Original article

Association between first language and SARS-CoV-2 infection rates, hospitalization, intensive care admissions and death in Finland: a population-based observational cohort study

Ville Holmberg [1, *], Heli Salmi [2], Salla Kattainen [2], Jukka Ollgren [3], Anu Kantele [1], Juulia Pynnönen [1], Asko Järvinen [1], Erik Forsblom [1], Suvi Silén [4, 5], Sanna-Maria Kivivuori [6], Atte Meretoja [7], Johanna Hästbacka [2]

[1] Department of Infectious Diseases, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
[2] Department of Anaesthesiology, Intensive Care and Pain Medicine, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
[3] Department of Health Security, Finnish Institute for Health and Welfare (THL), Finland
[4] Department of Otorhinolaryngology, Head and Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
[5] Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden
[6] Helsinki University Hospital, Helsinki, Finland
[7] General Administration, Helsinki University Hospital, Helsinki, Finland

A R T I C L E   I N F O

A B S T R A C T

Objectives: Motivated by reports of increased risk of coronavirus disease 2019 (COVID-19) in ethnic minorities of high-income countries, we explored whether patients with a foreign first language are at an increased risk of COVID-19 infections, more serious presentations, or worse outcomes.

Methods: In a retrospective observational population-based quality registry study covering a population of 1.7 million, we studied the incidence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), admissions to specialist healthcare and the intensive care unit (ICU), and all-cause case fatality in different language groups between 27th February and 3rd August 2020 in Southern Finland. A first language other than Finnish, Swedish or Sámi served as a surrogate marker for a foreign ethnic background.

Results: In total, 124 240 individuals were tested, and among the 118 300 (95%) whose first language could be determined, 4005 (3.4%) were COVID-19-positive, 623 (0.5%) were admitted to specialized hospitals, and 147 (0.1%) were admitted to the ICU; 254 (0.2%) died. Those with a foreign first language had lower testing rates (348, 95%CI 340–355 versus 758, 95%CI 753–762 per 10 000, p < 0.0001), higher incidence (36, 95%CI 33–38 versus 22, 95%CI 21–23 per 10 000, p < 0.0001), and higher positivity rates (103, 95%CI 96–109 versus 29, 95%CI 28–30 per 1000, p < 0.0001). There was no significant difference in ICU admissions, disease severity at ICU admission, or ICU outcomes. Case fatality by 90 days was 7.7% in domestic cases and 1.2% in those with a foreign first language, explained by demographics (age- and sex-adjusted HR 0.49, 95%CI 0.21–1.15).

Conclusions: The population with a foreign first language was at an increased risk for testing positive for SARS-CoV-2, but when hospitalized they had outcomes similar to those in the native, domestic language population. This suggests that special attention should be paid to the prevention and control of infectious diseases among language minorities. **Ville Holmberg, Clin Microbiol Infect 2022;28:107**

* Corresponding author: Ville Holmberg, Department of Infectious Diseases, University of Helsinki and Helsinki University Hospital, PO Box 340, Fi-00029, Helsinki, Finland.
E-mail address: ville.holmberg@hus.fi (V. Holmberg).

Introduction

Migrants and ethnic minority populations seem to experience a disproportionate burden of the coronavirus disease 2019 (COVID-19) pandemic [1]. Higher incidences have been reported in varying ethnic minorities in several countries, but regarding the outcomes the results have been conflicting [2–8]. Differences in health status, access to health care, housing conditions, family size, professional exposure, use of public transportation, the possibility of teleworking, and economic status have all been identified as factors contributing to increased exposure [1]. Ethnicity can be defined using various indicators such as the country of birth, nationality, migrant status, race, and self-definition. As an advantage of small language groups, in which the language is spoken only in highly restricted areas, an individual's first language can be used as a surrogate marker of ethnicity and used in primary screening to exclude native speakers as non-immigrants. Obviously, this approach does not exclude second-generation immigrants.

Healthcare interventions can address ethnicity-related differences in COVID-19 morbidity and mortality through a better understanding of their causes and consequences. One of the obvious challenges, language, influences all aspects of COVID-19 infection control and treatment, including implementation of public education, counselling and guidance, testing, treatment and contact tracking. At the same time, language problems, when acknowledged, may be overcome with relatively simple measures within the reach of healthcare providers.

We wanted to study whether the first language of a person is associated with the rates of COVID-19 testing, test positivity, hospitalization, ICU care and case fatality. We consider the first language of the patient to be an important aspect in the treatment cascade of COVID-19 patients, as it determines the success of the communication between the patient and the healthcare provider.

Methods

Setting

Finland is a Nordic welfare country with a public universal healthcare system for its residents. Inpatient treatment of COVID-19 patients and intensive care are exclusively provided within the public healthcare system. Testing and treatment of COVID-19 is free of charge for all patients, including migrants and tourists. Undocumented migrants, representing 0.1–0.2% of the population, are entitled to urgent healthcare and in some municipalities, including Helsinki, to all necessary care. Payment is never required prior to receiving health care. The first language might not have been registered for undocumented migrants and persons whose stay in the country is only temporary; thus these groups are probably underrepresented in this study. All residents can be identified with a personal identification code, which enabled us to link healthcare data in this study.

Finland, with a population of 5 525 292 at the end of 2019 [9], has two official national languages, Finnish and Swedish, and a third minority domestic language (Sámi). In the Uusimaa province in Southern Finland, encompassing the capital Helsinki with a population of 1 689 725 [9], Finnish was the first language of 78.2%, Swedish of 7.7% and Sámi for 137 persons (0.0%); foreign domestic languages represented 14.0% of the population ($n = 236,959$).

Testing for COVID-19 was available free of charge during the study period. However, during the timeframe included in the present study, access to testing was dependent on medical assessment and limited mainly to symptomatic persons. Screening or self-referral in walk-in or mobile testing units was not yet available. Indications for testing reflected changes in testing capacity and national policies.

Study population

This retrospective observational population-based quality registry study included all individuals in the capital province of Finland (Uusimaa) tested for COVID-19 between 27th February and 3rd August 2020 by the Helsinki University Hospital laboratory services. All patients admitted to specialist healthcare in any of the 22 hospitals in the capital province diagnosed during this period were included in the analyses of in-hospital treatment. Previously institutionalized and/or dependent patients were mainly treated in primary-care institutions and hospitals and were not included in the analyses of specialized in-hospital care.

Retrieval of data

We collected laboratory and clinical data from the electronic patient records of the Helsinki University Hospital district into a COVID-19 quality registry. Data documentation included sex, age, language, duration of hospital stay, date of death, basic laboratory test results and information from the ICU quality registry. Data on patient and ICU admission characteristics were retrieved from the Finnish Intensive Care Consortium (FICC) database and from the HUCH electronic patient data management systems (Miranda, PICIS, WebLab, Apotti). Based on these data, the Charlson comorbidity index (CCI) and sequential organ failure assessment scores (SOFA) during the first 24 h of ICU treatment, which were sporadically missing from the FICC database, were calculated. If the patient was transferred to another ICU the total length of the ICU stay was included as one admission. If the patient was readmitted to intensive care during the same hospital admission, only the first admission was included. Thus, a single patient was included only once. All-cause case-fatality data were automatically linked from the National Causes of Death Registry into the hospital's electronic patient records, allowing comprehensive post-hospital stay follow-up of possible fatal cases among all patients.

Statistical analyses

The p-values for categorical variables were calculated with two-sided Fisher's exact test and for age with Wilcoxon rank-sum (Mann–Whitney) test. The ICU data were first analysed with univariable models, and variables with $p < 0.2$ were included in the multivariable model, where a stepwise backward logistic regression model was used. For testing rates, positivity rates and incidence rates, 95% Poisson confidence intervals were calculated using the Stata 16.1 program. Mixed-effects (or fixed-effects) logistic regression was used to calculate the odds ratios for positivity, hospitalization and ICU admission. Ethnicity was taken into account either as a fixed effect or as a random effect, depending on the context. Survival during the first 90 days after the positive test date was analysed by the Kaplan–Meier estimator or the Cox proportional hazards model. In the Cox model, the potential clustering by different ethnicity groups was taken into account using robust standard error estimates. Age-adjusted risk of death differences were analysed using the Cox model or the Kaplan–Meier method with a stratified log-rank test. Tests were two-tailed. Analyses were carried out with IBM SPSS Statistics version 27 for Mac.

Institutional review and patient consent

The quality registry was institutionally approved without requirement for patient consent (approvals HUS/1049/2020/§4 and

Results

Characteristics of the cohorts

Altogether, 124 240 persons were tested for SARS-CoV-2 during this study, i.e. the first wave of the pandemic. The first language was registered for 118 300 patients (95.2%) (Fig. 1). The distribution of sex, age and 90-day case fatality in the four cohorts is shown in Table 1. Overall, 7.0% of tested patients, 21.7% of positive patients, 21.7% of patients admitted to specialist hospitals, and 25.9% of patients admitted to the ICU had a foreign (other than Finnish, Swedish or Sámi) first language (Fig. 1).

Testing rate and positivity

Those with a foreign first language had lower testing rates (348, 95%CI 340–355 versus 758, 95%CI 753–762 per 10 000, χ^2, p < 0.0001), higher incidence (36, 95%CI 33–38 versus 22, 95%CI 21–23 per 10 000, χ^2, p < 0.0001), and higher positivity rates (103, 95%CI 96–109 versus 29, 95%CI 28–30 per 1000, χ^2, p < 0.0001). Similar differences were seen in all age groups, suggesting that they are not explained by the younger age profile of the migrant population compared to the native population (Table 2). However, the findings were not uniform in all language groups. When adjusted to population size, the highest positivity rates were among native speakers of Somali and Albanian. Compared to domestic languages, higher proportions of positives among those tested were found in native speakers of Russian, Estonian, Arabic, Somali, English, Kurdish, Albanian and Turkish.

Odds ratios for COVID-19 and hospitalization

During the first wave, testing was focused on symptomatic patients and those needing in-hospital care, which was reflected in the increased odds ratio for testing positive for males and those older than 75 years (Table 3). Persons with a foreign first language were more likely to test positive, with the highest odds ratios when the first language was Somali, Albanian or Turkish. The odds ratio for hospitalization and need for intensive care were analysed in people under the age of 75 years, as older patients were often treated in primary care and were not included in our data. The odds ratio for hospitalization in specialist health care was higher for those with a foreign first language compared to the native population, but this is explained mainly by lower testing rates of persons with a foreign first language (Table 3). When the foreign language group was divided into specific languages, a distinct heterogeneity between languages was seen.

ICU outcomes

Among the 147 patients admitted to ICUs, 38 (26%) had a foreign first language. They were younger and had a lower Charlson comorbidity index than those with a domestic first language (Table 4). The unadjusted SOFA scores at 24 h from admission, the use of invasive mechanical ventilation, length of stay in the ICU, or case fatality did not differ between patients with foreign or domestic first languages (Table 4a). When comparing d-90 survivors and d-90 non-survivors in the ICU population in univariable analysis, age (p < 0.001), sex (p 0.007), CCI (<0.001), and SOFA score from the first 24 hours in the ICU (p 0.03) were included in the multivariable analysis, together with the language group (Table 4b). In the multivariable analysis, only CCI was significantly and independently associated with d-90 mortality (OR 1.697, 95%CI 1.30–2.21, p < 0.001).
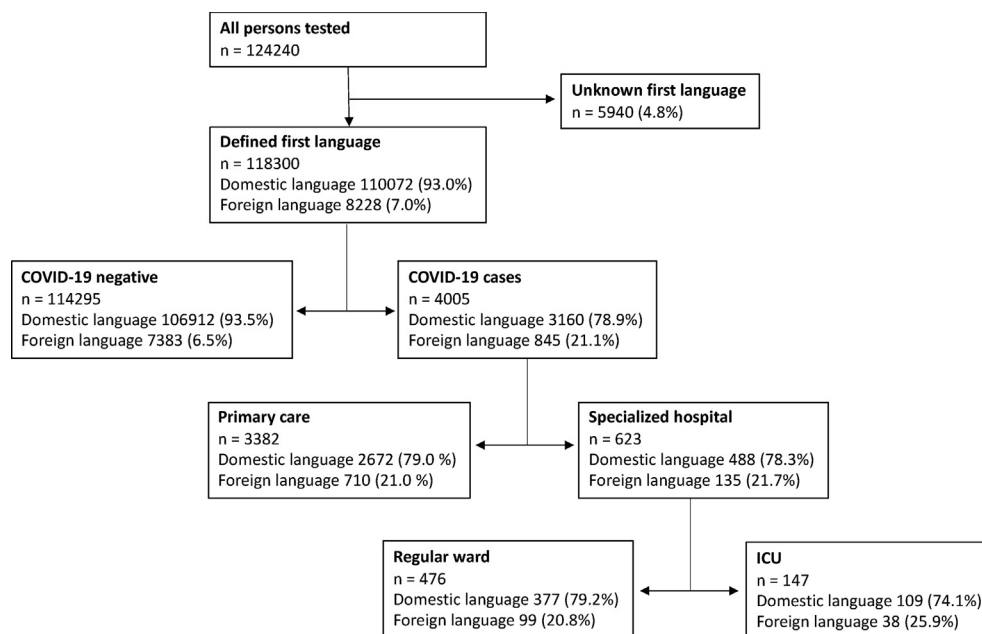
Fig. 1. Flow chart of the patients included in the study. The four cohorts analysed in the study were (1) coronavirus disease 2019 (COVID-19) negatives, (2) COVID-19 cases treated in primary care, nursing homes or at home, (3) COVID-19 cases treated in regular wards of specialized hospitals, and (4) COVID-19 cases treated at intensive care units (ICUs).

Table 1
Characteristics of the patients in each cohort. Domestic language is defined as Finnish, Swedish or Sámi spoken as the first language. All other languages are defined as foreign

	Domestic language	Foreign language	p-value
COVID-19 negative			
n (%)	106 912 (97.1)	7383 (89.7)	<0.001
Female sex, *n* (%)	66 902 (62.6)	4167 (56.4)	<0.001
Age (years), median (IQR)	38 (23–58)	32 (20–42)	<0.001
COVID-19, primary care			
n (%)	2672 (2.4)	710 (8.6)	<0.001
Female sex, *n* (%)	1520 (56.9)	341 (48.0)	<0.001
Age (years), median (IQR)	44 (29–60)	34 (24–45)	<0.001
D-90 case fatality, % (95%CI)	166 (6.2)	2 (0.3)	<0.001
COVID-19, specialist hospital ward, without ICU			
n (%)	377 (0.3)	99 (1.2)	<0.001
Female sex, *n* (%)	191 (50.7)	52 (52.5)	0.74
Age (years), median (IQR)	62 (50–70)	49 (36–57)	<0.001
D-90 case fatality, % (95%CI)	58 (15.4)	2 (2.0)	<0.001
COVID-19, specialist hospital ICU			
n (%)	109 (0.1)	38 (0.5)	<0.001
Female sex, *n* (%)	40 (36.7)	17 (44.7)	0.44
Age (years), median (IQR)	61 (51–70)	51 (41–57)	<0.001
D-90 Case fatality, % (95%CI)	20 (18.3)	6 (15.8)	0.81

COVID-19, coronavirus disease 2019.

Table 2
Rates of testing and positivity in different language groups

First language	Population	Tested (*n*)	Positives (*n*)	Testing rate per 10 000 population (95%CI)	Positives per 10 000 population (95%CI)	Positives per 1000 tested persons (95%CI)
Domestic	1 452 766	110 072	3160	758 (753–762)	22 (21–23)	29 (28–30)
0–19	306 590	21 662	301	707 (697–716)	10 (9–11)	14 (12–15)
20–39	385 133	35 887	909	932 (923–941)	24 (22–25)	25 (24–27)
40–59	384 687	27 056	1010	703 (695–711)	26 (25–28)	37 (35–40)
60–74	256 422	14 462	453	564 (555–573)	18 (16–19)	31 (28–34)
75+	119 934	11 003	487	917 (901–934)	41 (37–44)	44 (40–48)
Foreign	236 520	8228	845	348 (340–355)	36 (33–38)	103 (96–109)
0–19	59 992	1953	124	326 (311–340)	21 (17–24)	63 (53–74)
20–39	98 337	3641	371	370 (358–382)	38 (34–42)	102 (92–112)
40–59	62 071	2107	289	339 (325–354)	47 (41–52)	137 (122–152)
60–74	13 722	419	46	305 (277–334)	34 (24–43)	110 (80–140)
75+	2398	108	15	450 (367–533)	63 (31–94)	139 (74–204)
Finnish	1 321 800	101 796	2944	770 (766–775)	22 (21–23)	29 (28–30)
Swedish	130 829	8270	216	632 (619–645)	17 (14–19)	26 (23–30)
Other language	66 286	2531	261	382 (367–396)	39 (35–44)	103 (91–115)
Russian	40 218	1375	74	342 (324–360)	18 (14–23)	54 (42–66)
Estonian	33 458	946	78	283 (265–301)	23 (18–28)	82 (65–100)
Arabic	17 641	728	55	413 (383–442)	31 (23–39)	76 (56–95)
Somali	17 434	950	224	545 (511–579)	128 (112–145)	236 (209–263)
English	13 840	502	37	363 (332–394)	27 (18–35)	74 (51–97)
Chinese	8556	113	4	132 (108–156)	5 (0–9)	35 (1–69)
Kurdish	7852	273	28	348 (307–388)	36 (22–49)	103 (67–139)
Albanian	7562	262	50	346 (305–388)	66 (48–84)	191 (143–238)
Persian	7252	146	7	201 (169–234)	10 (3–17)	48 (13–83)
Vietnamese	6442	105	4	163 (132–194)	6 (0–12)	38 (1–75)
Spanish	5232	172	5	329 (280–377)	10 (1–18)	29 (4–54)
Turkish	4747	125	18	263 (218–309)	38 (20–55)	144 (82–206)
Sámi	137	6	0	438 (95–781)	0 (0–0)	0 (0–0)

Case fatality

Two hundred and fifty-four deaths within 90 days from the first positive test were recorded; in 244 cases the person's first language was a domestic one (7.7% overall case-fatality rate) and in ten (1.2%) it was a foreign one (Table 1). Of the deceased, 207 were older than 75 years, and only one of them had a first language that was foreign. The overall unadjusted 90-day case fatality of patients admitted to specialist healthcare hospitals was 78/486 (16.0%) among speakers of Finnish, Swedish or Sámi and 8/137 (5.8%) among speakers of a foreign language, with the latter group being more than 10 years younger (Table 1). The age- and sex-adjusted hazard ratio for 90-day case fatality was 0.49 (95%CI 0.21–1.15) in patients with foreign first languages (Cox regression, adjusted for clusters with a robust standard error estimator). Kaplan–Meier failure estimates showed similar case fatality in those with domestic and foreign first languages in individuals younger than 65 years of age (Fig. 2). In those older than 65 years of age, the case fatality was lower among individuals with a foreign first language, but the 95% confidence intervals were overlapping.

Discussion

Individuals with a foreign first language generally had a higher incidence, a higher proportion of positive test results, and a lower rate of testing than the domestic language population. The risk for

Table 3
Odds ratios (ORs) with coronavirus disease 2019 (COVID-19) positivity and hospitalization as outcomes. Patients hospitalized and intensive care unit (ICU) patients younger than 75 years of age were included in the analyses

	OR for COVID-19 among all tested (95%CI)	OR for hospitalization among COVID-19-infected (95%CI)	OR for ICU among hospitalized (95%CI)
Sex			
Female	1	1	1
Male	1.56 (1.39–1.75)	1.32 (0.99–1.76)	1.40 (0.91–2.14)
Age			
0–19	0.41 (0.35–0.48)	0.07 (0.02–0.22)	0.25 (0.03–2.06)
20–39	0.65 (0.58–0.72)	0.38 (0.27–0.54)	0.55 (0.28–1.10)
40–59	1	1	1
60–74	0.79 (0.68–0.92)	1.27 (0.78–2.05)	1.43 (0.91–2.14)
75+	1.43 (1.25–1.64)	—	—
First language			
Domestic	1	1	1
Foreign	3.53 (1.42–8.70)	1.74 (1.37–2.20)	2.60 (0.61–11.0)
Finnish	1	1	1
Swedish	0.89 (0.77–1.02)	0.72 (0.43–1.20)	—
Albanian	9.47 (6.90–12.9)	1.22 (0.46–3.22)	0.88 (0.09–8.64)
Arabic	2.86 (2.15–3.79)	2.75 (1.34–5.65)	0.67 (0.13.3.23)
English	2.66 (1.90–3.74)	2.29 (0.95–5.51)	0.57 (0.07–4.94)
Chinese	1.44 (0.53–3.93)	3.47 (0.30–41.0)	—
Kurdish	4.04 (2.71–6.00)	2.25 (0.80–6.35)	0.88 (0.09–8.52)
Persian	1.87 (0.87–4.01)	2.25 (0.25–20.5)	—
Somali	12.6 (10.8–14.8)	1.88 (1.25–2.83)	1.73 (0.79–3.74)
Russian	2.07 (1.63–2.63)	0.97 (0.46–2.03)	0.39 (0.05–3.23)
Estonian	3.08 (2.43–3.90)	0.71 (0.33–1.51)	4.80 (1.09–21.1)
Vietnamese	1.53 (0.56–4.16)	2.77 (0.25–30.4)	—
Turkish	5.72 (3.44–9.51)	4.21 (1.28–13.9)	1.28 (0.12–13.4)
Spanish	0.95 (0.39–2.31)	1	—
Other languages	4.13 (3.61–4.73)	2.04 (1.40–2.96)	0.89 (0.41–1.92)

Table 4
Characteristics of the 147 patients treated at intensive care units (ICUs). Data are presented as comparison between individuals with a domestic and foreign first language (A) and as a comparison based on 90-day survival

(A)	Domestic language	Foreign language	p-value
n	109	38	
Age, years, median (IQR)	61 (51.0–69.5)	50.5 (41.3–57.0)	<0.001
Female sex, *n* (%)	40 (36.7)	17 (44.7)	0.38
BMI, median (IQR)	29.4 (26.7–33.3)	29.2 (26.6–34.3)	0.73
Charlson comorbidity index, median (IQR)	2 (1–3)	1 (0–2)	<0.001
1st 24-hour SOFA, median (IQR)	6 (3–8)	7 (5–8)	0.74
Mechanical ventilation, *n* (%)	69 (63.3)	26 (68.4)	0.57
Duration of mechanical ventilation (days), median (IQR)	5 (0–14)	6 (0–15)	0.51
ICU length of stay (days), median (IQR)	12.3 (4.8–17.4)	13.5 (5.7–23.1)	0.45
D-90 case fatality, *n* (%)	20 (18.3)	6 (15.8)	0.72

(B)	D-90 survivors	D-90 non-survivors	p-value
First language			
Domestic (%)	89 (73.6)	20 (76.9)	0.72
Foreign (%)	32 (26.4)	6 (23.1)	
Age, years, median (IQR)	57 (48–65)	70.5 (57–77)	<0.001
Sex			
Male, *n* (%)	70 (57.9)	20 (76.9)	0.07
Female, *n* (%)	51 (42.1)	6 (23.1)	
Charlson comorbidity index, median (IQR)	2 (1–3)	3 (2–5)	<0.001
1st 24 h SOFA, median (IQR)	6 (3–8)	7 (5.8–9.3)	0.03

IQR, interquartile range; BMI, body mass index; SOFA, sequential organ failure assessment score.

hospitalization was increased in those with a foreign first language, which can be explained by a lower testing rate which resulted in only the more severe cases being detected. The odds ratio for intensive care admission, disease severity at ICU admission, and ICU outcomes did not differ between language groups, suggesting that there were no differences in disease presentation, and no imbalance in admission criteria, choice of treatment modality, or outcomes. After adjusting for age and sex, there was no significant difference in case fatality between patients with foreign and domestic first languages.

Previous studies [6–8] on the COVID-19 case-fatality risk for ethnic minorities have been conflicting. Access to high-quality health care might partially explain why case-fatality risk is increased in some settings but not in others. Our results are in line with previous studies reporting an increased infection rate for COVID-19 in different ethnic minorities in different healthcare systems and varying epidemic circumstances [1]. It is noteworthy that our results come from a setting where access to COVID-19 testing and high-level public health care at low or no cost were technically equal in the population. Yet, testing was likely affected
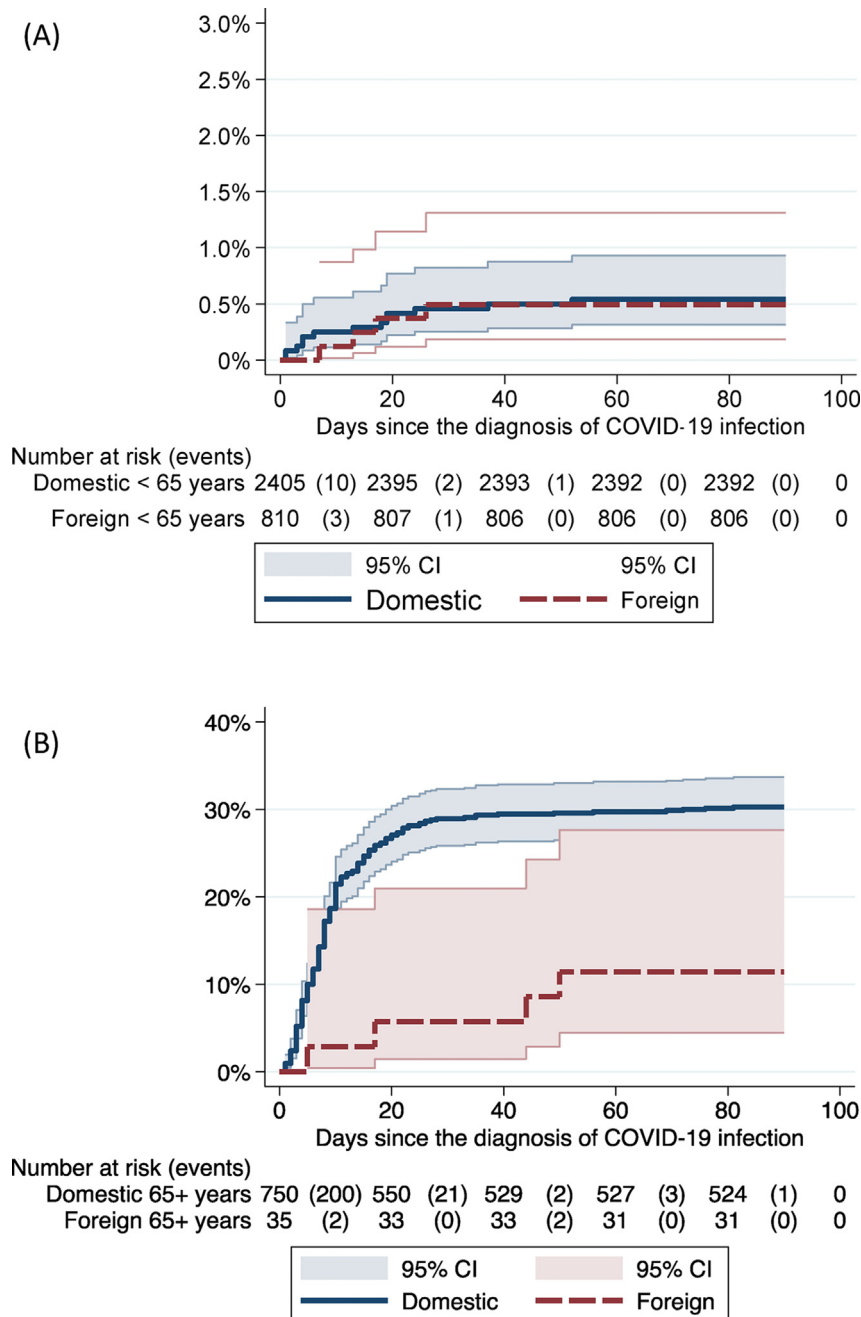
(A)

Number at risk (events)
Domestic < 65 years 2405 (10)   2395 (2)   2393 (1)   2392 (0)   2392 (0)   0
Foreign < 65 years   810 (3)   807 (1)   806 (0)   806 (0)   806 (0)   0

(B)

Number at risk (events)
Domestic 65+ years 750 (200)   550 (21)   529 (2)   527 (3)   524 (1)   0
Foreign 65+ years   35 (2)   33 (0)   33 (2)   31 (0)   31 (0)   0

Fig. 2. Kaplan−Maier failure estimates comparing case fatality in individuals with domestic and foreign first languages. Plot A describes the data for individuals younger than 65 years of age, and plot B for those older than 65 years.

by practical problems such as poor domestic language skills or by lower test-seeking behaviour, or healthcare workers might have been more prone to direct persons speaking domestic languages to testing. Foreign language speakers were overrepresented in a previous study on prehospital COVID-19 patients in Helsinki, which may indicate delayed seeking of health care [10].

An increased attack rate in ethnic minorities during epidemics is not limited to the most recent pandemic of SARS-CoV-2. Very similar findings can be traced back to the H1N1 pandemic in 1918 [11], and to the previous H1N1v pandemic in 2009 [12,13]. Quite analogously to our results, these studies also suggest that the problem does not lie in a genetic or biomedical susceptibility for a more severe illness but in poorer access to health care and increased risk of contracting

the virus due to diverse socioeconomic factors. Previous research may also provide encouraging tools for overcoming these differences, as they suggest that ethnic minorities may be more inclined to be vaccinated if offered the vaccine [14].

A major strength of our study was that the register data was of high quality since they were population-based [15]. In Finland, the self-reported first language of residents is registered in the Population Information System and thereby offers a useful indicator of ethnicity. However, the registered first language says more about the ancestry of the individual than their actual language skills, since many second-generation migrants may, often in addition to their own first language, speak fluent Finnish despite being registered as a speaker of another first language. Assuming that migrant

background influences the ability to use and access healthcare, including second-generation migrants is not expected to strengthen these findings, but rather to dilute this effect. Thus, the fact that we found differences, even when second-generation migrants were included, makes it even more likely that a true effect exists. Information about the first language was available for 95% of the tested individuals. Based on the names of the individuals with missing language data, we assume that most of them had a migrant background or were temporary visitors or tourists in Finland, which was why their first language was not registered. One limitation of the study was the lack of information about co-morbidities and the socioeconomic status of the outpatients, which could not be included in the multivariable analyses.

The study shows that language minorities have an increased attack rate for COVID-19, but after access to quality care the outcomes do not differ from those of the general population. Potential language barriers concerning situation awareness, access to testing and more efficient contact tracing should be addressed. This study did not provide data about the mechanisms underlying the increased susceptibility to COVID-19 in language minorities. Possible risk factors may include more crowded housing, larger family sizes, professional exposure, use of public transport, and lack of possibilities for teleworking [1]. The motivation for testing and adherence to isolation and quarantine recommendations might also be affected by access to social benefits, employment status and ability to understand instructions given in a foreign language, and even attitudes within the given group. Qualitative interviews with persons from language minorities could provide better understanding of the reasons for the increased incidences.

Studying ethnicity as a risk factor for infectious diseases includes ethical questions which need to be addressed [1]. The results must be interpreted and communicated with caution to avoid stigmatization of vulnerable groups. Still, experiences from HIV and tuberculosis research have shown that acknowledging risks specific to certain vulnerable groups can help in tailoring public health policies and programmes to reduce the disease burden in these populations [16,17]. In addition, using a person's first language as a risk factor may be less stigmatizing than relying on other markers of ethnicity, as language involves a practical aspect which can be addressed within the healthcare system. Differences in susceptibility to COVID-19 based on language group should encourage the implementation of concrete measures targeted at populations at risk, such as providing public guidance about infection control measures in relevant languages and using interpreters for contact tracing and patient information.

Author contributions

VH, HS, JH and AK conceptualized the study and its methodology. JO performed the statistical analyses. VH, HS and JH were responsible for writing the original draft. All authors reviewed and edited the manuscript and approved its final version.

Transparency declaration

Acknowledgements

References

[1] Melchior M, Desgrees du Lou A, Gosselin A, Datta GD, Carabali M, Merckx J, et al. Migrant status, ethnicity and COVID-19: more accurate European data are greatly needed. Clin Microbiol Infect 2021;27:160–2.

[2] Rentsch CT, Kidwai-Khan F, Tate JP, Park LS, King Jr JT, Skanderson M, et al. Patterns of COVID-19 testing and mortality by race and ethnicity among United States veterans: a nationwide cohort study. PLoS Med 2020;17: e1003379.

[3] Hansson E, Albin M, Rasmussen M, Jakobsson K. Large differences in excess mortality in March–May 2020 by country of birth in Sweden. Lakartidningen 2020;117:20113.

[4] Baqui P, Bica I, Marra V, Ercole A, van der Schaar M. Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. Lancet Glob Health 2020;8:e1018–26.

[5] Price-Haywood EG, Burton J, Fort D, Seoane L. Hospitalization and mortality among black patients and white patients with COVID-19. N Engl J Med 2020;382:2534–43.

[6] Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020;584:430–6.

[7] Mathur R, Rentsch CT, Morton CE, Hulme WJ, Schultze A, MacKenna B, et al. Ethnic differences in SARS-CoV-2 infection and COVID-19-related hospitalisation, intensive care unit admission, and death in 17 million adults in England: an observational cohort study using the OpenSAFELY platform. Lancet 2021;397:1711–24.

[8] Ogedegbe G, Ravenell J, Adhikari S, Butler M, Cook T, Francois F, et al. Assessment of racial/ethnic disparities in hospitalization and mortality in patients with COVID-19 in New York City. JAMA Netw Open 2020;3: e2026881.

[9] Official Statistics of Finland. Helsinki, Statistics Finland [referred 31.12.2019]. ISSN=1797-5379, http://www.stat.fi/til/vaerak/.

[10] Kuisma M, Harve-Rytsala H, Pirneskoski J, Boyd J, Laaperi M, Salo A, et al. Prehospital characteristics of COVID-19 patients in Helsinki—experience of the first wave of the pandemic. Scand J Trauma Resusc Emerg Med 2021;29: 95.

[11] Krishnan L, Ogunwole SM, Cooper LA. Historical insights on coronavirus disease 2019 (COVID-19), the 1918 influenza pandemic, and racial disparities: illuminating a path forward. Ann Intern Med 2020;173:474–81.

[12] Mertz D, Kim TH, Johnstone J, Lam PP, Science M, Kuster SP, et al. Populations at risk for severe or complicated influenza illness: systematic review and meta-analysis. BMJ 2013;347:f5061.

[13] Quinn SC, Kumar S, Freimuth VS, Musa D, Casteneda-Angarita N, Kidwell K. Racial disparities in exposure, susceptibility, and access to health care in the US H1N1 influenza pandemic. Am J Publ Health 2011;101:285–93.

[14] Bish A, Yardley L, Nicoll A, Michie S. Factors associated with uptake of vaccination against pandemic influenza: a systematic review. Vaccine 2011;29:6472–84.

[15] Forsblom E, Silén S, Kortela E, Ahava M, Kreivi H, Holmberg V, et al. Male predominance in disease severity and mortality in a low COVID-19 epidemic and low case-fatality area—a population-based registry study. Infect Dis 2021;53:789–99.

[16] Holmberg V, Soini H, Kivela P, Ollgren J, Ristola M. Epidemiology and outcome of HIV patients in Finland co-infected with tuberculosis 1998–2015. BMC Infect Dis 2019;19:264.

[17] Sullivan PS, Satcher Johnson A, Pembleton ES, Stephenson R, Justice AC, Althoff KN, et al. Epidemiology of HIV in the USA: epidemic burden, inequities, contexts, and responses. Lancet 2021;397:1095–106.