APPENDIX 1

**Table 1**

**Description of Registers**

*Multi-Generation Register*

The Multi-Generation Register is a register made up of persons who have been registered in Sweden at some time since 1961 and those who were born in 1932 or later. These are called index persons. The register contains connections between index persons and their biological parents. There are about 11 million index persons in the register. The Multi-Generation Register is a part of the register system for Total Population Register, where information comes from the National Tax Board. Every year, a new version of the register is created, including new index persons who immigrated or were born during the year. Information from the Multi-Generation Register may be disclosed for research and statistical purposes. For more information, see *Statistics Sweden, Background Facts, Population and Welfare Statistics 2017:2, Multi-generation register 2016. A description of contents and quality*

*National Patient Register*

In the 1960's the National Board of Health and Welfare started to collect information regarding in-patients at public hospitals, the National Patient Register (NPR). Initially it contained information about all patients treated in psychiatric care and approximately 16 percent of patients in somatic care. The register at that time covered six of the 26 county councils in Sweden. In 1984, the Ministry of Health and Welfare together with the Federation of County Councils decided a mandatory participation for all county councils. From 1987, NPR includes all in-patient care in Sweden. Since 2001, the register also covers outpatient doctor visits including day surgery and psychiatric care from both private and public caregivers. For more information, see *https://www.socialstyrelsen.se/en/statistics-and-data/registers/register-information/the-national-patient-register/*

*Primary Care Registry*

We also used information from our new Primary Care Registry (PCR), a research dataset including individual-level information on clinical diagnoses from primary health care centers from the following 15 of the 21 Swedish counties: Blekinge (2009-2016), Värmland (2005-2015), Kalmar (2007-2016), Sörmland (1997-2017), Uppsala (2005-2015), Västernorrland (2008-2015), Norrbotten (2009-2016), Gävleborg (2010-2016), Halland (2007-2014), Jönköping (2008-2014), Kronoberg (2006-2016), Skåne (1998-2013), Östergötland (1997-2014), Stockholm (2003-2016), and Västergötland (2000-2013). In 2016, these counties included 87% of the Swedish population. For more information see *Sundquist, J., Ohlsson, H., Sundquist, K. et al. Common adult psychiatric disorders in Swedish primary care where most mental health patients are treated. BMC Psychiatry 17, 235 (2017).*

*The Swedish Occupation register*

The Swedish Standard Classification of Occupations 2012 (SSYK 2012) is a system for classifying and aggregating data about occupations in administrative registers. SSYK 2012 is an update of the previous classification of occupations SSYK 96, which it also replaces. SSYK 2012 is based on the International Classification of Occupations 2008 (ISCO-08). SSYK 2012 is intended to cover all jobs on the Swedish labour market for which salary or other compensation is paid. Voluntary work is excluded, but it should be possible to classify unpaid work in the family business. In principle, it should only be possible to classify any given job in one way. The classification covers the following ten broad occupational fields (Major groups):

1. Managers

2. Occupations requiring advanced level of higher education

3. Occupations requiring higher education qualifications or equivalent

4. Administration and customer service clerks

5. Service, care and shop sales workers

6. Agricultural, horticultural, forestry and fishery workers

7. Building and manufacturing workers

8. Mechanical manufacturing and transport workers, etc.

9. Elementary occupations

0. Armed forces occupations

For more information, see *https://www.scb.se/contentassets/9f203b733c2942ec971fb098a7800417/ssyk-2019.pdf (In Swedish)*

**Table 2**

**OCCUPATION**

**Definition of occupation:**

The occupation recorded is the one with the highest taxable salary in November every year (2014 -2018). We selected the occupation that the individual was recorded in for the most number of years. If the individual was recorded in two occupations the same number of years, we selected the occupation that the individual was last recorded in.

**Strategy for excluding individuals that were registered with the disorder prior to entering the occupation:**

As we only have information on occupation for the years 2014 to 2018, we do not know if those individuals registered in their occupation in 2014 is their first year in the occupation. Therefore, in order to estimate the mean age at first registration in the occupation we only looked at individuals that had their first registration in the occupation during the years 2016 to 2018. In this sample, we estimated the mean age at first registration in the occupation and applied it to the entire sample (2014 to 2018). We then excluded individuals that had a first registration of the disorder prior to this estimated mean age at first registration in occupation.

**Strategy to test the hypothesis that individuals did not chose the occupation because individuals had a close relative registered with the disorder:**

In a separate database, we included all first-degree relatives to our probands, their age at registration of the ten disorders, their sex and year of birth. We then used the same underlying liability as in step 2 for the calculation of the FGRS to obtain sex and birth decade specific Z-scores for all relatives with the disorder and relatives without the disorder. We then weighted the relatives without the disorder using the same non-parametrically estimated distribution of *Age at first registration* as in step 1 in the calculation of the FGRS. We then average the Z-score across all first-degree relatives. Then for each of the significant occupations we performed a linear regression analysis, using the entire population, with the average Z-score as outcome and a dummy variable defining if the individual were registered in the specific occupation or not. In the regression analysis, we controlled for the original FGRS. If the dummy variable is positive and significant, it means that individuals in this occupation had overall more first-degree relatives with the disorder than individuals in other occupations, while controlling for their FGRS.

**Table 3**

**Familial genetic risk score (FGRS)**

*The FGRS was calculated for MD, AD, OCD, BD, SZ, AN, AUD, DUD, ADHD, ASD and Years of Education*

The dataset for the calculations includes:

Column1 = Identification number of the proband (Born 1932-1995)

Column2 = Identification number of the relative (1st to 5th degree relatives). The mean (SD) number of proband were as follows: 1st 4.55 (2.1); 2nd 7.63 (4.5); 3rd 8.61 (6.7); 4th 12.44 (10.2); 5th 6.90 (6.3).

Column3 = Proportion of shared additive genetic effects (0.03125 to 0.50) with the proband

Column4 = Year of Birth of relative

Column5 = Sex of relative

Column6 = Age at registration for disorder

Column7 = Age at end of follow-up (2017-12-31 or age at death, or age at emigration whichever came first)

**Step 1:** Using all unique relatives with a registration for the disorder of interest, we non-parametrically estimated the distribution of *Age at first registration*. The empirical distribution is used to obtain weights for relatives without a registration for the disorder, in order to account for the proportion of the time-at-risk period they had completed at the end of follow-up. For example, for relatives at age x at end of follow-up, the weight corresponds to the proportion of relatives registered for the disorder that had been registration at age x. For relatives born prior to 1958 we subtracted age at the end of follow-up with the following formula: 1958 - Year of birth of relative. This modification was done in order to control for registration effects (i.e, most registers in Sweden start in 1973 suggesting that relatives from early birth cohorts do not have the possibility to be registered at younger ages). Note that all relatives with the disorder are weighted one.

**Step 2:** Transform the binary variable (disorder yes/no) into a z-score based on the threshold for each disorder. The underlying liability of the individual is not assessable. Instead, we estimated the mean of the underlying liability to obtain sex and birth decade specific Z-scores for relatives with the disorder registration and relatives without the disorder. We generate n random numbers from a N(0, 1) distribution and estimate the mean for relatives registered with the disorder (i.e., mean of the observations above the threshold) and for relatives without a registration (i.e., mean of all observation below the threshold). The thresholds are calculated for each decade of birth and sex.

**Step 3**: Correct for cohabitation effects. To estimate the cohabitation effect (i.e. “shared environment”), we created a database with all individuals in the Swedish population born in Sweden 1955-1990. We also included the number of years, during ages 0-15, that individuals resided in the same household as their biological father. We thereby were able to define two kinds of families i) “not-lived-with” father families (offspring never resided for more than 1 year in the same household or in the same community as their biological father); ii) “lived-with” father (offspring resided a minimum of 13 year in the same household as their biological father. We performed a logistic regression model with the binary disorder in offspring as outcome and the binary disorder in father, type of father, and their interaction as predictors. We used the interaction term as the difference of effect between genes only and genes + environment. The same approach was performed for half-siblings where we compared those who were reared together versus reared apart. The following interaction terms were used in the calculations for our ten disorders and educational attainment:

|  |  |  |
| --- | --- | --- |
| Disorder | Parent/Children | Siblings |
| MD | 0.80 | 0.85 |
| AD | 0.87 | 0.81 |
| OCD | 0.79 | 0.74 |
| BD | 0.67 | 0.77 |
| SZ | 0.93 | 0.84 |
| AN | \* | 0.88 |
| AUD | 0.99 | 0.69 |
| DA | 0.92 | 0.52 |
| ADHD | 0.42 | 0.81 |
| ASD | 0.83 | 0.61 |
| YOE | 0.82 | 0.73 |
| \*No reliable estimate due to very low prevalence rates in males – we therefore used the mean among all other traits |

**Step 4:** Calculate the product for each relative using the four components:

1. Z-score (reflecting sex and year of birth adjusted rates)
2. Weight (reflecting the proportion of risk period they had completed)
3. Cohabitation effects
4. Proportion of shared genetic (0.03125 to 0.50) with the proband

**Step 5:** Average the product calculated in step 4 across all relatives to a proband

**Step 6**: Correct for the number of relatives. We multiplied the results from step 5 with a shrinkage factor. Shrinkage factor (SF): B / (B+A/C). It produces more shrinkage if B and C are small and A is large.

1. the variance of the z-score of the disorder across all relatives,
2. the variance in the mean z-score across all probands,
3. the weighted number of relatives for each proband (sum of Column 3 across each proband).

**Step 7:** Correct for difference by year of birth and county differences. There are 21 counties in Sweden. For each proband we used the county they had resided in during the maximum number of years (measured from 1969 and onwards) We standardized the risk score by year of birth and county of the proband into a z-score with mean 0 and SD 1. This was then used as the FGRS in the analyses.

**Method of Simulations**

To get realistic (Swedish population-like) simulations, after unsuccessful attempts to utilize the *R* pedigree simulating packages *pedSimulate* and *synbreed,* we implemented a de-novo pedigree simulation using Julia script because of its greater speed. For increased generality, the script was built to have numerous adjustable parameters:

1. Additive heritability of the trait [setting for this manuscript (SFTM) h2= {20%, 40%, 60%, 80%}],
2. variance of siblings’ trait that is explained by the common or shared environment (SFTM c2= {2.5%, 5%, 10%, 20%}) which was applied only for full siblings,
3. k=number of generations (SFTM: k=5, i.e., founder generation gen=0 and gen=1- 4 for subsequent generations),
4. vector of average number of children per couple in generations 0 to k-2 (SFTM: $μ=${2.1, 2.2, 1.7, 1.7}, as estimated from Swedish registries assuming average generation time is 25 years),
5. number of founders (SFTM: n=500K),
6. number of independent breeding groups (SFTM m=500, rather similar to villages),
7. (to avoid inbreeding) number of subgroups=k-1 for mothers in a subgroup to breed circularly with fathers from the next subgroup. (Children inherit the subgroup of mothers.)

The theoretical algorithm simulations were as follows:

1. Simulate independent True Breeding Values (TBV) for founder generation (gen=0), i.e. $TBV\_{j}=\sqrt{h^{2}}\* Z\_{j}$, where $Z\_{j}$ are independent standard normal (Gaussian) variates (j=1,…n),
2. For subsequent generations (gen=i>0)
	1. Within each group
		1. Permute mothers from one subgroups and fathers from the next,
		2. Pair mothers and fathers with the same rank,
		3. For each pair, simulate number of sibs m~ Poisson ($μ\_{i}$),
		4. If m>0, within each sibship
			1. Simulate sib’s j TBVj as the sum of parent’ average and mendelian sampling, i.e., sib $TBV\_{j}=\frac{TBV\_{mother}+TBV\_{father}}{2}+\sqrt{\frac{h^{2}}{2}} Z\_{j}$, where $Z\_{j}$ are independent standard normal (Gaussian) variates (j=1,…m),
			2. Simulate the common environment for all sibs within family as, $C= \sqrt{c^{2}} Z$, with $Z$ a single Gaussian variant for entire sibship,
			3. Simulate the independent environment for each sib within family as, $E\_{j}= \sqrt{1-h^{2}-c^{2}} Z\_{j}$, where $Z\_{j}$ are independent standard normal (Gaussian) variates (j=1,…m),
			4. Compute liability for each sib as $L\_{j}= TBV\_{j}+C+E\_{j}$
			5. Compute the affected status for each sib using the liability using a liability threshold model (for computational efficiency, computing affected status for multiple prevalences in a single pass).

Our simulations contained a mean (SD) of 324,656 (1,105) probands, each proband having a mean number of 3.7 (SD: 1.3) 1st degree relatives, 7.4 (SD: 1.8) 2nd degree relatives, 13.7 (SD: 4.3), 3rd degree relatives, and 23.3 (5.2) 4th degree relatives for a total mean number of relatives: 48.1 (SD:8.7).

Table 4

43 3-Digit Occupational Classes with No Evidence of Elevated FGRS for Psychiatric Disorders

|  |
| --- |
| Accountants, financial analysts and fund managers |
| Administration and planning managers |
| Administration and service managers not elsewhere classified |
| Architects and surveyors |
| Architectural and engineering managers |
| Biologists, pharmacologists and specialists in agriculture and forestry |
| Dentists |
| Designers |
| Education managers not elsewhere classified |
| Elderly care managers |
| Engineering professionals |
| Finance managers |
| Financial and insurance managers |
| Forestry and agricultural production managers |
| Health care managers |
| Hotel and conference managers |
| Human resource managers |
| Information and communications technology service managers |
| Information, communication and public relations managers |
| Legal professionals |
| Legislators and senior officials |
| Managing directors and chief executives |
| Marketing and public relations professionals |
| Mathematicians, actuaries and statisticians |
| Naprapaths, physiotherapists, occupational therapists |
| Organization analysts, policy administrators and human resource specialists |
| Other services managers not elsewhere classified |
| Other social services managers |
| Physicists and chemists |
| Preschool managers |
| Primary and secondary schools and adult education managers |
| Production managers in construction and mining |
| Production managers in manufacturing |
| Real estate and head of administration manager |
| Research and development managers |
| Retail and wholesale trade managers |
| Sales and marketing managers |
| Specialists in health care not elsewhere classified |
| Specialists within environmental and health protection |
| Sports, leisure and wellness managers |
| Supply, logistics and transport managers |
| Veterinarians |
| Vocational education teachers |

Figure 1 – Results of Simulations of Pedigrees Containing 1st-5th Degree Relatives Analyzed by FGRS as a Function of Heritability and Prevalence

Figure 2

For Figures 2-4, we included in our simulations, estimates of shared environment for siblings with c2 equal to, respectively, 2.5, 5 and 10%. The thick colored lines are the estimates with the addition of the shared environment. The dotted line are those calculaed with the c2 parameter added. We then ”correct” for that sibling effect with 4 values of ”down-weighting”: 0.8, 0.6, 0.4, 0.2, which are represented by the thinner lines in the figures.



Figure 3



Figure 4



Table 5

|  |  |  |
| --- | --- | --- |
|  |  | Number of lifetime Schizophrenia diagnoses in the registers |
|  |  | 1 | 2 | 3-5  | 6-10  | More than 10  |
| Number of lifetime Bipolar Disorder diagnoses in the registers | 1 | Last diagnosis | Last diagnosis | Most common diagnosis | Most common diagnosis | Most common diagnosis |
| 2 | Last diagnosis | Majority of last 3 diagnoses | Majority of last 3 diagnoses | Most common diagnosis | Most common diagnosis |
| 3-5  | Most common diagnosis | Majority of last 3 diagnoses | Majority of last 3 diagnoses | Majority of last 3 diagnoses | Majority of last 5 diagnoses |
| 6-10  | Most common diagnosis | Most common diagnosis | Majority of last 3 diagnoses | Majority of last 5 diagnoses | Majority of last 5 diagnoses |
| More than 10  | Most common diagnosis | Most common diagnosis | Majority of last 5 diagnoses | Majority of last 5 diagnoses | Majority of last 5 diagnoses |

**Table 6**

**RESULTS FOR MD *WITHOUT* HIERARCHY (RESULTS FROM MS)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **MD** | **95% CI LL** | **95% CI UL** |
| Ministers and deacons | 68% | 30% | 107% |
| Managers and leaders within religious bodies | 53% | 11% | 94% |
| Psychologists and psychotherapists | 47% | 21% | 74% |
| Creative and performing artists | 30% | 13% | 47% |
| Social work and counselling professionals | 28% | 13% | 44% |
| Authors, journalists and linguists | 25% | 10% | 39% |
| Medical doctors | 23% | 10% | 36% |
| Teaching professionals not elsewhere classified | 20% | 9% | 32% |
| Museum curators and librarians  | 20% | 7% | 33% |
| University and higher education teachers | 9% | 2% | 16% |
| Secondary education teachers | 7% | 1% | 14% |
| Nursing professionals | 4% | 1% | 7% |
|  |   |  |  |
| **RESULTS FOR MD *WITH* HIERARCHY**  | **MD** | **95% CI LL** | **95% CI UL** |
| Ministers and deacons | 74% | 37% | 110% |
| Managers and leaders within religious bodies | 67% | 20% | 114% |
| Psychologists and psychotherapists | 48% | 24% | 72% |
| Creative and performing artists | 32% | 15% | 48% |
| Social work and counselling professionals | 31% | 16% | 45% |
| Authors, journalists and linguists | 26% | 12% | 40% |
| Medical doctors | 24% | 12% | 37% |
| Managers in social and curative care | 23% | 5% | 41% |
| Teaching professionals not elsewhere classified | 18% | 9% | 28% |
| Museum curators and librarians and related professionals | 17% | 5% | 29% |
| University and higher education teachers | 14% | 5% | 22% |
| Nursing professionals | 10% | 4% | 15% |
|  |  |  |  |

**RESULTS FOR AD *WITHOUT* HIERARCHY (RESULTS FROM MS)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **AD** | **95% CI LL** | **95% CI UL** |
| Ministers and deacons | 54% | 22% | 85% |
| Creative and performing artists | 41% | 18% | 63% |
| Psychologists and psychotherapists | 39% | 17% | 61% |
| Authors, journalists and linguists | 28% | 12% | 44% |
| Museum curators and librarians  | 27% | 11% | 43% |
| Social work and counselling professionals | 25% | 11% | 38% |
| Teaching professionals not elsewhere classified | 22% | 10% | 35% |
| University and higher education teachers | 14% | 5% | 23% |
| Secondary education teachers | 9% | 3% | 16% |
|  |  |  |  |
| **RESULTS FOR AD *WITH* HIERARCHY**  | **AD** | **95% CI LL** | **95% CI UL** |
| Ministers and deacons | 62% | 29% | 96% |
| Creative and performing artists | 52% | 26% | 78% |
| Managers and leaders within religious bodies | 50% | 9% | 91% |
| Psychologists and psychotherapists | 37% | 17% | 57% |
| Social work and counselling professionals | 27% | 13% | 41% |
| Authors, journalists and linguists | 27% | 12% | 41% |
| Museum curators and librarians and related professionals | 26% | 11% | 41% |
| Teaching professionals not elsewhere classified | 24% | 12% | 37% |
| Managers in social and curative care | 22% | 6% | 39% |
| Secondary education teachers | 13% | 5% | 21% |
| University and higher education teachers | 11% | 3% | 18% |

**RESULTS FOR BD WITHOUT HIERARCHY (RESULTS FROM MS)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **BD** | **95% CI LL** | **95% CI UL** |
| Ministers and deacons | 18% | -4% | 40% |
| Psychologists and psychotherapists | 18% | -4% | 39% |
| Creative and performing artists | 12% | -3% | 27% |
| Museum curators and librarians  | 11% | -3% | 25% |
| Medical doctors | 10% | -2% | 21% |
| Social work and counselling professionals | 8% | -2% | 18% |
| University and higher education teachers | 7% | -2% | 16% |
| Authors, journalists and linguists | 7% | -2% | 16% |
| Teaching professionals not elsewhere classified | 5% | -1% | 11% |
|  |  |  |  |
|  |  |  |  |
| **RESULTS FOR BD *WITH* HIERARCHY** |  |  |  |
|  | **BD** | **95% CI LL** | **95% CI UL** |
| Psychologists and psychotherapists | 17% | -4% | 39% |
| Ministers and deacons | 17% | -4% | 38% |
| Creative and performing artists | 12% | -3% | 27% |
| Museum curators and librarians and related professionals | 10% | -3% | 23% |
| Medical doctors | 9% | -2% | 20% |
| Social work and counselling professionals | 8% | -2% | 17% |
| University and higher education teachers | 7% | -2% | 15% |
| Authors, journalists and linguists | 7% | -2% | 16% |
| Teaching professionals not elsewhere classified | 5% | -1% | 11% |

**RESULTS FOR SZ *WITHOUT* HIERARCHY (RESULTS FROM MS)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **SZ** | **95% CI LL** | **95% CI UL** |
| Museum curators and librarians  | 8% | -4% | 19% |
| Psychologists and psychotherapists | 6% | -3% | 16% |
| Creative and performing artists | 5% | -3% | 12% |
| Medical doctors | 4% | -2% | 11% |
| Authors, journalists and linguists | 4% | -2% | 11% |
| University and higher education teachers | 3% | -2% | 9% |
| Social work and counselling professionals | 3% | -2% | 9% |

**RESULTS FOR SZ *WITH* HIERARCHY**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **SZ** | **95% CI LL** | **95% CI UL** |
| Museum curators and librarians and related professionals | 7% | -4% | 18% |
| Psychologists and psychotherapists | 6% | -3% | 15% |
| Creative and performing artists | 5% | -3% | 12% |
| Medical doctors | 4% | -2% | 11% |
| Authors, journalists and linguists | 4% | -2% | 10% |
| University and higher education teachers | 4% | -2% | 9% |
| Social work and counselling professionals | 3% | -2% | 9% |