**MANUSCRIPT**

Interrogating an ICD Coded Electronic Health Records Database to Characterize the Epidemiology of Prosopagnosia, Pressl et al.

**SUPPLEMENTARY**

**Supplementary Table 1. Mixed effects logistic regression < 50 years of age**

|  |  |  |
| --- | --- | --- |
| **Variable**De-duplicated Patient Records queried n = 52,368Cases = 460 Controls = 51,908 | **Odds Ratio** (95% CI) | **P-Value** |
|  |  |  |
| Age | 0.99 (0.985, 0.999) | 0.0182 |
| Gender |  | 0.4264 |
|  Male | 1 [Reference] |  |
|  Female | 0.92 (0.76, 1.12) |  |
|  |  |  |
| Race |  | 0.0006\* |
|  White | 1 [Reference] |  |
|  Black or African American | 1.43 (1.06, 1.91) |  |
|  Other | 0.79 (0.59, 1.05) |  |
|  Unknown  | 0.93 (0.68, 1.27) |  |
|  |  |  |
| Ethnicity |  | 0.0929 |
|  Non-Hispanic | 1 [Reference] |  |
|  Hispanic | 1.41 (1.02, 1.93) |  |
|  Unknown | 1.02 (0.79, 1.31) |  |
|  |  |  |
| Comorbid Diagnoses  |  |  |
|  Adjustment Disorder | 1.42 (0.81, 2.48) | 0.2199 |
|  Anxiety Disorder | 3.11 (2.38, 4.07) | <0.0001\* |
|  Cerebral Artery Occlusion | 1.86 (0.80, 4.33) | 0.151 |
|  Concussion | 2.06 (0.65, 6.48) | 0.2185 |
|  Dementia | 0.70 (0.08, 6.54) | 0.7562 |
|  Depressive Disorder | 5.54 (4.23, 7.24) | <0.0001\* |
|  Developmental Disorders of Scholastic Skills | 1.91 (1.06, 3.45) | 0.0313 |
|  Epilepsy and Recurrent Seizures | 5.09 (3.58, 7.23) | <0.0001\* |
|  Head Injury | 0.66 (0.34, 1.25) | 0.1976 |
|  Intracranial Injury | 1.52 (0.28, 8.17) | 0.6269 |
|  Major Depressive Episode | 1.83 (1.25, 2.67) | 0.0017\* |
|  Panic Disorder | 1.98 (1.13, 3.49) | 0.0172 |
|  Personality Disorder | 6.58 (4.40, 9.83) | <0.0001\* |
|  PTSD | 1.48 (0.85, 2.58) | 0.1687 |
|  Trauma | 2.40 (0.76, 7.65) | 0.1375 |
| \* Significant at the Bonferroni adjusted significance level.Site is included in the model as a random effect with ICC = 0.0147 ± 0.02044 (se). Dyslexia and Alexia, Frontotemporal Dementia, Herpes Encephalitis, Malignant Neoplasm of Temporal Lobe, and Persistent Mental Disorders were not included in the model due to sparse counts. |

**Prosopagnosia query script**

**proc** **sql** noprint;

create table outdata.tobin\_prosopagnosia\_include as

select distinct patid,

dx as in\_dx label='inclusion diagnosis',

cdrn\_facilityid as in\_cdrn\_facilityid,

min(admit\_date) as earliest\_dxdt format date9.

from indata.diagnosis

WHERE dx in ('368.16','R48.3') and admit\_date >= **'01Jan2007'd**

group by patid

order by patid;**quit**;

\*Exclusion criteria -- Cases excluded from the included;

**proc** **sql** noprint;

create table outdata.tobin\_prosopagnosia\_exclude as

select

distinct patid ,

'exclude' as exclude

from indata.diagnosis

where patid in (select distinct patid from outdata.tobin\_prosopagnosia\_include)

and (dx in

( /\*keep\*/'290.0',/\*keep\*/'780.93',/\*keep\*/'G30',/\*keep\*/'I69.31' )

 /\*keep\*/or dx between '317' and '319'

 /\*keep\*/or dx between '360' and '367.99'

 /\*keep\*/or dx between '369' and '379.99'

 /\*keep\*/or dx between 'F01' and 'F99'

 /\*keep\*/or dx between 'H00' and 'H59'

 /\*keep\*/or dx like '299%'

 /\*keep\*/or dx like '780.0%'

 /\*keep\*/or dx like 'R41%')

order by **1**;**quit**;

**proc** **sql**;

create table outdata.tobin\_case as /\* patients met inclusion and had no exclusion \*/

select a.\*, 'no exclusion' as exclude

from outdata.tobin\_prosopagnosia\_include a

left join outdata.tobin\_prosopagnosia\_exclude b

on a.patid=b.patid

where b.patid is null;

**quit**;

**proc** **sql** ;

create table finalout.case as /\*final case ouput \*/

select a.\*

,b.dx

,b.admit\_date as surrogate\_dxdt format date9.

,min(b.admit\_date) as first\_surrogate\_dxdt format date9.

,b.cdrn\_facilityid

,

 **2006**-year(c.birth\_date) as age\_2007 label ="age at 1/1/2007"

,

c.sex,

c.hispanic,

d.enr\_basis

from outdata.tobin\_Case as a

left join

indata.diagnosis as b

on a.patid=b.patid and (b.dx in ('784.61','191.2','434.91','434.11','434.01','331.19','294','310.2','959.01',

 '058.29','054.3','311','296.2','296.3','300.0','293.84','300.05','309.0',

 '309.24','309.28','309.29','309.81','01.52','01.53')

 or b.dx between '296.21' and '296.26'

 or b.dx between '296.31' and '296.36'

 or b.dx like '345%'

 or b.dx like '290%'

 or b.dx like '850%'

 or b.dx like '854%'

 or b.dx like '301%'

 or b.dx like '300%'

 or b.dx like '315%'

 or b.dx like 'R48.%'

 or b.dx like 'F81.%'

 or b.dx in ('G40',

 'C71',

 'I66.29',

 'I66',

 'G31',

 'F06',

 'F07',

 'S06',

 'S02',

 'B10',

 'B00.4',

 'F32',

 'F33',

 'F41',

 'F48.8',

 'F43',

 'F81')

)

left join

indata.demographic as c

on a.patid=c.patid

left join

indata.enrollment as d

on a.patid=d.patid

group by a.patid

order by **1**,**3**;**quit**;