quasi-experiment, in which we use a difference-in-differences-estimator comparing immigrants from differing linguistic backgrounds arriving before and after this critical period[3]. Native-speaking immigrants, who learned English as their first language and whose language acquisition is independent of their age at arrival, act as a control group.

To clarify our identification strategy, we show a simple comparison of the age at arrival profiles of native and non-native-speaking immigrants in Figure 1. While the language acquisition of native-speaking immigrants is unaffected by their age at arrival (dark-grey line), the probability of being deficient as an adult is strongly increasing by age at arrival for immigrants with a non-English linguistic background (Panel A). This pattern of differences between native-speaking and non-native-speaking immigrants seems to be transmitted into an analogous pattern in their health status. While health scores are marginally increasing by age at arrival for mative speakers, non-native speaking immigrants seem to suffer from significant health disadvantages when they arrive at later ages.

Our main identification assumption is that differences in linguistic origin raise the costs of language acquisition, but have no further direct or indirect effect on health production (conditional on $\mathbf{X}_i$). Differences in health outcomes between the linguistic groups can be solely attributed to these differences in the costs of language acquisition such that the depicted patterns in Figure 1 motivate our difference-in-

---

[3]The empirical evidence on the precise critical period is ambiguous, locating it somewhere in early adolescence. Chiswick and Miller (2008) find a more or less continuously decreasing ability to acquire new languages

difference approach. Comparing across different linguistic groups and different ages at arrival enables us to partial out non-linguistic influences of the age-at-arrival (such as earlier acculturation, earlier access to host country educational institutions, health care and diet) from the effect of age at arrival on language acquisition. It is the interaction of age at arrival and linguistic distance that is used for identification, controlling for a rich set of control variables as well as country-of-origin fixed effects.

A distinctive feature of our identification approach is the use of a continuous measure of linguistic dissimilarities between home and host-country language. The identification approach originally proposed by Bleakley and Chin (2004) relies on a simple comparison of native and non-native speakers identified by country of birth (that is the identification depends on whether one is born in a predominantly English-speaking country). This assignment to control and treatment group masks a great amount of heterogeneity within the group of non-native speaking immigrants. Immigrants from European countries share a certain amount of language similarity with English-speaking immigrants (Scandinavian/German immigrants originate in the Germanic language family, but also Italian or Greek migrants come from an Indo-European language origin). Asian immigrants do not share these language family ties. We therefore rely on a continuous measurement of language differences described in detail in the data section. This summary statistic of linguistic distance provides an additional source of variation beyond the simple comparison of English/Non-English speakers that we will exploit for identification purposes.

Our data additionally allows us to address another threat of identification. The Grossman model specifies health production is a function of an initial health endowment. A low initial health endowment might arise as a result of childhood exposure to low quality health care and nutrition and would feasibly be related to the timing of migration to Australia. We have also to expect that these childhood health conditions differ systematically between immigrants from different linguistic backgrounds. Being both related to linguistic background and timing of migration, the initial health status is then correlated with both our instrument and the adult health status. Therefore, it

9

is necessary to control for initial health status to ensure the validity of our exclusion restriction.

Formally, we estimate a first stage explaining English deficiency as a function of age at immigration, linguistic background LD_i) and the interaction between both variables:

$$\widetilde{EDEF}_i = \gamma_0 + \gamma_1 \, AAI_i + \gamma_2 \, LD_i + \gamma_3 \left\{ AAI_i \times LD_i \right\} + \mathbf{X}_i^T \phi + \omega_i \tag{2}$$

Conditional on our identifying assumption that differences in linguistic origin affect health production only through increasing the costs of language acquisition, and controlling for a rich set of individual characteristics, the interaction term $\left\{ AAI_i \times LD_i \right\}$ is uncorrelated with the unobservable determinants of adult health. Consequently, estimating equation (2) will provide exogenous predictions $\widehat{EDEF}_i$ that we can use to explain the health outcomes in the second stage excluding the instrument:

$$H_i = \beta_0 + \beta_1 \, \widetilde{EDEF}_i + \beta_2 \, AAI_i + \beta_3 \, LD_i + \mathbf{X}_i^T \gamma + \mu_i \tag{3}$$

To allow for non-linearities in the age at arrival profiles, age at arrival enters our specifications as a categorical variable rather than a linear variable. We estimate this equation using 2SLS with standard errors clustered on the individual level to account for the usage of repeated observations of the same individual from the HILDA panel. We run a number of nested specifications, with our preferred specification including time- and country-of-origin fixed effects and a large set of control variables on the individual and regional level, as outlined in the data section.

3   Data

The main data source for this study is the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA is a comprehensive household survey, starting in 2001 and covering a broad range of information on social and economic well-being.

The panel follows approximately 20,000 individuals in about 8,000 households and is repeated on an annual basis.

Utilising data from 2001 to 2011, we construct a pooled cross section of childhood migrants who arrived in Australia at an age before 18 and who are aged between 18 and 65 at the point of the interview. Our outcome of interest is the individual's physical health score derived from the SF–36, a commonly applied health questionnaire based on 34 items on different domains of health. The SF–36 is included annually in the HILDA questionnaire and provides an excellent and quasi-objective assessment of the health status of individuals up to the age of 65. Table 1 provides a summary of the structure of the SF–36. The single items are grouped into 4 different domains of physical health: physical functioning, role limitations due to physical problems, bodily pain and general health perceptions. From these sub-domains a physical health score is derived using a standardized coding scheme. The physical score is defined on a 0–100 range, with a sample mean of 78.5 points and a standard deviation of 19.9 points.

As an additional health indicator, the HILDA data offers information on subjective childhood health status before the age of 15. As noted above, controlling for individual differences in childhood health might be crucial for identification—not controlling for childhood health might overstate the importance of English deficiency on adult health by attributing part of the detrimental effect of prolonged exposure to bad environments during childhood to English deficiency. Indeed, results excluding this control indicate a 50 % increase in the estimates. Lacking this important control might be a reason for the very large effects found by Guven and Islam (2014). To prevent this violation of our identification assumption, we condition in our regressions on a set of indicator variables of subjective childhood health status.[4]

Our main variable of interest is the English language proficiency of the individual. English proficiency is assessed as the self-reported ability to speak English. From this variable, we construct a binary indicator of English deficiency that takes on the value

[4]The childhood health status is only asked once in the observational period. But being inherently time-invariant, we carry forward and backward the observation within individuals.

of 0 if the respondent reports speaking English "very well" (the highest category) and a value of 1 otherwise. We further use biographical information on year of birth and year of migration to construct age at arrival.

The principal advantage of the HILDA data is that it provides a rich set of control variables. We control in increasingly comprehensive specifications for basic controls (gender, age (in categories of five years), and parental education), regional controls (state-fixed effects, urban environment, the SEIFA index of local socio-economic advantages and disadvantages), and a full set of country-of-origin fixed effects. Finally, we also include potentially endogenous common inputs in the health production function (marital status, children in the household, household income, individual education and the labor force status). This final set of explanatory variables is potentially a function of English deficiency itself, threatening the causal interpretation of our estimates. Nonetheless their inclusion provides some insight on the importance of socioeconomic status in mediating the effect of language deficiency in the health production. All specifications include year-fixed effects capturing common unobservable time-variant influences.

As our identification of the effect of English deficiency on health relies on a comparison across immigrant groups from different linguistic origin, we augment the HILDA data with a measure for linguistic dissimilarities between the mother tongue of an immigrant and English.[5] This measure of linguistic distance is taken from the Automatic Similarity Judgement Program (ASJP) of the German Max Planck Institute of Evolutionary Anthropology (Bakker et al., 2009). Originally developed to explain the historical development of languages, it offers a convenient way to summarize the various dimensions of language dissimilarities (such as vocabulary, pronunciation, phonetic inventories, grammar etc.) within a single summary statistic. The resulting continuous and cardinal index summarizes the initial hurdles raised by linguistic barriers in the language acquisition of immigrants (Isphording and Otten, 2013).

---

[5]The HILDA data only offers information whether English was the first language spoken as a child. We therefore approximate the individual mother tongue by the predominant language of the immigrant country of birth.

The measurement relies on an automatic comparison of a list of words chosen for their universal availability. This 40-item list, referred to as the Swadesh-list (originally, it was compiled by the US linguist Morris Swadesh (1952)) is summarized in the upper panel of Table 2. To derive the linguistic distance measure, each word in this list is transcribed into a phonetic script. For each transcribed word pair, the minimum number of sounds is calculated to transfer the one word into the other (the "Levenshtein distance"). Examples for this calculation between English and German are provided in the lower panel of Table 2: While the German and English word for fish are similarly pronounced (resulting in a minimum distance of 0), the English *mountain* and the German word *Berg*, transcribed in phonetic script as *maunt3n* and *bErk*, have no similarities, resulting in a minimum distance of seven. Finally, a normalized average is taken over the different word pairs to derive the summary measure of linguistic distance.

The Levenshtein distance behaves well as a predictor for expert opinions on language relations. Its good performance relies on the idea that similarity in pronunciation accounts for the number of *cognates*, words sharing a common ancestor word. The number of cognates again is correlated with the distance towards a common ancestor language from which both languages were derived. Therefore, the linguistic distance measured by ASJP, while focusing primarily on pronunciation, is highly correlated with further dimensions of dissimilarity, such as grammar or vocabulary (Wichmann et al., 2010). In our HILDA sample, the closest languages to English spoken by immigrants are Dutch, Swedish and Westvlaams (Belgium). In contrast, Vietnamese, Somali and Khmer (Cambodia) display the most distant languages.

After deleting observations with missing information in our variables of interest, our final sample consists of 5,706 person-year observations, including 4,420 observations with English as mother tongue (first language learned), and 1,286 observations with a foreign linguistic background.[6]

Table 3 provides summary statistics by linguistic origin and age at arrival. Based

---

[6]These person-year observations are based on 569 native speakers and 278 non-native speakers

on the unconditional means, no striking systematic differences in health scores can be observed, although scores seem to be marginally higher for late arrivers (after age of 12) for native-speaking immigrants, while they are marginally lower for non-native speakers arriving late. Unsurprisingly, the main difference between the samples can be observed in the probability of being deficient in the English language. While the probability of being deficient is negligible for native speakers, around 8 % of the young arriving non-native-speaking immigrants report to have problems in communicating in English, while this share is increasing to almost a third (32.5%) for late arrivers who did not learn English as their first language.

Examining the control variables, non-native-speakers are less likely married, appear more disadvantaged in terms of labor market outcomes, and are more likely to reside in urban and disadvantaged regions. Being a late arriver seems to have different influences for native and non-native speakers. While in both groups the household income is lower for late arrivers, the unemployment probability is lower for native-speaking immigrants arriving late (compared to native-speaking immigrants arriving in young ages), while it is larger for non-native-speaking late arrivers compared to their young arriving counterparts.

We further use auxiliary data from the Australian version of the Adult Literacy and Life Skills Survey (ALLS), which offers both self-reported measures and objective test scores on language proficiency to investigate the amount of misclassification error underlying our subjective deficiency indicator. The assessment of literacy skills involved questions that emphasized the implementation and use of literacy skills in

daily activities.[7]

4   Results

We start our analysis by first providing a descriptive overview of the relationship between adult health outcomes and language deficiency by the age at arrival of childhood immigrants. Table 4 provides OLS estimates for the effect of immigrant age at arrival on physical health for a set of specifications with an increasingly comprehensive set of controls. Panel A displays average age at arrival patterns for the total sample, not distinguishing by linguistic origin. At a glance, there is no evidence that age at arrival has a decreasing average effect on health outcomes.

In Panel B of Table 4, we allow for a different age at arrival profile for immigrants who report to speak English as their mother tongue (native speaker) and those who learned a language different from English as a first language (non-native speaker). We interact the age at arrival categories with our measure of linguistic dissimilarity towards the home country language. The main effects of age at arrival, interpreted as the pattern for native speakers, remain insignificant. The only exception is a moderate positive effect on physical health for those who immigrated from age 12 on. In contrast, we find a very distinctive negative relationship between age at arrival and the health outcomes for non-native speakers, indicated by the interaction terms of age-at-arrival with the linguistic distance. The estimated pattern is insensitive to the

---

[7]For the purposes of this paper, only the single skill domain prose literacy is used (the knowledge and skills needed to understand and use various kinds of information from text including editorials, news stories, brochures and instructions manuals). Proficiency is measured along a continuous scale ranging from 0 to 500 with individual's score denoting a point at which they have an 80 per cent chance of successfully completing tasks with a similar level of difficulty. The survey used item response theory and multiple imputation methods to generate five plausible values. This methodology is widely used in educational testing and large scale surveys such as the OECD's Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS), and Progress in International Reading Literacy Study (PIRLS). Given the set of item responses, each of the five plausible values are all equally valid estimates of an individual's skill. While the statistical procedures for working with plausible values is well understood (Programme for International Student Assessment, 2009), for the purposes of providing secondary evidence on the existence of non-classical measurement error, unbiased estimates are provided using only a single plausible value.

inclusion of potential additional confounders. For the latest arrival cohort (arriving at ages between 15 and 17), the average difference in health between a native and the most linguistically distant speaker (with $LD = 1$) amounts to about 18 points, almost one standard deviation of the physical health score. This broadly corresponds to the pattern in Figure 1, Panel A, using only a binary indicator for native and non-native speakers.

As the principal difference between native and non-native-speaking immigrants is likely to be found in their language skills (conditional on our comprehensive set of controls), we now examine the age at arrival profile in the incidence of language deficiency in Table 5. As expected, there is no systematic pattern between age at arrival and the likelihood of English deficiency for native speakers. Consistent with the critical period hypothesis, the likelihood of being deficient in English increases strongly with age at arrival for non-native speakers. Immigrants arriving before adolescence do not display a large probability of English deficiency. In line with the critical period hypothesis, rates are monotonically increasing from age 12 onwards. Linguistically distant immigrants in the latest arrival cohort display a difference of 32.5 probability points of being deficient in English, compared to native-speaking immigrants. These patterns are surprisingly insensitive towards the inclusion of control variables. Again, this corresponds to the pattern displayed in Figure 1, Panel A.

Using the identification strategy outlined in Section 2, Table 6 presents two stage least squares estimates of the causal effect of language skills on health. The first stage results in column (1) are similar to the specifications in Table 5 and show the significance of our instrument. The Kleibergen-Paap test statistic does not indicate a weak instrument problem. Column (2) reports the baseline OLS specification. Without taking endogeneity of language skills and measurement error of the binary English deficiency indicator into account, the results imply a lower health score of about 11.9 points for immigrants deficient in English. Columns (3) - (6) summarize the IV estimates, which are generally twice as large as the (naive) OLS estimates. Now, being deficient in English is associated with a decrease in health scores by about 21.5

points, about one standard deviation in the health scores. Controlling for an increasingly comprehensive set of control variables has only a small effect on the magnitude of the estimates.

This main effect is quite large when compared to the much studied education-health gradient. The unconditional mean difference in health between native-born tertiary educated individuals and native-born high school dropouts is slightly less than one half of a standard deviation in the health score.

We test the robustness of the IV results by estimating the effect of English deficiency separately on the single subdomains of the SF36 physical health score in Table 7 (recall the structure in Table 1). The different domains, again defined on a 0-100 scale, are physical functioning (related to everyday physical activities), role-physical (related to health hurdles on the job), bodily pain (related to general pain experience) and general health (subjective health measure). The effects on these sub-domains are significant with the exception of general health. The point estimates are comparable in magnitude with the effect on the total SF36 score, with marginally larger effects on the domains of the role-physical domain and the bodily pain. The estimated effects are robust to whether socioeconomic controls are excluded (Panel A) or included (Panel B).

As noted above, if childhood health differs by age at arrival and is systematically related to linguistic origin, our instrument would be invalid. This potential threat to identification is associated with childhood health as an omitted variable, which is likely related to both linguistic background and the timing of migration, and consequently correlated with both our instrument and adult health status. Indeed, our results are sensitive to the inclusion of childhood health as a control variable such that excluding controls for childhood health increases the estimates by approximately 50%.

Our exclusion restriction is that the interaction of linguistic origin and age at immigration increase the costs of language acquisition yet have no further direct or indirect effect on health production (conditional on our comprehensive controls). This implies that there should be no causal effect of current language skills upon childhood

17

health. A 'placebo test' using childhood health (a 0-4 ordinal subjective measure of health before the age of 15) as the dependent variable provides only a very small point estimate (0.12) and insignificant relationship between English deficiency and childhood health status, providing suggestive evidence to support the validity of our exclusion restriction, at least in terms of threats to identification based upon childhood health.

In the presence of (non-classical) measurement error in the binary indicator for English deficiency, the OLS estimator will be downward biased and cannot be corrected without knowledge of the misclassification rates in reported deficiency and the true probability that an individual is deficient. Under the assumption that, conditional upon the true probability of English deficiency, the measurement errors in both the endogenous variable and the instrument are independent of each other and physical health, it is readily shown (Kane et al., 1999) that, even with a valid instrument, the IV estimator is biased by a factor that depends only on the misclassification rates in the endogenous variable.[8]

We use the ALLS data on subjective and objective language proficiency to assess these probabilities. Restricting the sample to be analogous to the HILDA sample, we have 432 individuals with valid information on self reported language skills and an objective assessment of adult literacy skills based upon a battery of questions that emphasised the implementation and use of literacy skills in daily activities. Proficiency

---

[8]For example, consider a model with a binary endogenous regressor (English proficiency) and a binary instrument based upon the interaction of later arrivers and non-English mother tongue, both of which are subject to misclassification errors. Let w be a binary variable that takes on a value of one if an individual is truly proficient in English and let $\tilde{w}$ be a binary variable that takes on a value of one if an individual self-reports to be proficient in English. Define the correctly classified observations:

$$q_1 = \Pr(\tilde{w}_i = 1 | w_i = 1)$$
$$1 - q_0 = \Pr(\tilde{w}_i = 0 | w_i = 0)$$

and the misclassified observations:

$$q_0 = \Pr(\tilde{w}_i = 1 | w_i = 0)$$
$$1 - q_1 = \Pr(\tilde{w}_i = 0 | w_i = 1)$$

It can be shown that the IV estimator is biased by a factor $1/(q_1 - q_0)$.

is measured along a continuous scale from 0 to 500.

The ALLS data provide a secondary sample to estimate the 'true' probability of English deficiency. A skill score of 225 is regarded by the survey developers as the "minimum required for individuals to meet the complex demands of everyday life and work in the emerging knowledge-based economy" (Statistics Canada, 2005). Accordingly, individuals with a skill score below 225 would be deemed as functionally illiterate. Based upon this threshold, the ALLS sample provides an estimate of 0.880 for the 'true' probability of English proficiency.

As noted by Bleakley and Chin (2004), there is no obvious way to link a continuous variable into a binary indicator without arbitrarily assuming functional relationships. A straightforward approach is to estimate a Probit model of the binary indicator for English proficiency on the continuous literacy test score and to estimate the rates of misclassification according to arbitrarily chosen cutoffs based upon the predicted probabilities. For example, assigning a cutoff of 0.50 implies that all individuals with a predicted probability less than 0.5 are treated as truly English deficient and all individuals with a predicted probability at least as large as 0.50 are treated as truly English proficient. In this case, the estimates of the conditional probabilities q_1 and q_0 and the estimate for the 'true' probability of English proficiency imply the OLS estimator is downward biased by an approximate factor of 0.55 while the IV estimator would be upward biased by an approximate factor of 1.610. Increasing the cutoff threshold to 0.60, 0.75, and 0.80, implies the IV estimator would be upward biased by a factor of 2.0.15, 3.420, or 4.068 respectively. This analysis, based upon the auxiliary ALLS sample, provides suggestive evidence that our finding that the IV estimates are almost twice as large as the OLS estimates is consistent with the presence of non-classical measurement error in self-reported language skills. Consequently, the OLS results in the Table 6 provide a lower bound on the effect of language deficiency on health and the IV estimates provide an upper bound.

5 Mechanisms

As discussed in Section 2, incorporating language deficiency into a Grossman-style health production model implies a twofold role of language deficiency, analogous to the role of education. First, it influences the access to further inputs into the health production and second, it affects the efficiency of these inputs for health production. The first role, the effect of language deficiency on the input vector of health production, incorporates most findings of the detrimental effects of language deficiency on the socioeconomic integration and assimilation of immigrants.

Empirically well-established links between language proficiency and labor market outcomes have been shown for employment probabilities and earnings (Bleakley and Chin, 2004; Dustmann and van Soest, 2002; Chiswick and Miller, 1995). As other authors have pointed out (Peri and Sparber, 2009; Chiswick and Miller, 2010), immigrants deficient in the destination language also face significant hurdles in their occupational choice, ending up in jobs with a low communicative task content, but with a larger focus on manual work. These jobs are typically characterized by greater physical demands and dangerous work conditions, ultimately leading to adverse health effects which are only partially overcome by compensating wage differentials. Besides labor market influences, destination language proficiency also influences the access to social support systems. For example, Bleakley and Chin (2010) report lower rates of intermarriage, higher divorce rates and and a larger probability of local segregation for English-deficient childhood migrants in the US. Less socially integrated migrants lack a social feedback mechanism through relatives, friends or colleagues who otherwise would potentially comment on observable unhealthy behavior like drinking, smoking and physical activity.

Can we provide suggestive evidence on these indirect effects of language deficiency through health inputs? Our reduced-form estimates presented in Table 6 only provide 'black-box' causal effects that do not allow us to distinguish between these distinct roles of language in immigrant health production. Lacking comprehensive retrospective information on health behavior and health productive inputs, we can only utilize

contemporary information on both health outcomes and the inputs to infer the effects of past behavior and inputs on the health production path. Current health status as a stock is affected by current work conditions only to a minor degree, but is determined largely by the history of working conditions throughout one's career. As we only observe current occupation, current income and current health behavior, underlying long-term mediating relationships might be in principle not identifiable. Nonetheless, without claiming a causal interpretation, correlations between English deficiency and contemporary mediators provide some suggestive evidence for the importance of health productive inputs. Table 8 summarizes partial correlations of mediating labor market outcomes, social inclusion indicators and indicators for observable health behavior and the English deficiency. Although the point estimates have the expected signs, most of the coefficients are very small in magnitude and remain insignificant. Again, this does not eliminate the possibility that English deficiency has influenced health behavior in the past, which has an effect on contemporary health scores. Rather our data does not allow us to address this.

Taken together, these results provide no evidence that a more difficult and costly access to health inputs acts as mechanism for the estimated strong negative effect of English deficiency on the physical health scores. Acknowledging that English deficiency still might have had an effect on inputs in the past, we cautiously take this as suggestive evidence that the second role of language deficiency, that is the loss in health production efficiency, might play a more important role.[9]

Although inputs are not affected by language hurdles, individuals who are less proficient in communicating in the destination language face a lower efficiency in health production from given levels of the health inputs. This efficiency-decreasing effect of English deficiency is related to a large literature in medical science addressing destination language deficiency as a barrier in health care access (Anderson et al., 2003; Jacobs et al., 2006). Immigrants with inadequate language skills rely on friends and

---

[9]This conclusion is also consistent with the evidence presented in Cutler and Lleras-Muney (2006, 2010) for the relationship between education and health. They conclude that income, health insurance, and family background can account for at most 30 percent of the education-health gradient.

relatives for interpreting services, increasing the noisiness of the health information received. Murray and Skull (2005) highlight that the access to trained interpreters is a key factor in the successful provision of health care to English deficient immigrants in Australia, though the funding is often inadequate. The efficiency loss is also reflected in the discussion in social sciences on insufficient health literacy impeding personal choices on health investments and the effectiveness of public health policy measures (Nutbeam, 2000, 2008).

6   Conclusion

The acquisition of the destination language is crucial for the social and economic integration of migrants into the destination country, and its value in the labor market has been extensively analyzed in terms of earnings and employment probabilities. In this paper, we analyzed the importance of language skills on the health outcomes of immigrants.

To identify the causal effect of language on the health outcomes, we utilized a structural break in the ability of acquiring new languages during childhood. Information on linguistic dissimilarity between home and destination language and on the age at arrival allows us to construct an IV variable for language skills that resembles a difference-in-differences estimator comparing immigrants from different linguistic backgrounds arriving at younger and older ages.

Our estimates imply a large and distinctive negative causal effect of language deficiency on the physical health score of immigrants. We interpret this finding within the theoretical framework of a Grossman health production model. As we do not find evidence for a mediating role of the access to health inputs, we argue that language skills likely play an important role by shaping the efficiency of health production, analogously to the role of education.

The magnitude of the negative health-language gradient highlights the non-market returns to language skills which could lead to socially undesirable under-investments, imposing additional pressure on strained universal health care systems and a de-

creased fiscal net effect of immigration. This might advocate the importance of early supportive interventions like language classes and the provision of interpreter services to reduce the language barriers in health care.

References

Anderson, L. M., Scrimshaw, S. C., Fullilove, M. T., Fielding, J. E. and Normand, J. (2003). Culturally competent healthcare systems: A systematic review. *American Journal of Preventive Medicine* 24: 68–79.

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

Bakker, D., Müller, A., Velupillai, V., Wichmann, S., Brown, C. H., Brown, P., Egorov, D., Mailhammer, R., Grant, A. and Holman, E. W. (2009). Adding typology to lexicostatistics: A combined approach to language classification. *Linguistic Typology* 13: 169–181.

Bleakley, H. and Chin, A. (2004). Language skills and earnings: Evidence from childhood immigrants. *The Review of Economics and Statistics* 86: 481–496.

Bleakley, H. and Chin, A. (2010). Age at arrival, English proficiency, and social assimilation among US immigrants. *American Economic Journal: Applied Economics* 2: 165–192.

Brachet, T. (2008). Maternal smoking, misclassification, and infant health. MPRA Paper 21466, University Library of Munich, Germany.

Charette, M. and Meng, R. (1994). Explaining language proficiency: Objective versus self-assessed measures of literacy. *Economics Letters* 44: 313–321.

Chiswick, B. R. and Miller, P. (2008). A test of the critical period hypothesis for language learning. *Journal of Multilingual and Multicultural Development* Vol. 29, No. 1: 16–29.

Chiswick, B. R. and Miller, P. W. (1995). The endogeneity between language and earnings: International analyses. *Journal of Labor Economics* 13: 246–288.

Chiswick, B. R. and Miller, P. W. (2010). Occupational language requirements and the value of English in the US labor market. *Journal of Population Economics* 23: 353–372.

Coulon, A. de and Wolff, F.-C. (2007). Language proficiency of immigrants and misclassification. *Applied Economics Letters* 14: 857–861.

Cutler, D. M. and Lleras-Muney, A. (2006). Education and Health: Evaluating Theories and Evidence. NBER Working Paper 12352, National Bureau of Economic Research.

Cutler, D. M. and Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of Health Economics* 29: 1–28.

Dustmann, C. and van Soest, A. (2001). Language fluency and earnings: Estimation with misclassified language indicators. *The Review of Economics and Statistics* 83: 663–674.

Dustmann, C. and van Soest, A. (2002). Language and the earnings of immigrants. *Industrial and Labor Relations Review* 55: 473–492.

Gonzalez, A. (2003). The education and wages of immigrant children: the impact of age at arrival. *Economics of Education Review* 22: 203–212.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy* 80: 223–55.

Grossman, M. and Kaestner, R. (1997). Effects of education on health. In Behrman, J. and Stacey, N. (eds), *The Social Benefits of Education*. University of Michigan Press, 69–123.

Guven, C. and Islam, A. (2014). Age at migration, language proficiency and socioeconomic outcomes: Evidence from Australia. Unpblished manuscript, Monash University.

Isphording, I. E. (2014). Disadvantages of linguistic origin - Evidence from immigrant literacy scores. *Economics Letters* 123: 236–239.

Isphording, I. E. and Otten, S. (2013). The costs of babylon - linguistic distance in applied economics. *Review of International Economics* 21: 354–369.

Jacobs, E., Chen, A. H., Karliner, L. S., Agger-Gupta, N. and Mutha, S. (2006). The need for more research on language barriers in health care: a proposed research agenda. *Milbank Quarterly* 84: 111–133.

Kane, T. J., Rouse, C. E. and Staiger, D. (1999). Estimating returns to schoolong when schooling is misreported. NBER Working Paper 7235, National Bureau of Economic Research.

Murray, S. B. and Skull, S. A. (2005). Hurdles to health: immigrant and refugee health care in australia. *Australian Health Review* 29: 25–29.

Myers, D., Gao, X. and Emeka, A. (2009). The Gradient of Immigrant Age-at-Arrival Effects on Socioeconomic Outcomes in the U.S. *International Migration Review* 43: 205–229.

Newport, E. (2002). Critical periods in language development. In Nadel, L. (ed.), *Encyclopedia of Cognitive Science*. London: Macmillan Publishers Ltd./Nature Publishing Group.

Nutbeam, D. (2000). Health literacy as a public health goal: A challenge for contemporary health education and communication strategies into the 21st century. *Health Promotion International* 15: 259–267.

Nutbeam, D. (2008). The evolving concept of health literacy. *Social Science & Medicine* 67: 2072 – 2078.

Peri, G. and Sparber, C. (2009). Task Specialization, Immigration, and Wages. *American Economic Journal: Applied Economics* 1: 135–69.

Programme for International Student Assessment (2009). PISA Data Analysis Manual: SAS Second Edition. Organisation for Economic Cooperation and Development, Paris.

Statistics Canada (2005). Learning a Living: First Results of the Adult Literacy and Life Skills Survey. Minister of Industry, Canada and, Organisation for Economic Cooperation and Development (OECD), Paris.

Swadesh, M. (1952). Lexico-statistic dating of prehistoric ethnic contacts: With special reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96: 452–463.

van den Berg, G. J., Lundborg, P., Nystedt, P. and Rooth, D.-O. (2014). Critical periods during childhood and adolescence. *Journal of the European Economic Association* 12: 1521–1557.

Wichmann, S., Holman, E. W., Bakker, D. and Brown, C. H. (2010). Evaluating linguistic distance measures. *Physica A: Statistical Mechanics and its Applications* 389: 3632–3639.

Tables and Figures

Table 1: SF–36 domains of physical health

Items (34)	Domains (8)	Summary Measures (2)
Vigorous Activities Moderate Activities Lift, Carry, Groceries Climb Several Flights Climb One Flight Bend, Kneel Walk Mile Walk Several Blocks Walk One Block Bathe, Dress	*Physical functioning*	
Cut Down Time Accomplished Less Limited in Kind Had Difficulty	*Role-Physical*	**Physical Health**
Pain-Magnitude Pain-Interfere	*Bodily Pain*	
Sick Easier As Healthy as anyone Health To Get Worse Health Excellent	*General Health*	

Source: www.sf-36.org.

Table 2: The ASJP Linguistic distance measure

I	You	We	One
Two	Person	Fish	Dog
Louse	Tree	Leaf	Skin
Blood	Bone	Horn	Ear
Eye	Nose	Tooth	Tongue
Knee	Hand	Breast	Liver
Drink	See	Hear	Die
Come	Sun	Star	Water
Stone	Fire	Path	Mountain
Night	Full	New	Name

Word	English	German	Distance
fish	*fiS*	fiS	0
breast	*brest*	brust	1
hand	*hEnd*	hant	2
tree	*tri*	baum	4
Mountain	*maunt3n*	bErk	7

Source: Isphording and Otten (2013).

Table 3: Descriptive statistics by linguistic origin and age at arrival

	Total	Native speakers		Non-native speakers	
		Arrived before 12	Arrived after 12	Arrived before 12	Arrived after 12
Physical health (SF-36)	78.215	77.560	79.785	80.297	76.679
	(19.850)	(20.444)	(17.900)	(18.194)	(21.028)
English deficiency	0.041	0.001	0.015	0.066	0.351
	(0.160)	(0.013)	(0.104)	(0.160)	(0.364)
Linguistic distance	0.213	0.000	0.000	0.937	0.963
	(0.399)	(0.000)	(0.000)	(0.106)	(0.092)
Subjective childhood health status	3.324	3.335	3.362	3.279	3.245
	(0.976)	(0.967)	(1.027)	(1.000)	(0.904)
Female	0.529	0.495	0.506	0.663	0.592
	(0.499)	(0.500)	(0.500)	(0.473)	(0.492)
Age	41.399	42.497	45.317	34.947	37.057
	(12.355)	(11.308)	(12.367)	(13.413)	(12.840)
Age at arrival	7.444	4.972	14.276	6.469	15.169
	(5.085)	(3.487)	(1.960)	(3.210)	(1.748)
Married	0.698	0.751	0.716	0.531	0.550
	(0.459)	(0.432)	(0.451)	(0.499)	(0.498)
No. of children in household	0.643	0.683	0.464	0.675	0.617
	(1.016)	(1.026)	(0.804)	(1.168)	(0.972)
Household income (in 1,000 AUD)	93.524	93.931	86.465	104.799	84.031
	(83.385)	(83.876)	(62.257)	(107.586)	(60.602)
Full-time employed	0.529	0.549	0.548	0.449	0.478
	(0.499)	(0.498)	(0.498)	(0.498)	(0.500)
Part-time employed	0.226	0.220	0.248	0.263	0.161
	(0.418)	(0.415)	(0.432)	(0.441)	(0.368)
Unemployed	0.035	0.033	0.028	0.038	0.053
	(0.183)	(0.178)	(0.164)	(0.192)	(0.224)
Not in labor force	0.210	0.196	0.176	0.250	0.309
	(0.408)	(0.397)	(0.381)	(0.433)	(0.462)
Education: Tertiary	0.296	0.305	0.208	0.335	0.323
	(0.456)	(0.460)	(0.406)	(0.472)	(0.468)
Education: Certificate/Diploma	0.296	0.308	0.415	0.192	0.171
	(0.457)	(0.462)	(0.493)	(0.394)	(0.377)
Education: Year 12	0.167	0.154	0.112	0.246	0.235
	(0.373)	(0.361)	(0.315)	(0.431)	(0.424)
Education: Year 11 or below	0.241	0.234	0.265	0.228	0.271
	(0.428)	(0.424)	(0.442)	(0.420)	(0.445)
Father has higher education	0.414	0.441	0.346	0.396	0.362
	(0.493)	(0.497)	(0.476)	(0.489)	(0.481)
Mother has higher education	0.273	0.260	0.295	0.321	0.243
	(0.445)	(0.439)	(0.456)	(0.467)	(0.429)
Living in urban area	0.677	0.616	0.692	0.840	0.831
	(0.468)	(0.487)	(0.462)	(0.367)	(0.375)
SEIFA index of relative economic advantages and disadvantages	5.683	5.878	5.624	5.421	4.784
	(2.882)	(2.775)	(2.773)	(3.131)	(3.202)
Number of observations	5,706	3553	867	473	813

Notes: – Standard errors reported in parentheses –

Table 4: OLS results: Age at arrival and health outcomes by linguistic distance

	(1)	(2)	(3)	(4)
Panel A:				
Age at arrival (AAA)				
Ref. cat.: Age at arrival 0-2				
Age at arrival 3-5	-1.501	-1.200	-0.812	0.618
	(2.01)	(1.94)	(2.08)	(1.95)
Age at arrival 6-8	-0.295	-0.362	0.297	1.138
	(1.74)	(1.72)	(1.94)	(1.83)
Age at arrival 9-11	-3.606*	-3.418*	-2.793	-1.844
	(2.06)	(2.02)	(2.15)	(2.00)
Age at arrival 12-14	0.791	1.334	1.934	3.397
	(2.03)	(2.04)	(2.18)	(2.13)
Age at arrival 15-17	-1.217	-0.507	0.005	1.473
	(2.14)	(2.16)	(2.33)	(2.11)
Basic controls	yes	yes	yes	yes
Origin-fixed effects	no	yes	yes	yes
Regional effects	no	no	yes	yes
Socio-economic controls	no	no	no	yes
R^2	0.081	0.102	0.139	0.210
N	5706	5706	5706	5706
Panel B:				
Age at arrival (AAA) by linguistic distance (LD)				
Ref. cat.: Native speaker, Age at arrival 0-2				
Age at arrival 3-5	-1.424	-0.876	0.045	1.475
	(2.29)	(2.19)	(2.29)	(2.14)
Age at arrival 6-8	-0.055	0.013	0.923	1.996
	(2.04)	(2.01)	(2.21)	(2.05)
Age at arrival 9-11	-2.069	-1.859	-1.079	-0.223
	(2.42)	(2.36)	(2.47)	(2.25)
Age at arrival 12-14	2.310	3.021	3.944	5.191**
	(2.26)	(2.25)	(2.40)	(2.33)
Age at arrival 14-17	1.812	2.311	3.477	4.863*
	(2.85)	(2.82)	(2.94)	(2.58)
Linguistic distance	4.386*	4.815*	3.452	3.128
	(2.41)	(2.54)	(4.29)	(4.20)
LD × AAA 3-5	-1.938	-3.539	-6.924	-6.767
	(4.18)	(4.07)	(5.17)	(4.87)
LD × AAA 6-8	-3.243	-4.007	-5.609	-6.948*
	(3.24)	(3.29)	(4.00)	(3.91)
LD × AAA 9-11	-8.655**	-8.897**	-10.235**	-9.846**
	(4.28)	(4.28)	(4.82)	(4.45)
LD × AAA 12-14	-9.257**	-10.225**	-12.676***	-11.240**
	(4.14)	(4.12)	(4.87)	(4.79)
LD × AAA 15-17	-10.115**	-10.058**	-13.415***	-13.071***
	(4.15)	(4.14)	(4.87)	(4.37)
Basic controls	yes	yes	yes	yes
Origin-fixed effects	no	yes	yes	yes
Regional effects	no	no	yes	yes
Socio-economic controls	no	no	no	yes
R^2	0.087	0.107	0.146	0.216
N	5706	5706	5706	5706

Notes: – * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. – Standard errors reported in parentheses are clustered at the individual level. – Sample restricted to age 18 - 65. – The dependent health variables is based on SF-36 health scores ranged 0-100. – English deficiency is coded 1 if individuals report not to speak English "Very Well". Basic controls include gender, age (in categories), parental education, and time-fixed effects. – Regional controls include living in an urban environment, state indicators and SEIFA-IRSAD. – Socio-economic controls include marital status, individual education, number of children in household, household income, labor force status. –

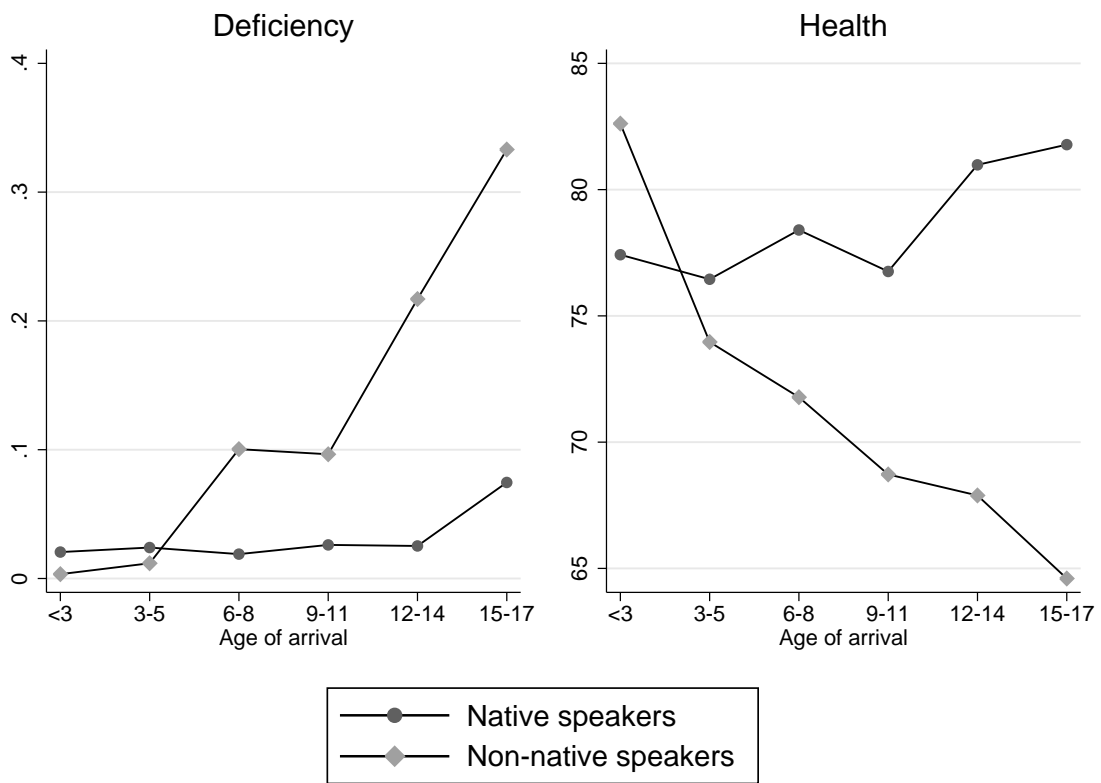Figure 1: Mean English deficiency and physical health by age at arrival

Table 5: OLS results: Age at arrival and English deficiency

	(1)	(2)	(3)	(4)
Panel A:				
Age at arrival (linear) by				
linguistic distance				
Age at arrival	0.001	0.001	0.001*	0.001
	(0.00)	(0.00)	(0.00)	(0.00)
Non-native speaker	-0.084***	-0.087***	-0.159***	-0.155***
	(0.03)	(0.03)	(0.04)	(0.04)
Interaction	0.027***	0.026***	0.027***	0.027***
	(0.00)	(0.00)	(0.00)	(0.00)
Basic controls	yes	yes	yes	yes
Origin-fixed effects	no	yes	yes	yes
Regional effects	no	no	yes	yes
Socio-economic controls	no	no	no	yes
R^2	0.366	0.376	0.531	0.541
N	5706	5706	5706	5706
Panel B:				
Age at arrival (AAA) by linguistic distance (LD)				
Ref. cat.: Native speaker, age at arrival 0-2				
Age at arrival 3-5	-0.002	-0.001	0.004	0.001
	(0.00)	(0.00)	(0.01)	(0.01)
Age at arrival 6-8	-0.005*	-0.006*	0.002	0.002
	(0.00)	(0.00)	(0.01)	(0.01)
Age at arrival 9-11	-0.004	-0.004	0.006	0.007
	(0.00)	(0.00)	(0.01)	(0.01)
Age at arrival 12-14	-0.002	-0.003	0.007	0.003
	(0.00)	(0.00)	(0.01)	(0.01)
Age at arrival 14-17	0.031	0.033	0.036*	0.032*
	(0.02)	(0.02)	(0.02)	(0.02)
Linguistic distance	0.047*	0.037	-0.050	-0.047
	(0.03)	(0.03)	(0.04)	(0.04)
LD × AAA 3-5	-0.036	-0.027	-0.011	-0.013
	(0.03)	(0.03)	(0.05)	(0.04)
LD × AAA 6-8	0.062	0.070	0.102*	0.103*
	(0.05)	(0.05)	(0.06)	(0.05)
LD × AAA 9-11	0.047	0.053	0.057	0.052
	(0.04)	(0.04)	(0.04)	(0.04)
LD × AAA 12-14	0.216**	0.221***	0.233***	0.226***
	(0.09)	(0.09)	(0.08)	(0.08)
LD × AAA 15-17	0.332***	0.332***	0.359***	0.357***
	(0.06)	(0.06)	(0.06)	(0.06)
Basic controls	yes	yes	yes	yes
Origin-fixed effects	no	yes	yes	yes
Regional effects	no	no	yes	yes
Socio-economic controls	no	no	no	yes
R^2	0.400	0.407	0.561	0.572
N	5706	5706	5706	5706

Notes: − * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. − Standard errors reported in parentheses are clustered at the individual level. − Sample restriction: age 18 - 65, age at arrival 0-17. − The dependent health variables is based on SF-36 health scores ranged 0-100. − English deficiency is coded 1 if individuals report not to speak English "Very Well". − Basic controls include gender, age (in categories), parental education, and time-fixed effects. − Regional controls include living in an urban environment, state indicators and SEIFA-IRSAD. − Socio-economic controls include marital status, individual education, number of children in household, household income, labor force status. −

Table 6: IV results: English deficiency and health outcomes

| | English Deficiency | Physical Health | | | | |
	OLS	OLS	IV 1	IV 2	IV 3	IV 4
English deficiency	–	-11.942***	-22.579**	-21.102**	-23.168**	-21.458**
		(3.81)	(10.35)	(10.28)	(10.53)	(9.52)
Ref. cat.: Age at arrival 0-2						
Age at arrival 3-5	0.001	0.888	-1.701	-1.323	-0.546	0.845
	(0.01)	(1.94)	(2.01)	(1.94)	(2.05)	(1.92)
Age at arrival 6-8	0.002	1.665	-0.197	-0.229	0.927	1.738
	(0.01)	(1.84)	(1.75)	(1.72)	(1.94)	(1.83)
Age at arrival 9-11	0.007	-1.318	-3.547*	-3.311*	-2.271	-1.342
	(0.01)	(1.99)	(2.05)	(1.99)	(2.13)	(1.96)
Age at arrival 12-14	0.003	4.334**	1.697	2.197	3.278	4.607**
	(0.01)	(2.14)	(2.07)	(2.06)	(2.21)	(2.14)
Age at arrival 14-17	0.032*	3.905*	2.514	2.999	4.024	5.191*
	(0.02)	(2.22)	(2.86)	(2.85)	(3.04)	(2.70)
Linguistic distance	-0.047	-4.209	2.089	1.608	-3.143	-3.354
	(0.04)	(3.09)	(1.96)	(1.90)	(3.01)	(3.03)
LD × AAA 3-5	-0.013	–	–	–	–	–
	(0.04)					
LD × AAA 6-8	0.103*	–	–	–	–	–
	(0.05)					
LD × AAA 9-11	0.052	–	–	–	–	–
	(0.04)					
LD × AAA 12-14	0.226***	–	–	–	–	–
	(0.08)					
LD × AAA 14-17	0.357***	–	–	–	–	–
	(0.06)					
Basic controls	yes	yes	yes	yes	yes	yes
Origin-fixed effects	yes	yes	no	yes	yes	yes
Regional effects	yes	yes	no	no	yes	yes
Socio-economic controls	yes	yes	no	no	no	yes
R^2	0.57	0.22	0.09	0.11	0.15	0.16
F-test of excluded instruments			12.62	12.46	9.75	10.25
N	5706	5706	5706	5706	5706	5706

Notes: – * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. – Standard errors reported in parentheses are clustered at the individual level. – The dependent health variables is based on SF-36 health scores ranged 0-100. – Sample restrictions identically to Table 4. – Basic controls include gender, age (in categories), parental education, and time-fixed effects. – Regional controls include living in an urban environment, state indicators and SEIFA-IRSAD. – Socio-economic controls include marital status, individual education, number of children in household, household income, labor force status. –

Table 7: IV results: English deficiency and health outcomes by SF-36 domain

	Physical functioning	Role-physical	Bodily pain	General health
Panel A: Baseline specification				
English deficiency	-20.070**	-25.591*	-27.863**	-19.146
	(10.02)	(15.02)	(11.88)	(12.11)
N	5706	5706	5706	5706
Panel B: Including socioeconomic controls				
English deficiency	-18.712**	-21.671	-27.308**	-18.142
	(9.03)	(13.57)	(11.38)	(11.54)
N	5706	5706	5706	5706

Notes: − * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. − Standard errors reported in parentheses are clustered at the individual level. − The dependent health variables are based on SF-36 health scores ranged 0-100. − Sample restrictions identically to Table 4. − − Basic controls include gender, age (in categories), parental education, and time-fixed effects. − Regional controls include living in an urban environment, state indicators and SEIFA-IRSAD. − Socio-economic controls include marital status, individual education, number of children in household, household income, labor force status. −

Table 8: English deficiency and health inputs

Panel A: *Labor market success*	Unemployed	Bluecollar Occupation	Household income (ln)
English deficiency	0.166**	0.047	-0.058
	(0.08)	(0.10)	(0.17)
N	4506.00	4304.00	4290.00
Panel B: *Social exclusion*	Noone to confide	Meeting Friends regularly	Noone to lean on
English deficiency	-0.009	-0.100	0.018**
	(0.01)	(0.09)	(0.01)
N	5706.00	5640.00	5706.00
Panel C: *Observable health behavior*	Physical Activity	Regular Smoking	Regular Drinking
English deficiency	-0.043	-0.046	0.087
	(0.08)	(0.07)	(0.09)
Basic controls	yes	yes	yes
Regional controls	yes	yes	yes
Origin-fixed effects	yes	yes	yes
N	5040.00	5697.00	5165.00

Notes: − * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. − Standard errors reported in parentheses are clustered at the individual level. − The dependent health variables is based on SF-36 health scores ranged 0-100. − Controls identically to Table 4. −