

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	<p>(a) Indicate the study’s design with a commonly used term in the title or the abstract</p> <p>(b) Provide in the abstract an informative and balanced summary of what was done and what was found</p>	<p>Study design reported in the abstract (page 2)</p> <p>Summary reported in abstract (page 2)</p>	<p>RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included.</p> <p>RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract.</p> <p>RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.</p>	<p>Type of data noted in the abstract (page 2)</p> <p>Geographic region and timeframe reported in abstract (page 2)</p> <p>Data linkage reported in abstract (page 2)</p>
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	Background and rationale reported on pages 3 and 4		
Objectives	3	State specific objectives, including any prespecified hypotheses	Objectives reported on page 4		

Methods					
Study Design	4	Present key elements of study design early in the paper	Summary of key methods presented in the abstract (page 2)		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Study setting described on page 5, other details of cohort creation reported on page 7		
Participants	6	<p>(a) <i>Cohort study</i> - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up</p> <p>(b) <i>Cohort study</i> - For matched studies, give matching criteria and number of exposed and unexposed</p>	<p>Described on pages 5-6</p> <p>Not applicable</p>	<p>RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided.</p> <p>RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided.</p> <p>RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.</p>	<p>Cohort creation described on pages 5-6, and a complete list of codes is available in Supplemental Table 2.</p> <p>Not applicable – no algorithms or codes were used to select the population</p> <p>Not applicable – study cohort based on one database</p>

Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	Described on pages 6-9, and in Supplemental Table 2	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Available in Supplemental Table 2
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Details of variable creation described on pages 6-9, and in Supplemental Table 2		
Bias	9	Describe any efforts to address potential sources of bias	Sensitivity analysis to assess information bias described on page 10		
Study size	10	Explain how the study size was arrived at	N/A – included all migrants who landed in Ontario over 20-year period		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	Described on pages 6-9		

Statistical methods	12	<p>(a) Describe all statistical methods, including those used to control for confounding</p> <p>(b) Describe any methods used to examine subgroups and interactions</p> <p>(c) Explain how missing data were addressed</p> <p>(d) <i>Cohort study</i> - If applicable, explain how loss to follow-up was addressed</p> <p>(e) Describe any sensitivity analyses</p>	Each of these elements can be found on pages 9-10		
Data access and cleaning methods		..		<p>RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.</p> <p>RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.</p>	<p>Described in acknowledgement section (page 19)</p> <p>Not applicable</p>
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Details of data linkages provided on pages 5 and 6, and details on linkage quality are provided on page 5

Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	Not applicable – study included all people who met the inclusion criteria	RECORD 13.1: Describe in detail the selection of the persons included in the study (i.e., study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Not applicable – study included all people who met the inclusion criteria
Descriptive data	14	(a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) <i>Cohort study</i> - summarise follow-up time (e.g., average and total amount)	Characteristics presented in Table 1 Information on handling missing data on page 9 Not applicable		
Outcome data	15	<i>Cohort study</i> - Report numbers of outcome events or summary measures over time	Presented in Figure 1		

Main results	16	<p>(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included</p> <p>(b) Report category boundaries when continuous variables were categorized</p> <p>(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period</p>	<p>Estimates presented in Figure 1 and online supplements 3-6</p> <p>N/A</p> <p>N/A</p>		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	Described on page 13 and Supplemental Tables 7-9		
Discussion					
Key results	18	Summarise key results with reference to study objectives	Summarized on page 14		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Described on page 16	RECORD 19.1: Discuss the implications of using data that were not collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Described on page 16

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Overall conclusions presented on page 17		
Generalisability	21	Discuss the generalisability (external validity) of the study results	Described in the last paragraph of page 16		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Source of funding described in acknowledgements (page 18)		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Addressed in acknowledgement section (page 20)

*Reference: Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015;12(10):e1001885.

*Checklist is protected under Creative Commons Attribution ([CC BY](https://creativecommons.org/licenses/by/4.0/)) license.

Variable Name	Source of Data	Description	Variable Name in Database	Values and Database Codes
Cohort Creation				
Migrant Status	IRCC	First-generation migrants to Ontario	LANDING_DATE	All people in the IRCC database with a landing date between January 1 1992 and December 31 2011
First Language	IRCC	First language	NAT_LANGUAGE	Categorized using one of 4 approaches, as described in the manuscript. Exact codes available from authors on request.
Sex	RPDB	Recorded sex	SEX	1 = Male 2 = Female
Country of Birth	IRCC	Country of birth	COUNTRY_BIRTH	Based on standardized ICES macro (\$CIC_COUNTRY_WORLDAREA_REGION) that maps the country code to world area regions
Education	IRCC	Educational attainment as of the landing date (among those aged 19+ years at landing)	EDUCATION_QUALIFICATION	1 = None (00) 2 = Secondary or Less (01, 7660) 3 = Formal Trade Certification or Apprenticeship (02, 7661) 4 = Non-University Certificate or Diploma (03, 7662) 5 = Some University, No Degree (04, 7663) 6 = Bachelor's Degree (05, 7664) 7 = Some Post-Grad, No Degree (06, 7665) 8 = Master's Degree (07, 7666) 9 = Doctorate (08, 7667)
Marital Status	IRCC	Marital status at the landing date (among those aged 19+ years at landing)	MARITAL_STATUS	1 = Single (1, 8871) 2 = Married (2, 8872) 3 = Widowed (3, 8873) 4 = Divorced (4, 7, 8874, 8877) 5 = Separated (5, 8875) 6 = Common-Law (6, 8876)
Age at Migration	RPDB, IRCC	Age at index date	BDATE LANDING_DATE	1 = Infancy (0-2 years) 2 = Early Childhood (3-6 years) 3 = Middle Childhood (7-12 years) 4 = Adolescence (13-18 years) 5 = Early Adulthood (19-29 years) 6 = Adulthood (30+ years)
Migrant Class	IRCC	Migrant class listed on application	IMMIGRATION_CATEGORY	Based on standardized ICES macro (\$CIC_IMMIGCATEG_MAINCATEG)
Rural Residence	RPDB	Rural place of residence at landing, defined using the Rurality Index of Ontario. Areas with score of 40 or above are considered rural.	RIO2008	1 = Rural 0 = Non-Rural
Income Quintile	RPDB	Neighbourhood-level income quintile at landing date	INCQUINT	1 = Lowest Income Quintile 5 = Highest Income Quintile
Canadian Language	IRCC	Knowledge of either of Canada's national languages as of the landing date	OFFICIAL_LANGUAGE	1 = English (1, 15220) 2 = French (2, 15221) 3 = Both (3, 15222) 4 = Neither (4, 15223)
Outcome Variables				
Diagnosis of Psychotic Disorder	DAD, OMHRS, OHIP, NACRS	Diagnosis of a non-affective disorder based on either one hospitalization or two outpatient visits with a diagnostic code for psychotic disorder within a 12 month period	DXCODE, DX10CODE (DAD) AXIS1_DSM4CODE_DISCH1 (OMHRS) DXCODE (OHIP) DX10CODE1 (NACRS)	1 = Non-Affective Psychotic Disorder (ICD-9 = 295.X, 298.X; ICD-10 = F20, F25, F29) 0 = No Diagnosis of Psychotic Disorder

Supplemental Table 3 – Results of the Poisson regression models for the effect of linguistic distance classified using language genealogies on the risk of non-affective psychotic disorder among first-generation migrant groups (n = 1,863,793).

Variables		Unadjusted IRR (95%CI)	Partially Adjusted* IRR (95%CI)	Fully Adjusted** IRR (95%CI)
Language Genealogies	English	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	Other Germanic Languages	0.48 (0.39 - 0.60)	0.65 (0.52 - 0.82)	0.63 (0.50 - 0.80)
	Non-Germanic Indo-European Languages	0.90 (0.86 - 0.94)	1.04 (0.98 - 1.10)	0.98 (0.92 - 1.05)
	Non-Indo-European Languages	0.86 (0.82 - 0.90)	1.13 (1.05 - 1.21)	1.08 (1.01 - 1.16)
Male Sex (Reference = Female)		1.22 (1.18 - 1.25)	1.19 (1.16 - 1.23)	1.20 (1.16 - 1.24)
Country of Origin	European	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	African	2.06 (1.94 - 2.18)	1.54 (1.44 - 1.65)	1.56 (1.46 - 1.67)
	Caribbean	1.51 (1.42 - 1.61)	1.40 (1.29 - 1.51)	1.43 (1.32 - 1.54)
	South Asian	1.02 (0.97 - 1.07)	0.95 (0.91 - 1.00)	0.97 (0.92 - 1.01)
	East Asian	0.64 (0.61 - 0.68)	0.65 (0.60 - 0.69)	0.64 (0.60 - 0.69)
	Latin American	1.08 (1.01 - 1.15)	0.95 (0.89 - 1.01)	0.97 (0.90 - 1.04)
	North Africa and Middle East	1.08 (1.02 - 1.15)	0.96 (0.90 - 1.02)	0.96 (0.90 - 1.02)
Age Group at Landing Date	0 to 2 years	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	3 to 6 years	1.08 (0.91 - 1.29)	1.03 (0.86 - 1.23)	1.04 (0.87 - 1.24)
	7 to 12 years	1.09 (0.93 - 1.28)	1.01 (0.86 - 1.19)	1.03 (0.87 - 1.22)
	13 to 18 years	1.00 (0.85 - 1.17)	0.90 (0.76 - 1.06)	0.92 (0.78 - 1.09)
	19 to 29 years	0.67 (0.58 - 0.79)	0.61 (0.52 - 0.72)	0.64 (0.54 - 0.75)
	30+ years	0.56 (0.48 - 0.66)	0.57 (0.48 - 0.67)	0.59 (0.50 - 0.70)
Migrant Class	Economic	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	Family	2.04 (1.83 - 2.27)	1.99 (1.78 - 2.22)	2.02 (1.81 - 2.25)
	Refugee	2.36 (2.27 - 2.46)	1.90 (1.82 - 1.99)	1.90 (1.82 - 1.99)
	Other or Not Stated	1.49 (1.44 - 1.55)	1.42 (1.36 - 1.47)	1.39 (1.33 - 1.44)
Neighbourhood Income	Quintile 5 (High)	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	Quintile 4	1.08 (1.01 - 1.17)	1.03 (0.95 - 1.11)	1.02 (0.95 - 1.11)
	Quintile 3	1.15 (1.07 - 1.23)	1.04 (0.97 - 1.12)	1.04 (0.97 - 1.12)
	Quintile 2	1.26 (1.18 - 1.34)	1.08 (1.01 - 1.16)	1.08 (1.01 - 1.15)
	Quintile 1 (Low)	1.55 (1.46 - 1.65)	1.21 (1.13 - 1.29)	1.20 (1.13 - 1.28)
Rural Place of Residence (Reference = Urban)		0.70 (0.60 - 0.81)	0.83 (0.70 - 0.98)	0.83 (0.70 - 0.98)
English Language at Arrival (Reference = No)		0.88 (0.85 - 0.91)	-	0.90 (0.87 - 0.93)

*Adjusted for sex, age at landing date, country of birth, migrant class, neighbourhood income quintile, rural residence

**Adjusted for knowledge of English, in addition to all variables in the partially adjusted model

Supplemental Table 4 – Results of the Poisson regression models for the effect of linguistic distance classified using estimated acquisition time on the risk of non-affective psychotic disorder among first-generation migrant groups (n = 1,626,450).

Variables		Unadjusted IRR (95%CI)	Partially Adjusted* IRR (95%CI)	Fully Adjusted** IRR (95%CI)
Estimated Acquisition Time	English	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	Category I Language (600-750 class hours)	0.88 (0.83 - 0.94)	0.94 (0.87 - 1.02)	0.88 (0.81 - 0.95)
	Category II Language (900 class hours)	0.71 (0.59 - 0.86)	0.89 (0.73 - 1.07)	0.86 (0.71 - 1.04)
	Category III Language (1100 class hours)	1.02 (0.98 - 1.06)	1.24 (1.15 - 1.33)	1.18 (1.10 - 1.27)
	Category IV Language (2200 class hours)	0.62 (0.59 - 0.66)	1.04 (0.95 - 1.14)	0.96 (0.88 - 1.06)
Male Sex (Reference = Female)		1.22 (1.18 - 1.26)	1.20 (1.16 - 1.24)	1.21 (1.17 - 1.25)
Country of Origin	European	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	African	2.16 (2.02 - 2.31)	1.75 (1.63 - 1.88)	1.79 (1.66 - 1.92)
	Caribbean	1.51 (1.42 - 1.61)	1.58 (1.45 - 1.72)	1.63 (1.49 - 1.78)
	South Asian	1.15 (1.09 - 1.20)	1.03 (0.98 - 1.08)	1.05 (1.00 - 1.11)
	East Asian	0.64 (0.61 - 0.68)	0.75 (0.70 - 0.79)	0.76 (0.71 - 0.81)
	Latin American	1.08 (1.01 - 1.15)	1.15 (1.05 - 1.25)	1.19 (1.09 - 1.30)
	North Africa and Middle East	1.09 (1.03 - 1.16)	1.08 (1.01 - 1.16)	1.11 (1.04 - 1.19)
Age Group at Landing Date	0 to 2 years	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	3 to 6 years	1.09 (0.92 - 1.30)	1.03 (0.86 - 1.24)	1.05 (0.87 - 1.26)
	7 to 12 years	1.09 (0.92 - 1.28)	1.01 (0.85 - 1.20)	1.04 (0.87 - 1.23)
	13 to 18 years	0.98 (0.84 - 1.16)	0.89 (0.75 - 1.06)	0.92 (0.78 - 1.09)
	19 to 29 years	0.68 (0.58 - 0.80)	0.62 (0.52 - 0.73)	0.66 (0.55 - 0.78)
	30+ years	0.56 (0.47 - 0.65)	0.57 (0.48 - 0.67)	0.60 (0.51 - 0.71)
Migrant Class	Economic	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	Family	2.02 (1.81 - 2.27)	2.08 (1.86 - 2.34)	2.13 (1.89 - 2.39)
	Refugee	2.34 (2.24 - 2.43)	1.85 (1.77 - 1.94)	1.85 (1.77 - 1.94)
	Other or Not Stated	1.53 (1.47 - 1.59)	1.45 (1.39 - 1.51)	1.42 (1.36 - 1.49)
Neighbourhood Income	Quintile 5 (High)	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	Quintile 4	1.09 (1.01 - 1.17)	1.03 (0.95 - 1.12)	1.03 (0.95 - 1.11)
	Quintile 3	1.16 (1.08 - 1.25)	1.05 (0.97 - 1.13)	1.05 (0.97 - 1.13)
	Quintile 2	1.28 (1.20 - 1.37)	1.09 (1.01 - 1.17)	1.08 (1.01 - 1.16)
	Quintile 1 (Low)	1.58 (1.48 - 1.68)	1.21 (1.13 - 1.29)	1.20 (1.12 - 1.28)
Rural Place of Residence (Reference = Urban)		0.67 (0.57 - 0.78)	0.77 (0.65 - 0.93)	0.78 (0.65 - 0.93)
English Language at Arrival (Reference = No)		0.89 (0.86 - 0.92)	-	0.87 (0.84 - 0.91)

*Adjusted for sex, age at landing date, country of birth, migrant class, neighbourhood income quintile, rural residence

**Adjusted for knowledge of English, in addition to all variables in the partially adjusted model

Supplemental Table 5 – Results of the Poisson regression models for the effect of linguistic distance classified using parametric comparison method (PCM) score on the risk of non-affective psychotic disorder among first-generation migrant groups (n = 1,163,474).

Variables		Unadjusted IRR (95%CI)	Partially Adjusted* IRR (95%CI)	Fully Adjusted** IRR (95%CI)
PCM Score	English (Score of 0)	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	Quartile 1	0.91 (0.86 - 0.96)	0.97 (0.90 - 1.04)	0.92 (0.85 - 0.99)
	Quartile 2	1.05 (1.00 - 1.12)	1.09 (1.00 - 1.19)	1.04 (0.95 - 1.14)
	Quartile 3	0.80 (0.75 - 0.85)	1.06 (0.94 - 1.20)	1.02 (0.90 - 1.15)
	Quartile 4	0.55 (0.51 - 0.58)	1.01 (0.86 - 1.19)	0.95 (0.81 - 1.13)
Male Sex (Reference = Female)		1.14 (1.10 - 1.19)	1.13 (1.08 - 1.17)	1.13 (1.09 - 1.18)
Country of Origin	European	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	African	1.29 (1.15 - 1.46)	1.26 (1.11 - 1.43)	1.26 (1.11 - 1.43)
	Caribbean	1.58 (1.48 - 1.68)	1.39 (1.28 - 1.51)	1.42 (1.30 - 1.54)
	South Asian	1.24 (1.17 - 1.32)	1.03 (0.94 - 1.12)	1.05 (0.96 - 1.15)
	East Asian	0.64 (0.60 - 0.68)	0.69 (0.59 - 0.81)	0.69 (0.60 - 0.81)
	Latin American	1.13 (1.05 - 1.21)	1.00 (0.93 - 1.07)	1.02 (0.95 - 1.10)
	North Africa and Middle East	1.00 (0.93 - 1.08)	0.89 (0.79 - 1.01)	0.90 (0.80 - 1.01)
Age Group at Landing Date	0 to 2 years	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	3 to 6 years	1.09 (0.89 - 1.33)	1.05 (0.85 - 1.29)	1.26 (0.08 - 0.05)
	7 to 12 years	1.03 (0.86 - 1.25)	0.97 (0.80 - 1.18)	0.98 (0.81 - 1.20)
	13 to 18 years	0.90 (0.74 - 1.08)	0.81 (0.67 - 0.99)	0.83 (0.68 - 1.01)
	19 to 29 years	0.61 (0.51 - 0.74)	0.56 (0.46 - 0.68)	0.58 (0.48 - 0.70)
	30+ years	0.53 (0.44 - 0.64)	0.53 (0.44 - 0.64)	0.55 (0.45 - 0.67)
Migrant Class	Economic	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	Family	1.99 (1.75 - 2.26)	2.07 (1.82 - 2.35)	2.11 (1.85 - 2.40)
	Refugee	2.08 (1.97 - 2.19)	1.78 (1.68 - 1.89)	1.79 (1.69 - 1.90)
	Other or Not Stated	1.55 (1.49 - 1.62)	1.43 (1.36 - 1.50)	1.41 (1.34 - 1.48)
Neighbourhood Income	Quintile 5 (High)	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	Quintile 4	1.07 (0.97 - 1.17)	1.03 (0.94 - 1.13)	1.03 (0.94 - 1.13)
	Quintile 3	1.13 (1.04 - 1.24)	1.04 (0.95 - 1.14)	1.04 (0.95 - 1.13)
	Quintile 2	1.30 (1.20 - 1.41)	1.14 (1.04 - 1.23)	1.13 (1.04 - 1.23)
	Quintile 1 (Low)	1.59 (1.48 - 1.72)	1.26 (1.17 - 1.37)	1.26 (1.16 - 1.36)
Rural Place of Residence (Reference = Urban)		0.67 (0.57 - 0.80)	0.72 (0.59 - 0.88)	0.72 (0.60 - 0.88)
English Language at Arrival (Reference = No)		0.98 (0.94 - 1.02)	-	0.90 (0.86 - 0.95)

*Adjusted for sex, age at landing date, country of birth, migrant class, neighbourhood income quintile, rural residence

**Adjusted for knowledge of English, in addition to all variables in the partially adjusted model

Supplemental Table 6 – Results of the Poisson regression models for the effect of linguistic distance classified using automatic similarity judgement program (ASJP) score on the risk of non-affective psychotic disorder among first-generation migrant groups (n = 1,849,789).

Variables		Unadjusted IRR (95%CI)	Partially Adjusted* IRR (95%CI)	Fully Adjusted** IRR (95%CI)
ASJP Score	English (Score of 0)	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	Quartile 1	0.92 (0.88 - 0.97)	1.02 (0.95 - 1.09)	0.97 (0.90 - 1.03)
	Quartile 2	0.89 (0.85 - 0.94)	1.04 (0.97 - 1.12)	0.99 (0.92 - 1.07)
	Quartile 3	0.87 (0.82 - 0.91)	1.07 (0.99 - 1.15)	1.02 (0.95 - 1.11)
	Quartile 4	0.79 (0.75 - 0.83)	1.19 (1.10 - 1.29)	1.15 (1.06 - 1.24)
Male Sex (Reference = Female)		1.22 (1.18 - 1.25)	1.19 (1.16 - 1.23)	1.20 (1.16 - 1.24)
Country of Origin	European	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	African	2.07 (1.94 - 2.20)	1.54 (1.44 - 1.66)	1.56 (1.45 - 1.68)
	Caribbean	1.51 (1.42 - 1.61)	1.41 (1.30 - 1.53)	1.44 (1.33 - 1.56)
	South Asian	1.02 (0.97 - 1.07)	0.97 (0.92 - 1.02)	0.98 (0.93 - 1.03)
	East Asian	0.64 (0.61 - 0.68)	0.64 (0.60 - 0.69)	0.64 (0.59 - 0.68)
	Latin American	1.08 (1.01 - 1.15)	0.96 (0.89 - 1.03)	0.98 (0.92 - 1.05)
North Africa and Middle East		1.08 (1.02 - 1.15)	0.99 (0.92 - 1.06)	0.99 (0.92 - 1.06)
Age Group at Landing Date	0 to 2 years	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	3 to 6 years	1.08 (0.91 - 1.28)	1.03 (0.86 - 1.23)	1.04 (0.87 - 1.24)
	7 to 12 years	1.07 (0.91 - 1.26)	1.00 (0.85 - 1.18)	1.02 (0.86 - 1.20)
	13 to 18 years	0.98 (0.83 - 1.14)	0.89 (0.75 - 1.05)	0.91 (0.77 - 1.07)
	19 to 29 years	0.67 (0.57 - 0.78)	0.61 (0.52 - 0.72)	0.64 (0.54 - 0.75)
	30+ years	0.56 (0.48 - 0.65)	0.56 (0.48 - 0.66)	0.59 (0.50 - 0.70)
Migrant Class	Economic	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	Family	1.99 (1.75 - 2.26)	2.06 (1.84 - 2.30)	2.09 (1.87 - 2.33)
	Refugee	2.08 (1.97 - 2.19)	1.93 (1.85 - 2.02)	1.93 (1.85 - 2.02)
	Other or Not Stated	1.55 (1.49 - 1.62)	1.42 (1.36 - 1.47)	1.39 (1.33 - 1.45)
Neighbourhood Income	Quintile 5 (High)	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	Quintile 4	1.08 (1.01 - 1.17)	1.03 (0.96 - 1.11)	1.03 (0.95 - 1.11)
	Quintile 3	1.15 (1.08 - 1.24)	1.05 (0.98 - 1.13)	1.05 (0.97 - 1.12)
	Quintile 2	1.25 (1.18 - 1.34)	1.08 (1.01 - 1.16)	1.08 (1.01 - 1.16)
	Quintile 1 (Low)	1.54 (1.45 - 1.64)	1.22 (1.14 - 1.30)	1.21 (1.13 - 1.29)
Rural Place of Residence (Reference = Urban)		0.69 (0.59 - 0.81)	0.76 (0.65 - 0.90)	0.76 (0.65 - 0.90)
English Language at Arrival (Reference = No)		0.88 (0.86 - 0.91)	-	0.90 (0.86 - 0.93)

*Adjusted for sex, age at landing date, country of birth, migrant class, neighbourhood income quintile, rural residence

**Adjusted for knowledge of English, in addition to all variables in the partially adjusted model

Supplemental Table 7 – 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, stratified by age at migration.

First Language Variable n classified (%)		<19 Years at Migration Fully Adjusted** IRR (95%CI)	19+ Years at Migration Fully Adjusted** IRR (95%CI)
Language Genealogy n = 1,863,793 (99.9%)	English	<i>Ref.</i>	<i>Ref.</i>
	Other Germanic Languages	0.72 (0.48 - 1.06)	0.61 (0.46 - 0.81)
	Non-Germanic Indo-European Languages	1.06 (0.94 - 1.18)	0.97 (0.89 - 1.04)
	Non-Indo-European Languages	1.15 (1.01 - 1.30)	1.05 (0.97 - 1.15)
Estimated Acquisition Time n = 1,626,450 (87.3%)	English	<i>Ref.</i>	<i>Ref.</i>
	Category I Language (600-750 class hours)	1.00 (0.86 - 1.16)	0.86 (0.77 - 0.95)
	Category II Language (900 class hours)	1.19 (0.88 - 1.62)	0.74 (0.57 - 0.94)
	Category III Language (1100 class hours)	1.21 (1.07 - 1.38)	1.17 (1.07 - 1.28)
	Category IV Language (2200 class hours)	0.96 (0.82 - 1.13)	0.98 (0.88 - 1.10)
PCM Score n = 1,163,474 (62.4%)	English (Score of 0)	<i>Ref.</i>	<i>Ref.</i>
	Quartile 1	0.96 (0.823 - 1.10)	0.90 (0.82 - 0.99)
	Quartile 2	1.07 (0.91 - 1.27)	1.03 (0.92 - 1.15)
	Quartile 3	1.06 (0.87 - 1.30)	1.02 (0.87 - 1.19)
	Quartile 4	0.75 (0.57 - 0.99)	1.12 (0.91 - 1.39)
ASJP Score n = 1,849,789 (99.2%)	English (Score of 0)	<i>Ref.</i>	<i>Ref.</i>
	Quartile 1	1.06 (0.94 - 1.19)	0.94 (0.87 - 1.03)
	Quartile 2	1.02 (0.90 - 1.16)	0.99 (0.90 - 1.08)
	Quartile 3	0.98 (0.86 - 1.12)	1.04 (0.95 - 1.15)
	Quartile 4	1.29 (1.12 - 1.48)	1.10 (0.99 - 1.21)

IRR = Incidence Rate Ratio; PCM = Parametric Comparison Method; ASJP = Automatic Similarity Judgement Program
 **Adjusted for sex, age at landing date, country of birth, neighbourhood income quintile, rural residence, and knowledge of English

Supplemental Table 8 – 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, stratified by migrant class.

First Language Variable n classified (%)		Economic Migrants Fully Adjusted** IRR (95%CI)	Sponsored Migrants Fully Adjusted** IRR (95%CI)	Refugee Migrants Fully Adjusted** IRR (95%CI)
Language Genealogy n = 1,863,793 (99.9%)	English	<i>Ref.</i>	<i>Ref.</i>	<i>Not Available due to Small Numbers</i>
	Other Germanic Languages	0.76 (0.57 - 1.02)	0.53 (0.36 - 0.78)	
	Non-Germanic Indo-European Languages	1.03 (0.93 - 1.13)	0.92 (0.83 - 1.02)	
	Non-Indo-European Languages	1.09 (0.97 - 1.23)	1.25 (1.10 - 1.43)	
Estimated Acquisition Time n = 1,626,450 (87.3%)	English	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	Category I Language (600-750 class hours)	0.85 (0.74 - 0.97)	0.89 (0.79 - 1.01)	0.84 (0.64 - 1.11)
	Category II Language (900 class hours)	0.83 (0.61 - 1.13)	0.79 (0.57 - 1.10)	0.99 (0.66 - 1.49)
	Category III Language (1100 class hours)	1.14 (1.03 - 1.26)	1.15 (1.00 - 1.31)	1.20 (0.97 - 1.50)
	Category IV Language (2200 class hours)	0.97 (0.85 - 1.10)	1.03 (0.87 - 1.23)	0.91 (0.70 - 1.17)
PCM Score n = 1,163,474 (62.4%)	English (Score of 0)	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	Quartile 1	0.99 (0.88 - 1.12)	0.88 (0.78 - 0.99)	0.84 (0.64 - 1.11)
	Quartile 2	0.99 (0.88 - 1.13)	1.24 (1.02 - 1.51)	0.79 (0.60 - 1.04)
	Quartile 3	0.94 (0.79 - 1.12)	1.08 (0.85 - 1.37)	1.02 (0.74 - 1.40)
	Quartile 4	0.83 (0.67 - 1.02)	1.42 (1.00 - 2.03)	0.94 (0.36 - 2.50)
ASJP Score n = 1,849,789 (99.2%)	English (Score of 0)	<i>Ref.</i>	<i>Ref.</i>	<i>Ref.</i>
	Quartile 1	0.96 (0.86 - 1.06)	0.94 (0.84 - 1.04)	1.05 (0.84 - 1.31)
	Quartile 2	1.07 (0.96 - 1.19)	0.91 (0.80 - 1.04)	1.04 (0.84 - 1.30)
	Quartile 3	1.00 (0.88 - 1.13)	1.22 (1.05 - 1.41)	0.94 (0.75 - 1.16)
	Quartile 4	1.25 (1.09 - 1.43)	1.14 (0.97 - 1.33)	1.11 (0.90 - 1.37)

IRR = Incidence Rate Ratio; PCM = Parametric Comparison Method; ASJP = Automatic Similarity Judgement Program

****Adjusted for sex, age at landing date, country of birth, neighbourhood income quintile, rural residence, and knowledge of English**

Supplemental Table 9 – 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 among people with complete data across all four language variables (n = 1,152,513).

Native Language Variable		Unadjusted IRR (95%CI)	Partially Adjusted* IRR (95%CI)	Fully Adjusted** IRR (95%CI)
Language Genealogy	English	Ref.	Ref.	Ref.
	Other Germanic Languages	0.45 (0.36 - 0.57)	0.63 (0.49 - 0.82)	0.61 (0.48 - 0.79)
	Non-Germanic Indo-European Languages	0.89 (0.85 - 0.94)	0.98 (0.91 - 1.05)	0.93 (0.87 - 1.00)
	Non-Indo-European Languages	0.76 (0.73 - 0.80)	1.18 (1.07 - 1.29)	1.13 (1.03 - 1.24)
Estimated Acquisition Time	English	Ref.	Ref.	Ref.
	Category I Language (approx. 600-750 class hours)	0.88 (0.83 - 0.94)	0.94 (0.87 - 1.02)	0.90 (0.82 - 0.97)
	Category II Language (approx. 900 class hours)	0.40 (0.29 - 0.55)	0.58 (0.42 - 0.80)	0.56 (0.41 - 0.78)
	Category III Language (approx. 1100 class hours)	1.02 (0.97 - 1.07)	1.13 (1.04 - 1.23)	1.09 (1.00 - 1.18)
	Category IV Language (2200 class hours)	0.62 (0.59 - 0.66)	1.07 (0.92 - 1.25)	1.03 (0.88 - 1.20)
PCM Score	English (Score of 0)	Ref.	Ref.	Ref.
	Quartile 1	0.91 (0.86 - 0.96)	0.97 (0.90 - 1.04)	0.93 (0.86 - 1.00)
	Quartile 2	1.05 (0.99 - 1.11)	1.09 (1.00 - 1.20)	1.05 (0.95 - 1.15)
	Quartile 3	0.80 (0.75 - 0.85)	1.07 (0.95 - 1.20)	1.03 (0.91 - 1.16)
	Quartile 4	0.55 (0.51 - 0.58)	1.02 (0.86 - 1.20)	0.96 (0.81 - 1.13)
ASJP Score	English (Score of 0)	Ref.	Ref.	Ref.
	Quartile 1	0.86 (0.82 - 0.91)	0.97 (0.91 - 1.04)	0.93 (0.86 - 1.00)
	Quartile 2	0.98 (0.89 - 1.07)	1.02 (0.91 - 1.14)	0.97 (0.87 - 1.09)
	Quartile 3	0.82 (0.78 - 0.87)	1.14 (1.04 - 1.25)	1.09 (0.99 - 1.21)
	Quartile 4	0.62 (0.57 - 0.66)	1.22 (1.10 - 1.37)	1.16 (1.04 - 1.30)

*Adjusted for sex, age at landing date, country of birth, migrant class, neighbourhood income quintile, rural residence

**Adjusted for knowledge of English, in addition to all variables in the partially adjusted model