Journal of Clinical and Translational Science

www.cambridge.org/cts

Interrogating an ICD-coded electronic health records database to characterize the epidemiology of prosopagnosia

Christina Pressl[1], Caroline S. Jiang[2], Joel Correa da Rosa[3,5], Maximilian Friedrich[4], Roger Vaughan[2,5], Winrich A. Freiwald[1,*] and Jonathan N. Tobin[5,6,*]

[1]Laboratory of Neural Systems, The Rockefeller University, New York, NY, USA; [2]Department of Biostatistics, The Rockefeller University, New York, NY, USA; [3]Department of Explorative Biology, LEO Pharma, Ballerup, Denmark; [4]Department of Neurology, University Hospital Würzburg, Würzburg, Germany; [5]Center for Clinical and Translational Science, The Rockefeller University, New York, NY, USA and [6]Clinical Directors Network (CDN), New York, NY, USA

Abstract

Introduction: Recognition of faces of family members, friends, and colleagues is an important skill essential for everyday life. Individuals affected by prosopagnosia (face blindness) have difficulty recognizing familiar individuals. The prevalence of prosopagnosia has been estimated to be as high as 3%. Prosopagnosia can severely impact the quality of life of those affected, and it has been suggested to co-occur with conditions such as depression and anxiety. *Methods:* To determine real-world diagnostic frequency of prosopagnosia and the spectrum of its comorbidities, we utilized a large database of more than 7.5 million de-identified electronic health records (EHRs) from patients who received care at major academic health centers and Federally Qualified Health Centers in New York City. We designed a computable phenotype to search the database for diagnosed cases of prosopagnosia, revealing a total of n = 902 cases. In addition, data from a randomly sampled matched control population (n = 100,973) were drawn from the database for comparative analyses to study the condition's comorbidity landscape. Diagnostic frequency of prosopagnosia, epidemiological characteristics, and comorbidity landscape were assessed. *Results:* We observed prosopagnosia diagnoses at a rate of 0.012% (12 per 100,000 individuals). We discovered elevated frequency of prosopagnosia diagnosis for individuals who carried certain comorbid conditions, such as personality disorder, depression, epilepsy, and anxiety. Moreover, prosopagnosia diagnoses increased with the number of comorbid conditions. *Conclusions:* Results from this study show a wide range of comorbidities and suggest that prosopagnosia is vastly underdiagnosed. Findings imply important clinical consequences for the diagnosis and management of prosopagnosia as well as its comorbid conditions.

Translational Research, Design and Analysis Research Article

Cite this article: Pressl C, Jiang CS, Correa da Rosa J, Friedrich M, Vaughan R, Freiwald WA, and Tobin JN. Interrogating an ICD-coded electronic health records database to characterize the epidemiology of prosopagnosia. *Journal of Clinical and Translational Science* **5**: e11, 1–9. doi: 10.1017/cts.2020.497

Received: 18 March 2020
Revised: 29 May 2020
Accepted: 9 June 2020

Keywords:
Electronic health records research; epidemiology; big data; rare diseases; prosopagnosia

Address for correspondence:
C. Pressl, MD, MS, Laboratory of Neural Systems, Laboratory of Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA. Tel: +1 212-327-7959. Email: cpressl@rockefeller.edu

*These authors share senior authorship.

Introduction

Prosopagnosic individuals struggle to recognize familiar persons by their face. Face blindness was first described by the neurologist Joachim Bodamer, who, in 1947, reported of three patients experiencing face recognition difficulties after suffering from brain damage [1]. Over the years, substantial progress has been made in understanding the neuronal networks and cognitive functions that underlie face perception [2–6]. Today we know of two major forms of face blindness: (1) acquired prosopagnosia occurring after brain damage [7–9] and (2) developmental prosopagnosia (DP) a form of face recognition impairment that is present where there has been no preceding comorbid or traumatic event [10]. In DP, the level of face recognition impairment can vary and difficulties range from an inability to recognize familiar faces when seen out of context, to the inability to recognize faces of family members [11–13].

While it has been estimated that DP may affect up to 3% of the general population [14–18], an objective assessment of the overall rate of diagnoses among a large cohort of patients has not been performed. Diagnosis of prosopagnosia can be difficult and there is no clinical gold standard available. Despite this fact, insights into the present diagnostic frequency provide crucial information regarding the current level of care for prosopagnosic individuals.

Living with face blindness can have severe impact on the affected individual's quality of life and result in elevated levels of anxiety and depression [19,20]. Emotional consequences may be socially debilitating, leading to social isolation, and can even result in decreased occupational competitive ability [20]. These adverse consequences impair psychosocial health and interpersonal relationships of people living with prosopagnosia, pose considerable negative impact on the affected individuals, their families, as well as the health care system and society as a whole. Nevertheless, to the authors' best knowledge, a systematic large-scale assessment of diagnostic

frequency, demographic and epidemiologic characteristics, as well as of the comorbidity landscape of prosopagnosia, has not been completed to date.

To address these evidence gaps, we queried a large, electronic health record (EHR)-containing database. We searched this database for diagnosed cases of prosopagnosia with the overall hypothesis that systematic investigation and comparative analyses would reveal insight into diagnostic frequency and allow for inferences on the epidemiology of prosopagnosia. We extracted data from the Patient-Centered Outcomes Research Institute (PCORI)-funded New York City Clinical Data Research Network (NYC-CDRN) [21], which, at the time of study, contained records from more than 7.5 million patients who had received some or all of their care from 12 NYC academic health centers and three Federally Qualified Health Centers networks. EHRs contain comprehensive longitudinal patient-level data including demographics, patient visits, clinical conditions and diagnoses, laboratory results, medications, and clinical procedures. Diagnoses were coded following version 9 and 10 of the International Classification of Diseases coding system (ICD-9-CM and ICD-10-CM). For the purpose of this study, we developed a computable phenotype for prosopagnosia, utilizing this ICD coding system (see Table 1). Since October 2015, the updated version of the ICD-9 coding system was implemented (ICD-10) and while, at time of data query, ICD-10 had already been in effect, no cases carrying the ICD10 main condition of interest (mCOI) could be identified. Therefore, data reported in this study also exclusively contain ICD-9 coded comorbid conditions of interest (cCOI). While the ICD-10 system provides an individual diagnosis code for prosopagnosia, in the ICD-9 system, the diagnosis code for prosopagnosia may also be applied when simultanagnosia is diagnosed. Simultanagnosia, along with optic ataxia and oculomotor apraxia, define Bálint's syndrome, a rare neurological condition in which patients cannot perceive more than one visually presented object at a time. Therefore, it is possible that the present case-cohort may contain some "misclassifications" of individuals diagnosed with simultanagnosia and not prosopagnosia or individuals suffering from both conditions. While Bálint's syndrome has been defined as a rare disease (https://www.orpha.net, ORPHA: 363746), prevalence rates of Balint's syndrome were not available at the time of this report. Moreover, prevalence rates for simultanagnosia were not available. Nevertheless, prevalence of simultanagnosia is expected to be low and cases of individuals suffering from simultanagnosia and prosopagnosia at the same time have been reported in the past [22]. Since the ICD-9 coding system provides no isolated code for prosopagnosia, possible inclusion of cases of simultanagnosia as well as, or instead of, prosopagnosia had to be accepted as a limitation in this study. Henceforth, it is inferred in this report that data extracted from cases detected reflect data from prosopagnosic individuals.

Guided by previously established recommendations to define prosopagnosia [23–25], we designed the computable phenotype by creating a list of inclusion and exclusion codes. This strategy allowed us to search for individuals carrying the mCOI and exclude individuals whose face recognition difficulties could be caused by other conditions such as autism [23], other cognitive and mental disorders, or conditions of eyes and adnexa. Utilizing the computable phenotype, we searched for diagnosed prosopagnosia cases and interrogated the database for frequencies and distributions of several selected cCOI. We utilized the extracted data to study diagnostic frequency, gain novel insight into epidemiologic and demographic characteristics of the case-cohort, and to perform

Table 1. *Prosopagnosia computable phenotype*

Inclusion diagnoses	ICD9-CM code
Main condition of interest (mCOI)	
Psychophysical visual disturbances (includes prosopagnosia)	368.16
Common coinciding conditions of interest (cCOI)	
Localization-related (focal) (partial) epilepsy and epileptic syndromes with complex partial seizures (focal temporal epilepsy)	345.4
Malignant neoplasm of temporal lobe	191.2
Cerebral artery occlusion	
Ischemic stroke	434.91
Embolic	434.11
Thrombotic	434.01
Dementia	290
Old age	
Alzheimer's disease	
Vascular	
Frontotemporal dementia other	331.1
Trauma	
Post-concussion syndrome	310.2
Concussion with unconsciousness < 30min	850.11
Concussion with unconsciousness > 30min	850.22
Herpesviral encephalitis other	058.2
Herpetic meningoencephalitis	054.3
Major depressive disorder	
Single episodes	296.20–296.26
Recurrent episodes	296.30–296.36
Anxiety disorder	
Atypical	300.00
Generalized	300.02
Associated with physical disorder	293.84
Panic disorder	300.01
Neurasthenia	300.05
Adjustment disorder	309.0
	309.24
	309.28
	309.29
Posttraumatic stress disorder	309.81
Procedures	**ICD9-CM PCS code**
Hemispherectomy	01.52
Lobectomy of brain	01.53
Exclusion diagnoses	**ICD9-CM code**
Disorders of the eye and adnexa	360–367 369–379

(*Continued*)

https://doi.org/10.1017/cts.2020.497 Published online by Cambridge University Press

Table 1. (*Continued*)

Exclusion diagnoses	ICD9-CM code
This group contains diagnoses for the following conditions: Pervasive developmental disorders, autistic disorder, childhood disintegrative disorder, other specified pervasive developmental disorders, unspecified pervasive developmental disorder	299.*
This group contains diagnoses for the following conditions: Alteration of consciousness, coma, transient alteration of awareness, persistent vegetative state, other alteration of consciousness	780.0*
Mild intellectual disabilities	317
Other specified intellectual disabilities	318
Unspecified intellectual disabilities	319
Senile dementia	290.0
Memory loss	780.93
	ICD10-CM code
Mental, behavioral and neurodevelopmental disorders	F01-F99
Diseases of the eye and adnexa	H00-H59
Cognitive deficits following cerebral infarction	I69.31
Alzheimer's disease	G30
Other symptoms and signs involving cognitive functions and awareness	R41

PCS, Procedure Classification System

comparative analyses by contrasting case data to data derived from a randomly sampled matched control-cohort to study the comorbidity landscape of prosopagnosia. This study sheds new light on the diagnostic rate of prosopagnosia in the NYC area and provides novel important insight into the frequency of its comorbid conditions.

Methods

De-identified data were extracted from a large NYC-based EHRs longitudinal database, the PCORnet-funded NYC-CDRN [21]. At the time of data extraction, the database contained ICD coded EHRs from 7,522,133 million de-duplicated individuals who received medical care in the NYC area between 2007 and 2015. The de-duplication process requires all sites contributing to the CDRN database to code each patient with a unique identifier, called a "proxyID." For each quarterly data refresh, CDRN sites submit a proxyID file to Healthix, a health information exchange, which matches patients across sites. Matching algorithms involve demographic variables, such as date of birth, gender, and race. The final output is a proxyID map file used to merge and de-duplicate patients. The de-duplicated database contains a unique patient table with one record per patient across all sites. The personID associated with one de-identified patient is used to link to clinical data tables.

We mined the database to extract EHRs from individuals who had received the mCOI, a diagnosis code to indicate psychophysical visual disturbances, such as prosopagnosia. Prosopagnosia is defined as the selective impairment of face recognition abilities, while visual processing and intellectual functioning remain intact. Therefore, individuals diagnosed with other disorders of the eye and adnexa, developmental disorders, autistic disorder, childhood disintegrative disorder, mental, behavioral, and neurodevelopmental disorders, cognitive deficits following cerebral infarction, intellectual disabilities, alteration of consciousness or awareness, coma or vegetative state, other symptoms and signs involving cognitive functions and awareness, senile dementia, memory loss, Alzheimer's disease were counted but were excluded from further analyses (Table 1). Exclusion codes were applied *a priori* to refine selection of cases and limit analyses to the population of interest. The group of individuals carrying the mCOI and not carrying any of the exclusion codes formed the case-cohort. To systematically mine the database for extractable information of interest on the case-cohort, we created a computable phenotype. This computable phenotype consisted of a list of ICD-9 codes, containing the mCOI, as well as selected diagnostic codes for common and known cCOIs. Additionally, we extracted information on whether procedures, such as hemispherectomy or lobectomy, had been performed. The complete list of inclusion and exclusion codes used to create the computable phenotype is provided in Table 1.

For comparative analyses, a randomly sampled control-cohort was formed through matched selection of individuals who: (a) did not carry the mCOI and (b) did not carry any exclusion code. With a goal of a 1:100 match between cases and controls, data were drawn from this randomly sampled control-cohort. Case and control subjects were matched on age, gender, and hospital site. The %match macro was used to perform the matching in SAS [26] (version Enterprise Guide 7.1). Data derived from the case-cohort were analyzed to explore frequency and distribution of common cCOIs. Mixed effects logistic regression was performed to estimate odds ratios (ORs) of cCOIs within the case-cohort in contrast to the matched control-cohort.

Demographic and clinical variables were summarized using frequency and percentage for categorical variables and mean ± standard deviation for continuous variables. We calculated the distribution of prevalence rates across hospital sites based on total number of records per site. cCOIs were binary coded (Yes/No). Mixed effects logistic regression analysis was performed with prosopagnosia diagnosis (Yes/No) as the outcome variable. The mixed model included age, gender, race, ethnicity, and all comorbid diagnoses as fixed effects; hospital site was modeled as a random effect to account for patient clustering within site using Proc Glimmix in SAS Studio Version 3.7. ORs with 95% confidence intervals and estimates of the intraclass correlation coefficient (ICC) for the proportion of variance explained by site were reported. Selected conditions, previously reported to be associated with alterations of the face perception network, face recognition or emotion recognition, were studied as conditions of interest. The list of cCOIs included epilepsy and temporal lobe epilepsy [27], malignant neoplasm of the temporal lobe, brain trauma [28], herpes viral encephalitis [29], cerebral artery occlusion [30], frontotemporal dementia [31], bipolar depression [32], anxiety, panic and adjustment disorder, as well as neurasthenia [20], and posttraumatic stress disorder [33]. Dyslexia and alexia, dementia, as well as personality disorders were included as additional conditions of interest.

We computed ORs separately for cohorts below the age of 50 years and compared these results to the resulting ORs when the entire cohort was used. As simultanagnosia most commonly results from bilateral parietal cortical damage due to stroke [34] or posterior cortical atrophy [35], an atypical form of Alzheimer's disease, patients under the age of 50 years are conceivably less likely to present with simultanagnosia.

https://doi.org/10.1017/cts.2020.497 Published online by Cambridge University Press

Post hoc data mining was performed to reveal the overall prevalence rate of the exclusion-code-complex and to uncover the number of individuals carrying the mCOI who were excluded a priori. Minor inconsistencies in the overall number of individuals contained within the database at different points in time (time point one: time of case and control-cohort data mining versus time point two: time of post hoc data mining for prevalence rate extraction of the exclusion-code-complex) can be explained through ongoing de-duplication efforts.

To study additive effects of multiple cCOIs, we counted the number of diagnoses for each individual who carried the prosopagnosia code. Counts were used to organize data in four groups; diagnosed with one, two, three, or four or more cCOIs. Percentages and ORs were calculated. The model included age, sex, race, and ethnicity as covariates, and site was included as a random effect.

Lastly, we investigated percent frequencies of comorbid conditions among patients who were diagnosed with multiple cCOIs. To gain a better understanding of the newly revealed high concordance rate of the triad of prosopagnosia, depression, and anxiety diagnoses, we pooled data from individuals diagnosed with four or more cCOIs. Given our list of captured cCOIs, we reasoned that diagnoses of trauma and brain damage, such as cerebral artery occlusion, concussion, head injury, intracranial injury, and trauma, could be causally related to the diagnoses of depression and anxiety. Therefore, we formed subgroups through which cases with captured trauma and brain damage diagnoses could be systematically excluded. Four subgroups of interest were formed: (1) prosopagnosics with four or more cCOIs, (2) prosopagnosics with four or more cCOIs but without captured diagnosis of trauma, (3) depressive patients with four or more cCOIs but without captured diagnosis of prosopagnosia or trauma, (4) individuals diagnosed with anxiety disorder and four or more cCOIs but without captured diagnosis of prosopagnosia or trauma.

Results

In this study, a computable phenotype containing inclusion and exclusion diagnoses codes was used to form the case-cohort. Matched selection of a cohort of control subjects was performed, revealing a total of n = 101,875 patients eligible. Among these individuals, we identified 902 cases - individuals who carried the mCOI - indicating the presence of prosopagnosia or simultanagnosia. The matched control-cohort contained n = 100,973 patients.

Investigations revealed a diagnostic frequency of the mCOI of 0.012% or 12 per 100,000 individuals. The matching algorithm produced a cohort with a close distribution between women and men (51% women), at a mean age of 48.1 years (see Table 2). The distribution of prevalence rates across hospital sites varied between 0.0058% and 0.048% (p-value < 0.0001) (0.0074% at facility 1, 0.0058% at facility 2, 0.0134% at facility 3, 0.0198% at facility 4, 0.048% at facility 5, 0.0135% at facility 6, 0.027% at facility 7, and 0.0468% at facility 8).

Data were acquired across eight racial groups (White, American Indian or Alaska Native, Asian, Black or African American, Multiple Race, Native Hawaiian or Other Pacific Islander, "Other," or "Unknown," where category "Other" was assigned when none of the other seven categories could be applied, "Unknown" was assigned when racial information was not known). Because of sparse case counts in three racial categories (American Indian or Alaska Native, 3 cases, Asian, 15 cases, Multiple Race, 3 cases, Native Hawaiian or Other Pacific Islander, 1 case) we collapsed cases from these three categories with cases from the category "Other." Mixed effects logistic regression across four racial categories (Black or African American, White, Other, Unknown) revealed decreased ORs for individuals categorized as "Other" (OR = 0.68, 95% CI: 0.56, 0.83) or "Unknown," (OR = 0.73, 95% CI: 0.59, 0.90). Among three ethnic groups (Hispanic, non-Hispanic, and "Unknown"), ORs were elevated for Hispanic (OR = 1.67, 95% CI: 1.33, 2.09) (see Table 3). These findings indicate that the mCOI was more likely to be assigned to individuals of Hispanic ethnicity and less likely to be assigned to individuals of unknown race as well as to individuals with a racial background that was different from the defined categories.

The mixed effects model, controlling for age, gender, race, ethnicity, and all surrogate diagnoses, and after controlling for multiple comparison using Bonferroni corrections, revealed that individuals carrying the prosopagnosia diagnosis code were more likely to also carry diagnoses codes for personality disorder (OR = 5.68, 95% CI: 4.03, 8.02), depressive disorder (OR = 4.36, 95% CI: 3.63, 5.25), epilepsy and recurrent seizures (OR = 4.11, 95% CI: 3.20, 5.27), anxiety disorder (OR = 2.51, 95% CI: 2.07, 3.04), panic disorder (OR = 2.06, 95% CI: 1.34, 3.16), and major depressive episode (OR = 1.73, 95% CI: 1.30, 2.32) (see Table 3). The ICC for site was 0.0175 ± 0.0165 (standard error).

Computations of ORs for cohorts below the age of 50 years revealed no additional diagnoses of significance. While OR for the diagnoses of frontotemporal dementia and herpesviral encephalitis could not be assessed due to low counts, panic disorder diagnoses do not remain significantly increased in a population below the age of 50 years (See Supplementary Table 1).

Analyses of additive effects of cCOIs revealed that cases, when compared to controls, showed increased odds of having multiple comorbid conditions, approximately doubling for each added condition (see Table 4). More precisely, we found that cases with one additional cCOI were 6.25 times more likely, cases with two cCOIs were 13.92 times more likely, cases with three cCOIs were 24.50 times more likely, and cases with four cCOIs were 39.02 times more likely to carry the mCOI. These findings either reflect difficulties in assigning the diagnosis of prosopagnosia, resulting in an accumulation of other nonspecific diagnoses, or reflect the true and complex comorbidity landscape of the condition and the challenges of managing multiple comorbid conditions.

In-depth analyses of the comorbidity landscape among individuals diagnosed with four or more comorbid conditions provided insight into the concurrent expression of a number of cCOIs and revealed high concordance rates among prosopagnosia, anxiety, and depression (see Table 5). Overall, n = 326 individuals with four or more cCOIs were found in our dataset. The four subgroups contained (1) n = 44 prosopagnosics with four or more cCOIs, (2) n = 33 prosopagnosics with four or more cCOIs but without captured diagnosis of trauma, (3) n = 27 depressive patients with four or more cCOIs but without captured diagnosis of prosopagnosia or trauma, (4) n = 27 individuals diagnosed with anxiety disorder and four or more cCOIs but without captured diagnosis of prosopagnosia or trauma. Interestingly, the 27 individuals who formed groups 3 and 4 were found to be identical, meaning that the same individuals who had no diagnosis of prosopagnosia or trauma were all diagnosed with depression and anxiety.

A post hoc interrogation for the overall prevalence of individuals carrying any diagnosis of the exclusion-code-complex revealed a prevalence of 15.9% (1,124,064 individuals at a time when the database contained 7,089,021 individuals total). At that

https://doi.org/10.1017/cts.2020.497 Published online by Cambridge University Press

Table 2. *Demographics*

Variable Patient records queried n = 7,522,133	Cases (n = 902) n (%)	Controls (n = 100,973) n (%)	p-Value
Prevalence	902 (0.012%)		
Age (mean +/- SD)	48.1 +/- 28.0	48.1 +/- 28.3	0.96
Female gender	460 (51.0%)	50,600 (50.1%)	0.60
Site			1.0
Number 1	70 (7.8%)	7700 (7.6%)	
Number 2	11 (1.2%)	1210 (1.2%)	
Number 3	110 (12.2%)	12,971 (12.9%)	
Number 4	148 (16.4%)	16,389 (16.2%)	
Number 5	412 (45.7%)	46,083 (45.6%)	
Number 6	3 (0.3%)	265 (0.3%)	
Number 7	89 (9.9%)	9790 (9.7%)	
Number 8	59 (6.5%)	6565 (6.5%)	
Race			<0.0001
Black or African American	182 (20.2%)	15,733 (15.6%)	
Other	245 (27.2%)	29,855 (29.6%)	
White	283 (31.4%)	26,432 (26.2%)	
Unknown	192 (21.3%)	28,953 (28.7%)	
Ethnicity			<0.0001
Hispanic	114 (12.6%)	7636 (7.6%)	
Non-Hispanic	504 (55.9%)	52,086 (51.6%)	
Unknown	284 (31.5%)	41,251 (40.9%)	
Comorbid diagnoses			
Adjustment disorder	25 (2.8%)	552 (0.6%)	<0.0001
Anxiety disorder	229 (25.4%)	4668 (4.6%)	<0.0001
Cerebral artery occlusion	43 (4.8%)	1694 (1.7%)	<0.0001
Concussion	5 (0.6%)	191 (0.2%)	0.03
Dementia	35 (3.9%)	1138 (1.1%)	<0.0001
Depressive disorder	285 (31.6%)	4672 (4.6%)	<0.0001
Dyslexia and alexia	0 (0%)	3 (0%)	1.0
Epilepsy and recurrent seizures	96 (10.6%)	1324 (1.3%)	<0.0001
Frontotemporal dementia	1 (0.1%)	26 (0.03%)	0.21
Head injury	37 (4.1%)	1899 (1.9%)	<0.0001
Herpes encephalitis	1 (0.1%)	13 (0.01%)	0.12
Intracranial injury	3 (0.3%)	63 (0.06%)	0.02
Major depressive episode	85 (9.4%)	982 (1.0%)	<0.0001
Malignant neoplasm of temporal lobe	0 (0%)	26 (0.03%)	1.0
Panic disorder	35 (3.9%)	347 (0.3%)	<0.0001
Persistent mental disorders	0 (0%)	2 (0%)	1.0
Personality disorder	64 (7.1%)	249 (0.3%)	<0.0001
Post-traumatic stress disorder	29 (3.2%)	255 (0.3%)	<0.0001
Developmental disorders of scholastic skills	15 (1.7%)	486 (0.5%)	<0.0001
Trauma	5 (0.6%)	125 (0.1%)	0.006

https://doi.org/10.1017/cts.2020.497 Published online by Cambridge University Press

Table 3. *Mixed effects logistic regression*

Variable De-duplicated patient records queried n = 7,522,133	Odds ratio (95% CI)	p-Value
Age	1.00 (0.995, 1.00)	0.0537
Gender		0.2044
Male	1 [Reference]	
Female	0.92 (0.80, 1.05)	
Race		<0.0001*
White	1 [Reference]	
Black or African American	1.13 (0.92, 1.38)	
Other	0.68 (0.56, 0.83)	
Unknown	0.73 (0.59, 0.90)	
Ethnicity		<0.0001*
Non-Hispanic	1 [Reference]	
Hispanic	1.67 (1.33, 2.09)	
Unknown	1.03 (0.86, 1.23)	
Comorbid diagnoses		
Adjustment disorder	1.34 (0.86, 2.10)	0.1928
Anxiety disorder	2.51 (2.07, 3.04)	<0.0001*
Cerebral artery occlusion	1.44 (1.02, 2.04)	0.0396
Concussion	1.22 (0.44, 3.42)	0.7039
Dementia	1.56 (1.06, 2.29)	0.0238
Depressive disorder	4.36 (3.63, 5.25)	<0.0001*
Developmental disorders of scholastic skills	2.18 (1.26, 3.77)	0.0056
Epilepsy and recurrent seizures	4.11 (3.20, 5.27)	<0.0001*
Frontotemporal dementia	2.14 (0.28, 16.62)	0.4667
Head injury	0.99 (0.68, 1.45)	0.9761
Herpes encephalitis	3.05 (0.39, 24.15)	0.2904
Intracranial injury	1.68 (0.45, 6.35)	0.4433
Major depressive episode	1.73 (1.30, 2.32)	0.0002*
Panic disorder	2.06 (1.34, 3.16)	0.0010*
Personality disorder	5.68 (4.03, 8.02)	<0.0001*
Post-traumatic stress disorder	1.63 (1.00, 2.66)	0.0510
Trauma	1.82 (0.67, 4.94)	0.2398

*Significant at the Bonferroni-adjusted significance level of 0.0024 (0.05/21 variables)
Site is included in the model as a random effect with ICC = 0.0175 ± 0.0165 (se). Dyslexia and Alexia, Malignant Neoplasm of Temporal Lobe, and Persistent Mental Disorders were not included in the model due to sparse counts

time the overall mCOI prevalence was 0.038% (38 individuals per 100,000, 2727 individuals total), while 0.025% (25 individuals per 100,000, 1776 individuals total) carried any of the exclusion codes.

Discussion

Integration of clinical data across health systems has created new research databases [21,36] allowing accumulation of information. In the past, large EHR databases have been utilized to screen medical notes to study rare psychiatric conditions, such as Capgras delusion [37], a condition where familiar individuals are perceived as impostors [38]. The cited study revealed 187 cases suffering from Capgras delusion among the 250,000 records screened (74.8 cases per 100,000). In the current study, we were able to digitally screen more than 7.5 million patients' EHRs through utilization of a computable phenotype. This strategy allowed us to identify patients with the diagnosis of interest, revealing novel epidemiologic and demographic characteristics and through comparison to a matched control-cohort, permitted investigation of the condition's comorbidity landscape. Results derived from this study provide proof of principle for the feasibility of mining large clinical databases to identify sufficiently large case-cohorts and to employ strategies to identify comparator populations.

Data mining revealed 902 cases with a similar rate of male and female cases. This frequency is consistent with a prevalence rate of 0.012% prosopagnosia cases (12 per 100,000) detected among the sampled 7.5 million patients who had been seen across NYC academic health centers. The observed diagnostic rate reflects a prevalence rate that is lower than previously reported and one that is, unexpectedly, even lower than the prevalence rate of Capgras syndrome. This finding may be a reflection of difficulties to diagnose prosopagnosia as, for example, in comparison to Capgras syndrome [39], symptoms of prosopagnosia may not be as apparent, clear, and specific.

The detected case-cohort presumably contains acquired as well as developmental cases of prosopagnosia. While the two forms of prosopagnosia are easily distinguishable when patient history is available, the current study design does not allow for distinction between developmental and acquired cases. Although dates associated with diagnosis may be recorded, time and age of onset are not consistently captured. Further investigations of the temporal association of prosopagnosia diagnosis assignment in relation to the diagnosis of comorbid conditions would provide insight into the sequence of diagnoses and shed light onto the proportion of acquired versus developmental cases. Moreover, information on diagnostic criteria or the specialty of the diagnosing clinicians were not available but would provide valuable additional information to allow for a more detailed study of the formation of the case-cohort. Lastly, in this study, patients with a diagnosis of autism spectrum disorder (ASD) were excluded a priori; however, future studies should elucidate potential relationships between ASD, prosopagnosia, and related diagnoses.

Overall, the prevalence rate observed in this study is lower than previously published estimates of up to 3% for DP alone [16,17,40]. While our result could be a reflection of the true prevalence, it could alternatively be due to underdiagnosis and/or under-coding of true cases in the NYC area. The first assumption, that the prevalence estimated in this study indeed reflects an unbiased and accurate prevalence, implies that previously reported prevalence rates are overestimates of the true prevalence. This explanation seems not entirely implausible, since the majority of previous studies based their assessment on self-report alone [16,17]. Since individuals may underestimate their own ability to recognize faces [41–43], it is plausible that studies based on self-report would lead to overestimation. Nonetheless, even when diagnoses were based on the most commonly used objective measure of face recognition abilities [40], "The Cambridge Face Memory Test" (CFMT) [44], prevalence rates were 2–2.9%. In those studies, however, CFMT scores provided the only measure of face recognition abilities and, therefore, no additional or confirming evidence was available for cases identified. Hence, it is possible that a number

https://doi.org/10.1017/cts.2020.497 Published online by Cambridge University Press

Table 4. *Additive effects*

Number of comorbid conditions*	Cases (n = 902) n (%)	Controls (n = 100,973) n (%)	Adjusted OR (95% CI)	p-Value
One versus none	248 (39.6%)	9610 (9.9%)	6.25 (5.31, 7.37)	< 0.0001*
Two versus none	152 (28.7%)	2688 (3.0%)	13.92 (11.44, 16.94)	< 0.0001*
Three versus none	80 (17.5%)	828 (0.9%)	24.50 (18.96, 31.64)	< 0.0001*
Four or more versus none	44 (10.4%)	282 (0.3%)	39.02 (27.83, 54.71)	< 0.0001*

*Significant at the Bonferroni-adjusted significance level of 0.0063 (0.05/8 variables)
From mixed effects logistic regression with age, gender, race, and ethnicity as fixed effects, and site as a random effect with ICC = 0.0241 ± 0.0203 (se)

Table 5. *Summary for patients with four or more comorbid conditions*

	Frequency				
Comorbid conditions	All (n = 326)	Proso (n = 44)	Proso w/o trauma (n = 33)	Depression w/o trauma or Proso (n = 27)	Anxiety w/o trauma or Proso (n = 27)
Adjustment disorder	21.5%	6.8%	9.1%	25.9%	25.9%
Anxiety disorder	86.2%	88.6%	87.9%	100.0%	100.0%
Cerebral artery occlusion	25.2%	11.4%	/	/	/
Concussion	5.8%	0.0%	/	/	/
Dementia	21.5%	4.6%	3.0%	25.9%	25.9%
Depressive disorder	90.5%	95.5%	97.0%	100.0%	100.0%
Developmental disorders of scholastic skills	4.3%	2.3%	3.0%	0.0%	0.0%
Dyslexia and alexia	0.0%	0.0%	0.0%	0.0%	0.0%
Epilepsy and recurrent seizures	30.4%	36.4%	27.3%	33.3%	33.3%
Frontotemporal dementia	0.9%	0.0%	0.0%	0.0%	0.0%
Head injury	28.2%	9.1%	/	/	/
Herpes encephalitis	0.0%	0.0%	0.0%	0.0%	0.0%
Intracranial injury	2.5%	0.0%	/	/	/
Major depressive episode	53.4%	70.5%	75.8%	74.1%	74.1%
Malignant neoplasm of the temporal lobe	0.0%	0.0%	0.0%	0.0%	0.0%
Panic disorder	22.7%	43.2%	39.4%	44.4%	44.4%
Persistent mental disorders	0.0%	0.0%	0.0%	0.0%	0.0%
Personality disorder	23.9%	50.0%	54.6%	63.0%	63.0%
Post-traumatic stress disorder	21.5%	34.1%	36.4%	55.6%	55.6%
Trauma	5.2%	4.6%	/	/	/

Proso standing for prosopagnosics, w/o = without
Cells marked by "/" marks conditions that were excluded for the analyses of cases w/o trauma diagnoses

of low-performing individuals, who were categorized as being affected by prosopagnosia in those studies, may in fact not be symptomatic in everyday life, which would also result in an overestimation of the prevalence even though an objective measure was used. On the other hand, if previous estimates of prosopagnosia prevalence were accurate, our results imply that a substantial diagnostic gap exists between actual and diagnosed cases of prosopagnosia: less than one in a hundred prosopagnosics would actually be diagnosed. In this scenario, we would further argue that underdiagnosis could potentially and in part be a result of the fact that, while guidelines for definition and diagnosis of prosopagnosia have recently been published [45], there is still no clinical gold standard available. Moreover, clinical diagnoses may require prosopagnosic individuals to be aware of their own difficulties, while self-awareness and accurate assessment of one's own face recognition abilities have been reported to be limited [43,46]. Therefore, it is conceivable that when diagnoses depend on patients' self-awareness and self-report, under-reporting and underdiagnoses may be the result. Considering the evidence for either one of the two possibilities, we conclude that results derived from our study strongly suggest that

https://doi.org/10.1017/cts.2020.497 Published online by Cambridge University Press

diagnosis of prosopagnosia is infrequent and that the condition has been largely underdiagnosed over the years covered by our records. Despite the known imprecision of diagnostic coding and the likely widespread use of non-standardized clinical evaluation procedures, we conducted this large-scale study to examine the demographic and epidemiologic characteristics of prosopagnosia and to investigate frequencies and distribution of known comorbid conditions. To the best of the authors' knowledge, this report provides the first systematic quantitative evaluation of a large multi-ethnic urban cohort of potential prosopagnosia cases and allows novel insight into demographic characteristics and comorbid conditions.

The finding of equal gender distribution of diagnoses provides novel information and, most importantly, our comparative analyses revealed elevated ORs for several psychiatric conditions. Moreover, this study uncovered an elevated OR for receiving the diagnosis of prosopagnosia in Hispanic individuals. To the authors' best knowledge, this is the first time that increased rates of prosopagnosia diagnoses have been reported for a specific ethnic group and further investigations will be needed to elucidate the potential underpinnings of this finding. Analyses of additive effects revealed that ORs increased with increasing number of comorbid diagnoses. This may have important diagnostic implications, as this finding either reflects difficulties in assignment of the correct diagnoses or alternatively imply that prosopagnosia tends to co-occur with certain comorbid conditions and, therefore, increased awareness may be warranted when patients with certain comorbidities are seen in the clinic. Moreover, assessment of frequency of comorbid conditions showed, for the first time, high concordance rates of prosopagnosia, depression, and anxiety with largely overlapping comorbidity landscapes for the three conditions. Interestingly, we revealed a 100% overlap of those individuals diagnosed with four or more comorbidities including depression and anxiety when excluding prosopagnosia and trauma diagnoses. To gain a deeper understanding of the relationship between prosopagnosia, depression, and anxiety, future studies should investigate the temporal relationship between the emergence of face recognition difficulties and these conditions.

Results reported in this study provide crucial insights into the clinical, epidemiologic, and demographic characteristics of prosopagnosia and emphasize the need for enhanced awareness and diagnostic standardization for the assessment of face recognition abilities.

Acknowledgments. We would like to thank the INSIGHT Clinical Research Network/New York City Clinical Data Network (NYC-CDRN) and their data analysis team, in especially Yiyuyan Wu, for their help with data acquisition. We would like to thank the Clinical Research Support Office at The Rockefeller University's Center for Clinical and Translational Science for their assistance and support.

Supported in part by grant # UL1 TR001866 from the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH), and the Patient-Centered Outcomes Research Institute (PCORI) PCORnet grant to the NYC-CDRN grant # CDRN-106-03961.

Disclosures. The authors have no financial interests to disclose.

Data availability statement. Data have been provided by the INSIGHT Clinical Research Network and can be made available upon request. The INSIGHT Governance Board will review each request.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/cts.2020.497.

References

1. **Bodamer J**. Die Prosop-Agnosie (Die Agnosie des Physiognomieerkennens). *Arch Psychiatr Nervenkrank* 1947; **179**: 6–53.
2. **Kanwisher N, McDermott J, Chun MM.** The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 1997; **17**: 4302–4311.
3. **Kanwisher N, Stanley D, Harris A.** The fusiform face area is selective for faces not animals. *Neuroreport* 1999; **10**: 183–187.
4. **Kanwisher N, Yovel G.** The fusiform face area: a cortical region specialized for the perception of faces. *Philos Trans Royal Soc Lond B* 2006; **361**: 2109–2128.
5. **Tsao DY, Freiwald WA, Tootell RB, Livingstone MS.** A cortical region consisting entirely of face-selective cells. *Science* 2006; **311**: 670–674.
6. **Tsao DY, Moeller S, Freiwald WA.** Comparing face patch systems in macaques and humans. *Proc Natl Acad Sci USA* 2008; **105**: 19514–19519.
7. **Barton JJ.** Disorders of face perception and recognition. *Neurolog Clin* 2003; **21**: 521–548.
8. **Barton JJ.** Prosopagnosia associated with a left occipitotemporal lesion. *Neuropsychologia* 2008; **46**: 2214–2224.
9. **Barton JJ.** Structure and function in acquired prosopagnosia: lessons from a series of 10 patients with brain damage. *J Neuropsychol* 2008; **2**: 197–225.
10. **Behrmann M, Avidan G.** Congenital prosopagnosia: face-blind from birth. *Trends in Cognitive Sciences* 2005; **9**: 180–187.
11. **Damasio AR, Castro-Caldas A.** Letter: Prosopagnosia. *The Lancet* 1974; **2**: 892–893.
12. **Damasio AR, Damasio H, Van Hoesen GW.** Prosopagnosia: anatomic basis and behavioral mechanisms. *Neurology* 1982; **32**: 331–341.
13. **Duchaine B.** Developmental prosopagnosia: cognitive, neural, and developmental investigations. In: Rhodes G, Calder A, Johnson M, Haxby JV, eds. *Oxford Handbook of Face Recognition*. Oxford: Oxford University Press, 2011.
14. **Kennerknecht I, Plumpe N, Edwards S, Raman R.** Hereditary prosopagnosia (HPA): the first report outside the Caucasian population. *J Human Genetics* 2007; **52**: 230–236.
15. **Kennerknecht I, Pluempe N, Welling B.** Congenital prosopagnosia–a common hereditary cognitive dysfunction in humans. *Front Biosci* 2008; **13**: 3150–3158.
16. **Kennerknecht I, Ho NY, Wong VC.** Prevalence of Hereditary Prosopagnosia (HPA) in Hong Kong Chinese population. *Am J Med Genetics – Part A* 2008; **146A**: 2863–2870.
17. **Kennerknecht I, Grueter T, Welling B, *et al.*** First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). *Am J Med Genet A* 2006; **140**: 1617–1622.
18. **Fine DR.** A life with prosopagnosia. *Cogn Neuropsychol* 2012; **29**: 354–359.
19. **Dalrymple KA, Fletcher K, Corrow S, *et al.*** "A room full of strangers every day": the psychosocial impact of developmental prosopagnosia on children and their families. *J Psychosom Res* 2014; **77**: 144–150.
20. **Yardley L, McDermott L, Pisarski S, Duchaine B, Nakayama K.** Psychosocial consequences of developmental prosopagnosia: a problem of recognition. *J Psychosom Res* 2008; **65**: 445–451.
21. **Kaushal R, Hripcsak G, Ascheim DD, *et al.*** Changing the research landscape: the New York City Clinical Data Research Network. *Journal of the American Medical Informatics Association* 2014; **21**: 587–590.
22. **Sakurai Y, Hamada K, Tsugawa N, Sugimoto I.** Ventral simultanagnosia and prosopagnosia for unfamiliar faces due to a right posterior superior temporal sulcus and angular gyrus lesion. *Neurocase* 2016; **22**: 122–129.
23. **Barton JJS, Corrow SL.** The problem of being bad at faces. *Neuropsychologia* 2016; **89**: 119–124.
24. **Corrow SL, Dalrymple KA, Barton JJ.** Prosopagnosia: current perspectives. *Eye Brain* 2016; **8**: 165–175.
25. **Dalrymple KA**. Prosopagnosia. *Encyclopedia of the Neurological Sciences* 2014; **3**: 989–991.
26. **Erik J, Bergstralh JLK**. *Computerized Matching of Cases to Controls*. Mayo Foundation, 1995.
27. **Ahs F, Engman J, Persson J, *et al.*** Medial temporal lobe resection attenuates superior temporal sulcus response to faces. *Neuropsychologia* 2014; **61**: 291–298.

28. **Davies-Thompson J, Pancaroglu R, Barton J.** Acquired prosopagnosia: structural basis and processing impairments. *Front Biosc* 2014; **6**: 159–174.
29. **Fox CJ, Iaria G, Duchaine BC, Barton JJ.** Residual fMRI sensitivity for identity changes in acquired prosopagnosia. *Front Psychol* 2013; **4**: 756.
30. **Cousins R.** Prosopagnosia after stroke: potentials for impairment and treatment. *Top Stroke Rehabi* 2013; **20**: 471–477.
31. **Kumfor F, Hutchings R, Irish M**, ***et al.*** Do I know you? Examining face and object memory in frontotemporal dementia. *Neuropsychologia* 2015; **71**: 101–111.
32. **Robinson LJ, Gray JM, Burt M, Ferrier IN, Gallagher P.** Processing of facial emotion in bipolar depression and euthymia. *J Int Neuropsycholog Soc* 2015; **21**: 709–721.
33. **Schwarzbold M, Diaz A, Martins ET**, ***et al.*** Psychiatric disorders and traumatic brain injury. *Neuropsych Disease Treatment* 2008; **4**: 797–816.
34. **Rafal R.** *Neurological foundations of cognitive neuroscience - Chapter 2 Bálint's Syndrome: A Disorder of Visual Cognition*. MIT Press, Cambridge, MA, 2003.
35. **Schott JM, Crutch SJ**. Posterior cortical atrophy. *Continuum (Minneap Minn)* 2019; **25**: 52–75.
36. **Kho AN, Hynes DM, Goel S**, ***et al.*** CAPriCORN: Chicago Area Patient-Centered Outcomes Research Network. *J Am Med Inform Assoc* 2014; 21: 607–611.
37. **Bell V, Marshall C, Kanji Z, Wilkinson S, Halligan P, Deeley Q.** Uncovering Capgras delusion using a large-scale medical records database. *BJPsych Open* 2017; **3**: 179–185.
38. **Thiel CM, Studte S, Hildebrandt H, Huster R, Weerda R.** When a loved one feels unfamiliar: a case study on the neural basis of Capgras delusion. *Cortex* 2014; **52**: 75–85.
39. **Barrelle A, Luaute JP.** Capgras syndrome and other delusional misidentification syndromes. *Front Neurol Neurosci* 2018; **42**: 35–43.
40. **Bowles DC, McKone E, Dawel A**, ***et al.*** Diagnosing prosopagnosia: effects of ageing, sex, and participant-stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cogn Neuropsychol* 2009; **26**: 423–455.
41. **Murray E, Bate S.** Self-ratings of face recognition ability are influenced by gender but not prosopagnosia severity. *Psycholog Assess* 2019; **31**: 828–832.
42. **Murray E, Hills PJ, Bennetts RJ, Bate S.** Identifying hallmark symptoms of developmental prosopagnosia for non-experts. *Sci Rep* 2018; **8**: 1690.
43. **Bindemann M, Attard J, Johnston RA, Walla P.** Perceived ability and actual recognition accuracy for unfamiliar and famous faces. *Cogent Psychol* 2014; **1**: 986903.
44. **Duchaine B, Nakayama K.** The Cambridge Face Memory Test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia* 2006; **44**: 576–585.
45. **Bate S, Tree JJ.** The definition and diagnosis of developmental prosopagnosia. *Quart J Exp Psychol* 2017; **70**: 193–200.
46. **Laguesse R, Hall B, Irons J**, ***et al.*** Subjective self-assessment of face recognition ability is only weakly related to objective measures of face recognition performance. *J Vis* 2013; **13**: 979–979.

https://doi.org/10.1017/cts.2020.497 Published online by Cambridge University Press