

Original Article

Submitted To: Psychological Medicine, PSM-D-23-01496.R1

**Cite this article:** Anderson KK, Khan JA, Edwards J, Le B, Longobardi G, Witt I, Alonso-Sánchez MF, Palaniyappan L (2024). Lost in translation? Deciphering the role of language differences in the excess risk of psychosis among migrant groups. *Psychological Medicine* 54, 3063–3070. <https://doi.org/10.1017/S003329172400117X>

Received: 23 October 2023  
Revised: 18 April 2024  
Accepted: 26 April 2024  
First published online: 22 May 2024

**Keywords:** incidence; language; migration; psychotic disorders; public mental health

**Corresponding author:** Kelly K. Anderson;  
Email: [kelly.anderson@schulich.uwo.ca](mailto:kelly.anderson@schulich.uwo.ca)

# Lost in translation? Deciphering the role of language differences in the excess risk of psychosis among migrant groups

Kelly K. Anderson<sup>1,2,3</sup> , Jahin Ali Khan<sup>1</sup>, Jordan Edwards<sup>4</sup>, Britney Le<sup>3</sup>, Giuseppe Longobardi<sup>5</sup>, Ivan Witt<sup>2</sup>, María Francisca Alonso-Sánchez<sup>2,6</sup> and Lena Palaniyappan<sup>2,7,8</sup>

<sup>1</sup>Department of Epidemiology and Biostatistics, Schulich School of Medicine and Dentistry, Western University, London Ontario Canada; <sup>2</sup>Department of Psychiatry, Schulich School of Medicine and Dentistry, Western University, London Ontario Canada; <sup>3</sup>ICES Western, London Ontario Canada; <sup>4</sup>Offord Centre for Child Studies, McMaster University, Hamilton Ontario Canada; <sup>5</sup>Department of Language and Linguistic Science, University of York, York UK; <sup>6</sup>CIDCL, Escuela de Fonoaudiología, Universidad de Valparaíso, Chile; <sup>7</sup>Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Montreal PQ Canada and <sup>8</sup>Robarts Research Institute, Western University, London Ontario, Canada

## Abstract

**Background.** Migration is a well-established risk factor for psychotic disorders, and migrant language has been proposed as a novel factor that may improve our understanding of this relationship. Our objective was to explore the association between indicators of linguistic distance and the risk of psychotic disorders among first-generation migrant groups.

**Methods.** Using linked health administrative data, we constructed a retrospective cohort of first-generation migrants to Ontario over a 20-year period (1992–2011). Linguistic distance of the first language was categorized using several approaches, including language family classifications, estimated acquisition time, syntax-based distance scores, and lexical-based distance scores. Incident cases of non-affective psychotic disorder were identified over a 5- to 25-year period. We used Poisson regression to estimate incidence rate ratios (IRR) for each language variable, after adjustment for knowledge of English at arrival and other factors.

**Results.** Our cohort included 1 863 803 first-generation migrants. Migrants whose first language was in a different language family than English had higher rates of psychotic disorders (IRR = 1.08, 95% CI 1.01–1.16), relative to those whose first language was English. Similarly, migrants in the highest quintile of linguistic distance based on lexical similarity had an elevated risk of psychotic disorder (IRR = 1.15, 95% CI 1.06–1.24). Adjustment for knowledge of English at arrival had minimal effect on observed estimates.

**Conclusion.** We found some evidence that linguistic factors that impair comprehension may play a role in the excess risk of psychosis among migrant groups; however, the magnitude of effect is small and unlikely to fully explain the elevated rates of psychotic disorder across migrant groups.

## Background

The excess rates of psychotic disorders among migrant and ethnic minority groups have been called a ‘public health tragedy’ (Morgan & Hutchinson, 2010), and these inequities have persisted for nearly a century with little progress toward prevention. The most recent meta-analytic estimates suggest a more than two-fold greater risk of psychotic disorder among first-generation migrants, with persistence of risk into the second generation (Selten, van der Ven, & Termorshuizen, 2020).

One factor that may hold promise for improving our understanding of the relationship between migration and psychotic disorders is language. Indeed, language impairments are a key feature of schizophrenia and other psychotic disorders, with evidence for the central role of language spanning from etiology to diagnosis and therapeutics (Covington et al., 2005). It has been hypothesized that schizophrenia arose as a manifestation of the genetic evolution towards the capacity for language (Berlim, Mattevi, Belmonte-de-Abreu, & Crow, 2003), and disruptions in discourse are a key feature of thought disorder (Covington et al., 2005). Among people who are bilingual, the clinical presentation of psychotic symptoms may differ based on the language used for assessment, with more severe symptoms present when assessments are in the first language (Erkoreka, Ozamiz-Etxebarria, Ruiz, & Ballesteros, 2020). Advances in linguistic computational approaches suggest that language abnormalities are a potential biomarker for diagnosis and prognosis of psychotic disorders (De Boer, Brederoo, Voppel, & Sommer, 2020). Looking further upstream, early life socio-economic conditions

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



also have a pervasive effect across multiple domains of language use in adult life (Rowe, 2018), and key social determinants include social class, ethnicity, migrant status, parental education, and size of a family's social network (Perkins, Finegood, & Swain, 2013; Rowe, 2018). Many of these early life social conditions are also important markers of risk for psychotic disorders, yet the social pathogenesis of psychosis remains poorly understood (Shah, Mizrahi, & McKenzie, 2011).

Although it has been extensively studied for non-psychotic disorders (Montemitro et al., 2021), evidence on the role of language as an explanatory factor for psychotic disorders among migrant groups is limited. A recent case-control study by Jongsma and colleagues used a binary indicator that combined fluency in the host country language and linguistic distance (Jongsma et al., 2021), which is defined as the degree of relatedness between a migrant's first language and the dominant language in the host country, and is an important determinant of language acquisition (Ispording & Otten, 2014). Adjustment for this indicator attenuated the association between migrant status and psychotic disorders, and there was a nearly two-fold greater odds of psychosis among people with linguistic distance and/or low fluency in the majority language (Jongsma et al., 2021). Similarly, we have previously shown that first-generation migrants to Ontario who spoke neither of Canada's official languages (English and French) at arrival had a 13% higher risk of psychotic disorder, relative to people who spoke English at arrival, and this effect was specific to psychotic disorders and not a marker of risk for mood and anxiety disorders among migrant groups (Anderson, Le, & Edwards, 2022). Finally, it has also been hypothesized that diglossia may be important to understanding the etiology of psychotic disorders (Alherz, Almusawi, & Barry, 2019), which refers to a linguistic context where there is a 'high' form of language used for more formal communication – such as educational or employment settings – and a 'low' form of language that is used for everyday discourse (Schiffman, 2017). Migration imposes an induced diglossic environment, whereby the first language becomes the 'low' form of language, and the dominant language in the host country is used as the 'high' form of language in educational and employment contexts (Alherz, 2022). Diglossia has recently been shown to be associated with prodromal symptoms of psychosis among first-generation migrant groups (Alherz, Almusawi, & Alsayegh, 2022).

To advance knowledge in this nascent field, there is a need for longitudinal research that considers the separate effects of linguistic distance and fluency in the host language, accounting for timing of language acquisition by considering factors such as age at migration (Alherz, 2022; Alherz et al., 2019). Thus, we sought to explore the association between linguistic distance of the first language and the risk of psychotic disorders among first-generation migrant groups in Ontario, Canada.

## Methods

We followed the REporting of studies Conducted using Observational Routinely collected health Data (RECORD) guidelines (Benchimol et al., 2015) (online Supplement 1). The data were obtained from ICES, which is an independent, non-profit research institute whose legal status under Ontario's health information privacy law allows it to collect and analyze health care and demographic data, without consent, for health system evaluation and improvement. The use of the data in this project was authorized under section 45 of Ontario's Personal Health Information

Protection Act (PHIPA) and does not require review by a Research Ethics Board.

## Study design and source of data

We used linked population-based health administrative data from ICES to create a retrospective cohort of first-generation migrants who landed in Ontario over a 20-year period (1992 to 2011), which has been described previously (Anderson et al., 2022). The study cohort was predominantly based on data from the Immigration, Refugee, and Citizenship Canada Permanent Resident (IRCC-PR) database, which contains data on migrants who landed in Ontario after 1985 linked to the health administrative data (linkage rate = 86%) (Chiu et al., 2016). The IRCC-PR includes information from federal immigration records on the characteristics of migrants at the time of landing, including country of origin, migrant class, first language, and knowledge of Canada's official languages (English and French).

The cohort of first-generation migrants was linked to the health administrative data, which includes information on outpatient physician visits, emergency department visits, and hospitalizations covered under the universal Ontario Health Insurance Program (OHIP). Person-time follow-up began as of the 14th birthday for people aged 0–13 years at the time of migration, and as of the landing date for people aged 14+ years at migration, due to the low risk of psychotic disorder prior to age 14 (<3% of cases) (Solmi et al., 2022). The cohort was followed to the end of 2016, age 65, death, or the end of OHIP eligibility, which would indicate a move outside of Ontario or emigration from Canada. We had 5–25 years of follow-up available, depending on the landing date. A complete description of the databases and variables used is available in online Supplement 2. These datasets were linked using unique encoded identifiers and analyzed at ICES.

## Exposure variable

Language is a highly complex construct made up of different objects (e.g. sounds, words, syntax, etc.), and there are several approaches in the literature for classifying languages into categories reflecting linguistic distance (Jongsma et al., 2022). To identify classification approaches for the purposes of the current study, we sought advice through an international email listserv of linguists (<https://linguistlist.org>), and received a number of suggestions for classification approaches. Based on the advice we received, we classified first language using one of four approaches:

- i. *Language Genealogies* – Using a similar approach to the prior study by Jongsma et al. (2021), we operationalized linguistic distance in relation to English using a genealogical language tree, which is based on the classical lexical-etymological method. Each language was assigned the following relatedness score: 0 = first language is English; 1 = first language is on the same 'branch' as English within the same language family (i.e. Other Germanic Languages); 2 = first language is on a different branch but within the same language family as English (i.e. Non-Germanic Indo-European Languages); 3 = first language is in a different language family from English (i.e. Non-Indo-European Languages).
- ii. *Estimated Acquisition Time* – The Foreign Service Institute has developed an estimate of the time and difficulty associated with the acquisition of various languages (<https://>

[www.state.gov/foreign-language-training/](http://www.state.gov/foreign-language-training/)). This approach classifies 66 languages into four categories based on the time required to reach a professional working proficiency. Categories range from I, which are languages closely related to English (e.g. Dutch, Afrikaans) to IV, which represent languages which are exceptionally difficult relative to English (e.g. Japanese, Arabic, Chinese, Korean).

- iii. *Parametric Comparison Method (PCM) Score* – The PCM score calculates the distance between 54 different languages by comparing their properties on 94 binary syntactic parameters (Irimia et al., *In Press*). PCM transcends the established genealogical distinctions and provides scores ranging from 0 to 1, with higher scores reflecting greater syntactic distance from English. The distribution of PCM scores was bimodal, therefore we divided the non-zero scores into quartiles, and used English (score = 0) as the reference category in our main analyses.
- iv. *Automated Similarity Judgement Program (ASJP)* – The ASJP calculates the distance between two languages by comparing a set of core vocabulary. The score represents the normalized average of the number of additions, deletions, and substitutions required to transfer a word from one language to another – in other words, the similarity of synonymous words across different families (Wichmann, Holman, & Brown, 2022). Scores ranged from 0 to 105, with higher scores reflecting greater lexical distance from English. The distribution of scores was highly skewed, therefore we divided the non-zero scores into quartiles, and used English (score = 0) as the reference category in our main analyses.

There were nearly 900 different languages, including dialects, listed in the IRCC database, and these were classified by a member of the research team (JAK), with dual coding of a random sample of 10% of the languages by a second member of the team (IW), with a high level of agreement. Some languages were unable to be classified by one or more of the four approaches (range 0.1–37.6%), and there were a small number of languages that were unable to be classified by any approach (e.g. language isolate, sign language). We also obtained information on whether migrants could speak English at the time of arrival (yes/no), which is the dominant language in Ontario (93%) (Government of Canada, 2023). This variable was based on an indicator in the immigration record of whether migrants could speak one of Canada's national languages (English and French), which did not consider level of fluency or language proficiency.

### Outcome variable

We followed cohort members in the health administrative data to identify incident cases of non-affective psychotic disorder (schizophrenia, schizoaffective disorder, psychosis not otherwise specified) using a validated algorithm (Kurdyak, Lin, Green, & Vigod, 2015) (online Supplement 2). Cases were identified based on the presence of an inpatient hospitalization with a discharge diagnosis of non-affective psychotic disorder, or two outpatient physician visits for non-affective psychotic disorder within a 12-month period. A modified version of this algorithm has been previously validated against medical charts, and found to have high levels of sensitivity (94%) and adequate positive predictive value (62%) (Kurdyak et al., 2015).

### Other variables

We obtained information on other factors previously shown to be associated with the risk of psychotic disorders among migrants in the study cohort (Anderson et al., 2022) (online Supplement 2). Age at migration was classified as infancy (0–2 years), early childhood (3–6 years), middle childhood (7–12 years), adolescence (13–18 years), early adulthood (19–29 years), and adulthood (30+ years) (Anderson & Edwards, 2020). Binary sex was classified as male and female. Country of birth was classified as European (including Russia), African (excluding North Africa), Caribbean, South Asian, East Asian (including Southeast Asian), Latin American (including Central and South American), and North Africa & Middle East (Statistics Canada, 2010). Migrant class was categorized as economic, sponsored (including family reunification migrants), and refugee. We also obtained information on post-migration place of residence (urban and rural), and census-based neighborhood income quintile.

### Data analysis

We summarized the characteristics of the cohort using counts and proportions for categorical variables, and means and standard deviations for continuous variables. There was a low proportion of the sample with missing data (<6%), and those with missing data were excluded.

We used modified Poisson regression models with robust variance estimators (Zou, 2004) to estimate incidence rate ratios (IRR) for the association between each language variable and the risk of developing a psychotic disorder. We first estimated the unadjusted association between each language variable and the risk of psychotic disorder, followed by a partially adjusted model that included age at migration, sex, country of birth, migrant class, rurality of residence, and neighborhood income quintile. Finally, we estimated a fully adjusted model that included all variables from the partially adjusted model, in addition to knowledge of English at the time of arrival.

We conducted subgroup analyses by age at migration (<19 years *v.* 19+ years), as age is responsible for approximately 30% of the variance in second language acquisition (Granena & Long, 2013). We also stratified our analyses by migrant class (economic, sponsored, and refugee), as economic migrants to Canada must demonstrate proficiency in English or French, whereas sponsored migrants and refugees do not have this requirement. We conducted sensitivity analyses, including (i) removing people whose first language was English from the sample, and using PCM and ASJP scores as a continuous variable; and (ii) restricting the sample to people with complete data across all four language variables and repeating the analyses. This was done to assess the impact of comparing models with different samples in the main analysis.

All analyses were conducted using SAS Version 9.4 (SAS Institute, Cary, North Carolina), and results are presented as IRRs with 95% confidence intervals (CI). There were few differences between the partially adjusted and fully adjusted models, therefore our presentation of the findings will focus on the unadjusted and fully adjusted models. Full parameter estimates from the multivariable models are presented in online Supplements 3 to 6.

### Results

Our study cohort included 1 863 803 migrants who landed in Ontario, Canada between 1992 and 2011 (Table 1). The largest

proportion of migrants came from South Asian (27.2%) and East Asian (25.6%) countries, followed by migrants from European countries (17.1%) and from North Africa and the Middle East (10.9%). There was a relatively small proportion of migrants from Latin America (6.6%), the Caribbean (5.2%), or from African countries (5.0%). Approximately half of the cohort came as an economic migrant (48.5%) and were over 30 years of age at the landing date (51.8%). Over 60% of the sample spoke English at the time of entry into Canada. We identified 16 771 incident cases of psychotic disorder over the follow-up period, which have been described in detail previously (Anderson et al., 2022).

### Language genealogies

Nearly the entire sample (99.9%) was classified using the classical language genealogy approach grounded in lexical etymologies. In the unadjusted models, all non-English first language categories were associated with a lower risk of psychotic disorder, relative to English (Fig. 1). This difference remained in the fully adjusted model for other Germanic languages (IRR = 0.63, 95% CI 0.50–0.80), however there was no longer a significant effect for non-Germanic Indo-European languages (IRR = 0.98, 95% CI 0.92–1.05). People whose first language was in a non-Indo-European language family had significantly higher rates of psychotic disorder, relative to those whose first language was English (IRR = 1.08, 95% CI 1.01–1.16) (Fig. 1).

### Estimated acquisition time

We were able to classify 87.3% of the sample based on estimated acquisition time of the first language. Nearly all categories of language acquisition time had a lower risk of psychotic disorder in the unadjusted models, relative to English (Fig. 1). In the fully adjusted models, both Category I languages (approximately 600–750 class hours, IRR = 0.88, 95% CI 0.81–0.95) and Category II languages (approximately 900 class hours; IRR = 0.86, 95% CI 0.71–1.04) were associated with lower rates of psychotic disorder, relative to English, although the latter includes the possibility of a null effect. Category III languages (approximately 1100 class hours) were associated with a higher risk of psychotic disorder (IRR = 1.18, 95% CI 1.10–1.27), relative to English, however there was no increased risk for Category IV languages (approximately 2200 class hours; IRR = 0.96, 95% CI 0.88–1.06) (Fig. 1).

### Parametric comparison method (PCM) score

PCM scores were available for 62.4% of the study sample. Nearly all quartiles of PCM score were associated with a lower risk of psychotic disorder in the unadjusted models, relative to English (Fig. 1). In the fully adjusted model, languages with the closest syntactic distance to English (i.e. Quartile 1) were associated with a lower risk of psychotic disorder, relative to English (IRR = 0.92, 95% CI 0.85–0.99). We did not observe significant effects for the other quartiles of PCM score in the fully adjusted models (Fig. 1). In sensitivity analyses, where we used the PCM score as a continuous variable, we again did not find significant effects for PCM score in the fully adjusted model (IRR = 1.37, 95% CI 0.94–1.99).

### Automated Similarity Judgement Program (ASJP) score

ASJP scores were available for 99.0% of the study sample. In the unadjusted models, all quartiles of ASJP score had a lower risk of psychotic disorder, relative to English (Fig. 1). In the fully adjusted model, languages with the greatest lexical distance from English (i.e. Quartile 4) were associated with a higher risk of psychotic disorder, relative to English (IRR = 1.15; 95% CI 1.06–1.24). We did not observe significant effects for the other quartiles of ASJP score in the fully adjusted models. In sensitivity analyses, where we used the ASJP score as a continuous variable, we found that each one-point increase in ASJP score was associated with a 1% increase in the risk of psychotic disorder (IRR = 1.01, 95% CI 1.01–1.02).

### Subgroup and sensitivity analyses

The findings of our subgroup analysis stratified by age at migration (<19 years v. 19+ years) were largely aligned with our main analyses. Of exception, in adjusted analyses the protective effects of having a first language with lower acquisition times were only observed in people 19 years of age and older at migration, and not among those who were under the age of 19 at migration (online Supplement 7). We also conducted subgroup analyses stratified by migrant class (economic, sponsored, refugee). These were again largely aligned with our main analyses, with some evidence of a stronger magnitude of effect among sponsored migrants for some indicators of linguistic distance (online Supplement 8).

We conducted a sensitivity analysis restricted to people with complete data on all four indicators of linguistic distance, and the results were aligned with our main analyses (online Supplement 9).

### Discussion

Canada has a large and diverse migrant population from a wide range of geographic locations with different migratory patterns relative to other countries, thus providing a unique perspective on the epidemiology of psychotic disorder among migrant groups. Our analysis of a large, population-based cohort of migrants with detailed information on language enabled us to conduct a prospective analysis of the role of first language on the risk of psychotic disorders, accounting for knowledge of English at the landing date and a range of other markers of psychosis risk. Our findings suggest that linguistic distance of the first language may play some role in the excess risk of psychosis among migrants, although effects were varied across the different indicators of linguistic distance. In addition, the magnitude of effect was small, and marked differences in psychosis risk persisted for African and Caribbean groups after multivariable adjustment (online Supplement 3 to 6). If linguistic distance has an etiological role in psychotic disorder among migrant groups, we might have expected to see a gradient effect with higher psychosis risk with increasing linguistic distance, which was not observed. Furthermore, there may be opposing mechanisms at play – for example, East Asian languages such as Chinese, Japanese, and Korean have some of the highest scores for linguistic distance, but we have previously shown that East Asian groups have lower rates of psychotic disorder than the general population (Anderson et al., 2022). Conversely, Caribbean migrant groups have some of the highest rates of psychotic disorder, both in Canada and internationally (Anderson et al., 2022; Selten et al., 2020), and English and other languages with a smaller linguistic distance from English, such as Dutch and

**Table 1.** Sociodemographic characteristics of the cohort of first-generation migrant groups who landed in Ontario, Canada between 1992 and 2011 ( $n = 1\,863\,803$ )

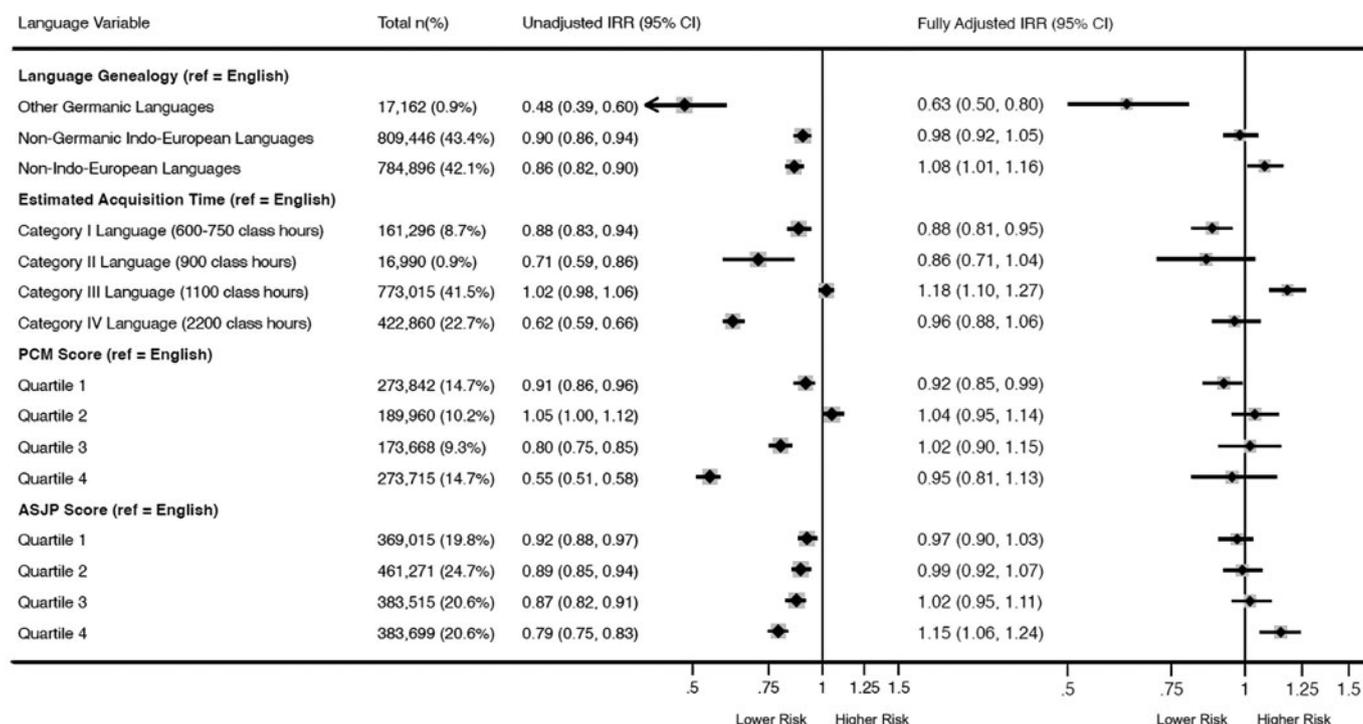
	Total Sample ( $n = 1\,863\,803$ )	Psychotic Disorder ( $n = 16\,771$ )	No Psychotic Disorder ( $n = 1\,847\,032$ )
<b>Male sex, <math>n</math> (%)</b>	898 575 (48.2%)	8959 (53.4%)	889 616 (48.2%)
Age at landing date, mean $\pm$ s.d.	31.18 $\pm$ 15.86	25.65 $\pm$ 12.67	31.23 $\pm$ 15.88
<b>Age group at landing date, <math>n</math> (%)</b> 0–2 years	19 770 (1.1%)	161 (1.0%)	19 609 (1.1%)
3–6 years	63 185 (3.4%)	682 (4.1%)	62 503 (3.4%)
7–12 years	150 318 (8.1%)	2102 (12.5%)	148 216 (8.0%)
13–18 years	172 351 (9.2%)	2544 (15.2%)	169 807 (9.2%)
19–29 years	492 799 (26.4%)	4820 (28.7%)	487 979 (26.4%)
30+ years	965 380 (51.8%)	6462 (38.5%)	958 918 (51.9%)
<b>Migrant class, <math>n</math> (%)</b> Economic	904 209 (48.5%)	6164 (36.8%)	898 045 (48.6%)
Sponsored	672 073 (36.1%)	6089 (36.3%)	665 984 (36.1%)
Refugee	255 764 (13.7%)	4169 (24.9%)	251 595 (13.6%)
Other or not stated	31 757 (1.7%)	349 (2.1%)	31 408 (1.7%)
<b>Country of origin, <math>n</math> (%)</b> European	318 448 (17.1%)	3008 (17.9%)	315 440 (17.1%)
African	93 002 (5.0%)	1716 (10.2%)	91 286 (4.9%)
Caribbean	97 718 (5.2%)	1495 (8.9%)	96 223 (5.2%)
South Asian	506 867 (27.2%)	4452 (26.5%)	502 415 (27.2%)
East Asian	477 017 (25.6%)	2706 (16.1%)	474 311 (25.7%)
Latin American	123 268 (6.6%)	1228 (7.3%)	122 040 (6.6%)
North Africa and Middle East	203 399 (10.9%)	1855 (11.1%)	201 544 (10.9%)
<b>Neighborhood income, <math>n</math> (%)</b> Quintile 1 (low)	690 340 (37.0%)	7611 (45.4%)	682 729 (37.0%)
Quintile 2	430 183 (23.1%)	3743 (22.3%)	426 440 (23.1%)
Quintile 3	315 726 (16.9%)	2431 (14.5%)	313 295 (17.0%)
Quintile 4	239 452 (12.8%)	1698 (10.1%)	237 754 (12.9%)
Quintile 5 (high)	179 282 (9.6%)	1189 (7.1%)	178 093 (9.6%)
<b>Rural place of residence, <math>n</math> (%)</b>	24 896 (1.3%)	167 (1.0%)	24 729 (1.3%)
<b>Knowledge of English at landing, <math>n</math> (%)</b>	1 137 091 (61.0%)	9947 (59.3%)	1 127 144 (61.0%)

Spanish, are predominant in the Caribbean. Thus, if linguistic distance is a contributing factor to the excess risk of psychotic disorder among migrants, the relationship is likely complex and multifactorial, and unlikely to fully account for the elevated rates of psychotic disorders among some migrant groups.

We note that lexico-semantic differences between languages (i.e. genealogical and ASJP approaches) were more influential than syntactic variations (i.e. PSM) from English. As the similarities in vocabulary are more critical for comprehension than syntactic structure *per se* (Longobardi & Guardiano, 2009), linguistic factors that impair understanding of the message communicated by a speaker may be a more relevant risk factor for psychosis. Prior research has found an independent effect of fluency or proficiency in the language of the host country on the odds or risk of psychotic disorders among migrant groups (Haasen, Lambert, Mass, & Krausz, 1998; Jongasma et al., 2021; Tarricone et al., 2022), whereas others have found no association (Garrido-Torres et al., 2022). These findings have typically been interpreted from a sociocultural perspective, whereby a lack of proficiency in the language of the host country increases marginalization, impedes social and occupational functioning, and prevents full

participation in society (Jongasma et al., 2021). Furthermore, a lack of proficiency in English may increase the likelihood of experiencing racism and discrimination (De Souza, Pereira, Camino, De Lima, & Torres, 2016), which has been shown to increase the risk of psychotic symptoms and disorders among migrant and ethnic minority groups (Bardol et al., 2020; Pearce, Rafiq, Simpson, & Varese, 2019). We did not find that adjusting for knowledge of English at the landing date had an appreciable impact on our observed estimates, although it was associated with a lower risk of psychotic disorder across all models (online Supplement 3–6).

From a neurobiological perspective, switching between languages engages multiple brain regions and cognitive control plays a crucial role in proficient bilingual language processing. The extent of the impact of switching languages on cognitive performance varies based on task difficulty (Köpke et al., 2021; Luque & Morgan-Short, 2021). The cognitive demands associated with languages that are more distant may lead to a different brain activation pattern (Cargnelutti, Tomasino, & Fabbro, 2022). Difficulties in predicting the communicative intent of a foreign language speaker may increase the computational cost in a social



IRR = Incidence Rate Ratio; PCM = Parametric Comparison Method; ASJP = Automatic Similarity Judgement Program; English Reference Group: n = 252,289 (13.5%)  
 \*Full model adjusted for sex, age at landing date, country of birth, migrant class, neighbourhood income quintile, rural residence, and knowledge of English

**Figure 1.** Distribution of first languages in the study sample, and results of the Poisson regression models for the four different approaches to classifying linguistic distance on the risk of non-affective psychotic disorder among first-generation migrant groups.

discourse, and reduce proactive processes that normally aid in comprehension (Kuperberg, Ditman, & Choi Perrachione, 2018). It has been hypothesized that cultural misunderstandings in verbal communication exacerbate experiences of social exclusion and discrimination among migrants, thus leading to stress-associated impacts on dopaminergic neurotransmission (Henssler et al., 2020).

Our measure of English knowledge was assessed at the time of migration and was limited to a binary indicator that did not consider level of fluency or language proficiency, nor did we consider longitudinal changes in English language capacity after migration. It has been hypothesized that the cognitive impairments that characterize the early course of psychotic illness, in addition to genetic variations (Vaughn & Hernandez, 2018), could impede second language acquisition, and therefore a lack of language proficiency could represent a pre-migratory marker of vulnerability to psychotic disorder (Montemitro et al., 2021); however, prior research did not find that second language acquisition was impaired in people with schizophrenia (Bersudsky, Fine, Gorjaltan, Chen, & Walters, 2005). Evidence from non-psychotic mental disorders suggests that language proficiency may have differential impacts at different time points in the post-migration phase (Montemitro et al., 2021), and a lack of dominant language proficiency has been shown to correlate with symptoms of paranoia, particularly persecutory ideation (Thomas, Bentall, Hadden, & O’Hara, 2017). Future research should explore how second language acquisition and proficiency throughout the post-migration period may impact the risk of psychotic disorder among migrant groups.

### Limitations

The most notable limitation of our findings is the high degree of complexity associated with quantifying language-related variations. There is likely extensive heterogeneity in English fluency, pre-migration English exposure, and individual aptitude for second language acquisition among the migrants in our cohort. Furthermore, there is no validated measure to classify linguistic distance, and we used a binary indicator of English language knowledge. We have attempted to minimize the impact of this by exploring multiple approaches to classifying linguistic distance. We were also unable to classify all languages using each approach. We explored the effect of this in sensitivity analyses, with little effect on our findings, but caution should be exercised when interpreting the findings from the PCM scores. Linguistic distance *per se* is a coarse measure of challenges in comprehension and production that may be faced by migrant groups; sociolinguistic differences (e.g. phonology, vocabulary, and dialectic differences) are critical factors (Palaniyappan, 2021) that we were not able to examine using an administrative database.

Additionally, our classification using estimated acquisition time was based on the time required for English speakers to learn a second language, rather than for people who speak other languages to learn English; we have assumed that the acquisition times would be similar, which may not be valid. Low English proficiency or a lack of translation services may be a barrier to accessing mental health services, which may have led to a detection bias or a differential likelihood of misdiagnosis for some groups. We are missing information on non-physician mental health services, such as psychologists and community mental

health providers, as well as important contextual factors, such as same-group ethnic density (Baker, Jackson, Jongasma, & Saville, 2021). We also did not adjust for education at landing in our multivariable models, which has been shown to be associated with language acquisition among migrant groups (Chiswick & Miller, 2001). Finally, we are unable to identify affective psychotic disorders in the health administrative data due to a lack of specificity in the outpatient diagnostic codes.

## Conclusions

There has been little progress toward prevention of the excess rates of psychosis among migrant groups. Our findings suggest that proficiency in the language of the host country may hold some promise for informing our understanding of psychotic disorders among migrant groups. Importantly, this represents a *modifiable* marker of risk for psychosis that could be the target of public mental health strategies. Further research to elucidate the role of linguistic distance and host language proficiency in the excess risk of psychotic disorders among migrant groups is warranted.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S003329172400117X>.

**Data availability statement.** The dataset from this study is held securely in coded form at ICES. While legal data sharing agreements between ICES and data providers (e.g. healthcare organizations and government) prohibit ICES from making the dataset publicly available, access may be granted to those who meet pre-specified criteria for confidential access, available at [www.ices.on.ca/DAS](http://www.ices.on.ca/DAS) (email: [das@ices.on.ca](mailto:das@ices.on.ca)). The full dataset creation plan and underlying analytic code are available from the authors upon request, understanding that the computer programs may rely upon coding templates or macros that are unique to ICES and are therefore either inaccessible or may require modification. The ICES analyst (Britney Le) had full access to the study database.

**Acknowledgements.** This study was supported by ICES, which is funded by an annual grant from the Ontario Ministry of Health (MOH) and the Ministry of Long-Term Care (MLTC). This study also received funding from Department of Psychiatry at Western University. Kelly Anderson is supported by a Tier 2 Canada Research Chair in Public Mental Health Research. Lena Palaniyappan is supported by the Monique H. Bourgeois Chair in Developmental Disorders and Graham Boeckh Foundation (Douglas Research Centre, McGill University) and a salary award from the Fonds de recherche du Québec-Santé (FRQS).

This document used data adapted from the Statistics Canada Postal CodeOM Conversion File, which is based on data licensed from Canada Post Corporation, and/or data adapted from the Ontario Ministry of Health Postal Code Conversion File, which contains data copied under license from ©Canada Post Corporation and Statistics Canada. Parts of this material are based on data and information compiled and provided by: Ontario Ministry of Health, and the Canadian Institute of Health Information (CIHI). The analyses, conclusions, opinions, and statements expressed herein are solely those of the authors and do not reflect those of the funding or data sources; no endorsement is intended or should be inferred. Parts or whole of this material are based on data and/or information compiled and provided by Immigration, Refugees and Citizenship Canada (IRCC) current to March 17, 2020. However, the analyses, conclusions, opinions and statements expressed in the material are those of the author(s), and not necessarily those of IRCC. We thank the Toronto Community Health Profiles Partnership for providing access to the Ontario Marginalization Index, and Dr Wolfram Hinzen of the Grammar and Cognition Lab at the Universitat Pompeu Fabra (Barcelona) for early discussions on classifying language families.

**Competing interests.** L. Palaniyappan has received personal fees from Janssen Canada, Otsuka Canada, SPMM Course Limited UK, and the

Canadian Psychiatric Association; editorial stipend from the Canadian Medical Association Journals ([jpn.ca](http://jpn.ca)); book royalties from the Oxford University Press; and investigator-initiated educational grants from Sunovion, Janssen Canada and Otsuka Canada. No other competing interests were declared.

## References

- Alherz, M. (2022). On the existence of a linguistic distance in schizophrenia. *Psychological Medicine*, 52(4), 796–797. doi:10.1017/S0033291720002585
- Alherz, M., Almusawi, H., & Alsayegh, A. (2022). Diglossia correlates with prodromal symptoms of psychosis among first-generation migrants. *Schizophrenia Bulletin Open*, 3(1), sgac066. doi:10.1093/schizbullopen/sgac066
- Alherz, M., Almusawi, H., & Barry, D. S. (2019). Diglossia in the etiology of schizophrenia: A hypothesis. *Journal of Nervous and Mental Disease*, 207(11), 987–992. doi:10.1097/NMD.0000000000001067
- Anderson, K. K., & Edwards, J. (2020). Age at migration and the risk of psychotic disorders: A systematic review and meta-analysis. *Acta Psychiatrica Scandinavica*, 141(5), 410–420. doi:10.1111/acps.13147
- Anderson, K. K., Le, B., & Edwards, J. (2022). Comparing risk factors for non-affective psychotic disorders with common mental disorders among migrant groups: A 25-year retrospective cohort study of 2 million migrants. *Schizophrenia Bulletin*, 48(5), 1011–1020. doi:10.1093/schbul/sbac021
- Baker, S. J., Jackson, M., Jongasma, H., & Saville, C. W. N. (2021). The ethnic density effect in psychosis: A systematic review and multilevel meta-analysis. *The British Journal of Psychiatry*, 219(6), 632–643. doi:10.1192/bjp.2021.96
- Bardol, O., Grot, S., Oh, H., Poulet, E., Zeroug-Vial, H., Brunelin, J., & Leaute, E. (2020). Perceived ethnic discrimination as a risk factor for psychotic symptoms: A systematic review and meta-analysis. *Psychological Medicine*, 50(7), 1077–1089. doi:10.1017/S003329172000094X
- Benchimol, E. I., Smeeth, L., Guttman, A., Harron, K., Moher, D., & Petersen, I., ... RECORD Working Committee. (2015). The REporting of studies conducted using observational routinely-collected health data (RECORD) statement. *PLOS Medicine*, 12(10), e1001885. doi:10.1371/journal.pmed.1001885
- Berlim, M. T., Mattevi, B. S., Belmonte-de-Abreu, P., & Crow, T. J. (2003). The etiology of schizophrenia and the origin of language: Overview of a theory. *Comprehensive Psychiatry*, 44(1), 7–14. doi:10.1053/comp.2003.50003
- Bersudsky, Y., Fine, J., Gorjaltan, I., Chen, O., & Walters, J. (2005). Schizophrenia and second language acquisition. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 29(4), 535–542. doi:10.1016/j.pnpbp.2005.01.004
- Cargnelutti, E., Tomasino, B., & Fabbro, F. (2022). Effects of linguistic distance on second language brain activations in bilinguals: An exploratory coordinate-based meta-analysis. *Frontiers in Human Neuroscience*, 15, 744489. doi:10.3389/fnhum.2021.744489
- Chiswick, B. R., & Miller, P. W. (2001). A model of destination-language acquisition: Application to male immigrants in Canada. *Demography*, 38(3), 391–409.
- Chiu, M., Lebenbaum, M., Lam, K., Chong, N., Azimae, M., Iron, K., ... Guttman, A. (2016). Describing the linkages of the immigration, refugees and citizenship Canada permanent resident data and vital statistics death registry to Ontario's administrative health database. *BMC Medical Informatics and Decision Making*, 16(1), 135. doi:10.1186/s12911-016-0375-3
- Covington, M. A., He, C., Brown, C., Naçi, L., McClain, J. T., Fjordbak, B. S., ... Brown, J. (2005). Schizophrenia and the structure of language: The linguist's view. *Schizophrenia Research*, 77(1), 85–98. doi:10.1016/j.schres.2005.01.016
- De Boer, J. N., Brederoo, S. G., Voppel, A. E., & Sommer, I. E. C. (2020). Anomalies in language as a biomarker for schizophrenia. *Current Opinion in Psychiatry*, 33(3), 212–218. doi:10.1097/YCO.0000000000000595
- De Souza, L. E. C., Pereira, C. R., Camilo, L., De Lima, T. J. S., & Torres, A. R. R. (2016). The legitimizing role of accent on discrimination against immigrants: Legitimization of accent on discrimination. *European Journal of Social Psychology*, 46(5), 609–620. doi:10.1002/ejsp.2216
- Erkoreka, L., Ozamiz-Etxebarria, N., Ruiz, O., & Ballesteros, J. (2020). Assessment of psychiatric symptomatology in bilingual psychotic patients: A systematic review and meta-analysis. *International Journal of Environmental Research and Public Health*, 17(11), 4137. doi:10.3390/ijerph17114137

- Garrido-Torres, N., Alameda, L., Cristóbal, J. P., Padilla, M. V., Robles, C. S., Canal-Rivero, M., ... Ruiz-Veguilla, M. (2022). Examining the relationship between psychosis and immigration in Spain: The effect of cannabis use and language barrier in a large psychosis sample. *Journal of Psychiatric Research*, *145*, 361–365. doi:10.1016/j.jpsychires.2021.11.005
- Government of Canada. (2023). Statistics on official languages in Canada. Retrieved 21 February 2023, from <https://www.canada.ca/en/canadian-heritage/services/official-languages-bilingualism/publications/statistics.html>
- Granena, G., & Long, M. (2013). Introduction and overview. In G. Granena, & M. Long (Eds.), *Sensitive periods, language aptitude and ultimate L2 attainment* (pp. vii–viii). Philadelphia, PA: John Benjamin Publishing Co.
- Haasen, C., Lambert, M., Mass, R., & Krausz, M. (1998). Impact of ethnicity on the prevalence of psychiatric disorders among migrants in Germany. *Ethnicity & Health*, *3*(3), 159–165. doi:10.1080/13557858.1998.9961858
- Hensler, J., Brandt, L., Müller, M., Liu, S., Montag, C., Sterzer, P., & Heinz, A. (2020). Migration and schizophrenia: Meta-analysis and explanatory framework. *European Archives of Psychiatry and Clinical Neuroscience*, *270*(3), 325–335. doi:10.1007/s00406-019-01028-7
- Irimia, M. A., Ceolin, A., Crisma, P., Guardiano, C., Kazakov, D., & Longobardi, G. (In Press). The parametric comparison method: A linguistic introduction to deep history. In *Oxford studies in historical and diachronic linguistics*. Oxford, UK: Oxford University Press.
- Ispording, I. E., & Otten, S. (2014). Linguistic barriers in the destination language acquisition of immigrants. *Journal of Economic Behavior & Organization*, *105*, 30–50. doi:10.1016/j.jebo.2014.03.027
- Jongsma, H. E., Gayer-Anderson, C., Tarricone, I., Velthorst, E., van der Ven, E., Quattrone, D., ... Kirkbride, J. B. (2021). Social disadvantage, linguistic distance, ethnic minority status and first-episode psychosis: Results from the EU-GEI case-control study. *Psychological Medicine*, *51*(9), 1536–1548. doi:10.1017/S003329172000029X
- Jongsma, H. E., Gayer-Anderson, C., van der Ven, E., Quattrone, D., di Forti, M., Arango, C., ... Kirkbride, J. B. (2022). Authors' reply to 'on the existence of a linguistic distance in schizophrenia'. *Psychological Medicine*, *52* (4), 798–799. doi:10.1017/S00332917200003013
- Köpke, B., Howells, R. K. R., Cortelazzo, F., Péran, P., De Boissezon, X., & Lubrano, V. (2021). Functional and structural differences in brain networks involved in language processing and control in highly proficient early and late bilinguals. *Journal of Neurolinguistics*, *59*, 100988. doi:10.1016/j.jneuroling.2021.100988
- Kuperberg, G. R., Ditman, T., & Choi Perrachione, A. (2018). When proactivity fails: An electrophysiological study of establishing reference in schizophrenia. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(1), 77–87. doi:10.1016/j.bpsc.2017.09.007
- Kurdyak, P., Lin, E., Green, D., & Vigod, S. (2015). Validation of a population-based algorithm to detect chronic psychotic illness. *The Canadian Journal of Psychiatry*, *60*(8), 362–368. doi:10.1177/070674371506000805
- Longobardi, G., & Guardiano, C. (2009). Evidence for syntax as a signal of historical relatedness. *Lingua. International Review of General Linguistics. Revue internationale De Linguistique Generale*, *119*(11), 1679–1706. doi:10.1016/j.lingua.2008.09.012
- Luque, A., & Morgan-Short, K. (2021). The relationship between cognitive control and second language proficiency. *Journal of Neurolinguistics*, *57*, 100956. doi:10.1016/j.jneuroling.2020.100956
- Montemitto, C., D'Andrea, G., Cesa, F., Martinotti, G., Pettorosso, M., Di Giannantonio, M., ... Tarricone, I. (2021). Language proficiency and mental disorders among migrants: A systematic review. *European Psychiatry*, *64*(1), e49. doi:10.1192/j.eurpsy.2021.2224
- Morgan, C., & Hutchinson, G. (2010). The social determinants of psychosis in migrant and ethnic minority populations: A public health tragedy. *Psychological Medicine*, *40*(5), 705–709. doi:10.1017/S0033291709005546
- Palaniyappan, L. (2021). More than a biomarker: Could language be a bio-social marker of psychosis? *Npj Schizophrenia*, *7*(1), 42. doi:10.1038/s41537-021-00172-1
- Pearce, J., Rafiq, S., Simpson, J., & Varese, F. (2019). Perceived discrimination and psychosis: A systematic review of the literature. *Social Psychiatry and Psychiatric Epidemiology*, *54*(9), 1023–1044. doi:10.1007/s00127-019-01729-3
- Perkins, S. C., Finegood, E. D., & Swain, J. E. (2013). Poverty and language development: Roles of parenting and stress. *Innovations in Clinical Neuroscience*, *10*(4), 10–19.
- Rowe, M. L. (2018). Understanding socioeconomic differences in parents' speech to children. *Child Development Perspectives*, *12*(2), 122–127. doi:10.1111/cdep.12271
- Schiffman, H. F. (2017). Diglossia as a sociolinguistic situation. In F. Coulmas (Ed.), *The handbook of sociolinguistics* (pp. 205–216). Oxford, UK: Blackwell Publishing. doi:10.1002/9781405166256.ch12
- Selten, J.-P., van der Ven, E., & Termorshuizen, F. (2020). Migration and psychosis: A meta-analysis of incidence studies. *Psychological Medicine*, *50*(2), 303–313. doi:10.1017/S0033291719000035
- Shah, J., Mizrahi, R., & McKenzie, K. (2011). The four dimensions: A model for the social aetiology of psychosis. *The British Journal of Psychiatry: The Journal of Mental Science*, *199*(1), 11–14. doi:10.1192/bjp.bp.110.090449
- Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar de Pablo, G., ... Fusar-Poli, P. (2022). Age at onset of mental disorders worldwide: Large-scale meta-analysis of 192 epidemiological studies. *Molecular Psychiatry*, *27*(1), 281–295. doi:10.1038/s41380-021-01161-7
- Statistics Canada. (2010). Standard Classification of Countries and Areas of Interest. Retrieved from <https://www.statcan.gc.ca/eng/subjects/standard/sccai/2010/sccai>
- Tarricone, I., D'Andrea, G., Jongsma, H. E., Tosato, S., Gayer-Anderson, C., Stilo, S. A., ... Morgan, C. (2022). Migration history and risk of psychosis: Results from the multinational EU-GEI study. *Psychological Medicine*, *52* (14), 2972–2984. doi:10.1017/S003329172000495X
- Thomas, J., Bentall, R. P., Hadden, L., & O'Hara, L. (2017). Ethnic identity and paranoid thinking: Implicit out-group preference and language dominance predict paranoia in Emirati women. *Journal of Behavior Therapy and Experimental Psychiatry*, *56*, 122–128. doi:10.1016/j.jbtep.2016.10.004
- Vaughn, K. A., & Hernandez, A. E. (2018). Becoming a balanced, proficient bilingual: Predictions from age of acquisition & genetic background. *Journal of Neurolinguistics*, *46*, 69–77. doi:10.1016/j.jneuroling.2017.12.012
- Wichmann, S., Holman, E. W., & Brown, C. H. (2022). The ASJP Database. Retrieved from <https://asjp.cld.org>
- Zou, G. (2004). A modified Poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology*, *159*(7), 702–706. doi:10.1093/aje/kwh090