APPENDIX

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).*

Table 2 Definitions of Phenotypes

|  |  |  |
| --- | --- | --- |
|  | Registers Used | Definition |
| Major Depression (MD) | The Swedish Hospital Discharge Register (coverage 1973-2017); Outpatient Care Register (national coverage 2001-2017); Primary Care Registry (Partly coverage from 1999-2017) | ICD-8: 296.2, 298.0, 300.4; ICD-9: 296.2, 296.4, 298.0, 300.4; ICD-10: F32, F33.  **Note**: all individuals with a registration for BD were excluded. |
| Drug Use Disorder (DUD) | The Swedish Hospital Discharge Register (coverage 1973-2017); Outpatient Care Register (national coverage 2001-2017); Primary Care Registry (Partly coverage from 1999-2017); the Swedish Drug Register (2005-2017); the Swedish Mortality Register, and the Swedish Criminal Register (1973-2017) and the Swedish Suspicion Register (1998-2017) | Drug Use Disorder (DUD) was identified in the Swedish medical and mortality registries by ICD codes (ICD8: Drug dependence (304); ICD9: Drug psychoses (292) and Drug dependence (304); ICD10: Mental and behavioral disorders due to psychoactive substance use (F10-F19), except those due to alcohol (F10) or tobacco (F17)); in the Suspicion Register by codes 3070, 5010, 5011, and 5012, that reflect crimes related to DUD; and in the Crime Register by references to laws covering narcotics (law 1968:64, paragraph 1, point 6) and drug-related driving offences (law 1951:649, paragraph 4, subsection 2 and paragraph 4A, subsection 2). DUD was identified in individuals (excluding those suffering from cancer) in the Prescribed Drug Register who had retrieved (in average) more than four defined daily doses a day for 12 months from either of Hypnotics and Sedatives (Anatomical Therapeutic Chemical (ATC) Classification System N05C and N05BA) or Opioids (ATC: N02A). |
| ADHD | The Swedish Hospital Discharge Register (coverage 1973-2017); Outpatient Care Register (national coverage 2001-2017); Primary Care Registry (Partly coverage from 1999-2017) | ICD-9: 314; ICD-10: F90 |
|  |  |  |
| Anxiety Disorder (AD) | The Swedish Hospital Discharge Register (coverage 1973-2017); Outpatient Care Register (national coverage 2001-2017); Primary Care Registry (Partly coverage from 1999-2017) | ICD-8: 300.0, 300.2 ; ICD-9: 300A, 300C; ICD-10: F40, F41 |
| Bipolar Disorder (BD) | The Swedish Hospital Discharge Register (coverage 1973-2017); Outpatient Care Register (national coverage 2001-2017); Primary Care Registry (Partly coverage from 1999-2017) | ICD-8: 296.1, 296.3, 296.8, 296.9, 298.1; ICD-9: 296A, 296C, 296D, 296E, 296W, 298B; ICD-10: F30, F31. **Note**: BP and SZ were separated based on table below. |
| Schizophrenia (SZ) | The Swedish Hospital Discharge Register (coverage 1973-2017); Outpatient Care Register (national coverage 2001-2017); Primary Care Registry (Partly coverage from 1999-2017) | ICD-8: 295.1, 295.2, 2953, 295.9, 295.6; ICD-9: 295B, 295C, 295D, 295G, 295X; ICD-10: F200, F201, F202, F203, F205, F209. **Note**: BP and SZ were separated based on table below. |
| Alcohol Use Disorder (AUD) | The Swedish Hospital Discharge Register (coverage 1973-2017); Outpatient Care Register (national coverage 2001-2017); Primary Care Registry (Partly coverage from 1999-2017); the Swedish Drug Register (2005-2017); the Swedish Mortality Register, and the Swedish Criminal Register (1973-2017) and the Swedish Suspicion Register (1998-2017) | Alcohol Use Disorder (AUD) was identified in the Swedish medical and mortality registries by ICD codes: ICD9: V79B, 305A, 357F, 571A-D, 425F, 535D, 291, 303, 980; ICD 10: E244, G312, G621, G721, I426, K292, K70, K852, K860, O354, T51, F10); in the Crime Register by codes 3005, 3201, which reflect crimes related to alcohol abuse; in the Suspicion Register by codes 0004, 0005 (Only those individuals with at least two alcohol-related crimes or suspicion of crimes from both Crime Register and Suspicion Register were included); in the Prescribed Drug Register by the drugs disulfiram (Anatomical Therapeutic Chemical (ATC) Classification System N07BB01), acamprosate (N07BB03), and naltrexone (N07BB04). |
| Autism spectrum disorder (ASD) | The Swedish Hospital Discharge Register (coverage 1973-2017); Outpatient Care Register (national coverage 2001-2017); Primary Care Registry (Partly coverage from 1999-2017) | ICD-9: 299; ICD-10: F840, F841, F845, F849 |

Table 3 - Definition of clinical features

|  |  |
| --- | --- |
| **Clinical Feature** | **Definition** |
| Age at Onset (AAO) | Age at first registration in either of the registers |
| Number of recurrences | Total number of registrations in all registers. Note that registrations within 90 days from the previous registration is not counted. |
| Type of ascertainment 1 (DUD) | Registration in the criminal registers, the medical registers or the prescriptions register. Here we used a hierarchy – (1) criminal, (2) medical, (3) prescription - for individuals with multiple registrations.  Gr 1: 55.2%  Gr2: 37.3%  Gr3: 7.7% |
| Type of ascertainment 2 (DUD) | For individuals with a DUD registration in the medical registers we used a hierarchy and categorized individuals into three groups; (1) at least one registration in the inpatient register, (2) registration in the specialist care, and (3) registration in primary care only.  Gr 1: 63.4%  Gr2: 25.1%  Gr3: 6.6% |
| Type of ascertainment (ADHD) | For individuals with an ADHD registration we used a hierarchy and categorized individuals into three groups; (1) at least one registration in the inpatient register, (2) registration in the specialist care, and (3) registration in primary care only.  Gr 1: 8.2%  Gr2: 82.8%  Gr3: 9.0% |
| Type of ascertainment (MD) | For individuals with a MD registration we used a hierarchy and categorized individuals into four groups; (1) individuals who had a Suicide Attempt at the same day as their MD registration; (2) at least one registration in the inpatient register, (3) registration in the specialist care, and (4) registration in primary care only.  Gr 1: 0.9%  Gr2: 7.7%  Gr3: 24.0%  Gr4: 67.5%  Suicide attempts were defined using the following ICD codes in the Swedish medical registers: ICD-8 codes E950-E959, E980-987  ICD-9 codes E950-E959, E980-987  ICD-10 codes X60-X84, Y10-Y34 |
| Clinical severity for MD | Only measured in individuals registered with MD during the period ICD10 was used (1997 and onwards). We categorized individuals into three groups, also a hierarchy; severe, (ICD10: F32.2 – N: 32,223 (21.2%)); moderate (ICD10: F32.1; N: 79,098 (51.9%)) and mild (ICD10: F32.0 – N: 40.982 (26.9%)) single episode. 152,303 individuals were included in this analysis. |
| Treatment (MD) | Electroconvulsive therapy (ECT) information was only available from 2006 and onwards, so this analysis was only performed among individuals who had an MD registration during this period. ECT was defined using the following operation codes: DA006, DA024, and DA025. Individuals with at least one registration with ECT was categorized as treated.  588,656 individuals were included in this analysis. 7,7706 (1.31%) were treated with ECT |
| Treatment (ADHD) | Stimulant users were defined using the Swedish prescription register. ATC codes: N06BA04, N06BA09, N06BA01, N06BA02. 54,720 (80.3%) were registered as stimulants users and 13,394 (19.7%) as no stimulant users. In total 68,114 individuals were included in this analysis. |
|  |  |

Table 4 -Steps for the Calculation of the FGRS

|  |
| --- |
| 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)  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 trait  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, 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 trait 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 (trait yes/no) into a z-score based on the threshold for each trait. 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 trait registration and relatives without the trait. 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 trait in offspring as outcome and the binary trait 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 each of our 8 main disorders:   |  |  |  | | --- | --- | --- | |  | Parent/Children | Siblings | | MD | 0.80 | 0.85 | | AD | 0.87 | 0.81 | | BD | 0.67 | 0.77 | | SZ | 0.93 | 0.84 | | AUD | 0.99 | 0.69 | | DUD | 0.92 | 0.52 | | ADHD | 0.42 | 0.81 | | ASD | 0.83 | 0.61 | |
| **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 effects (0.0625 - 1) 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. |
| **Note – The authors will be pleased to provide our SAS code to calculate the FGRS to interested investigators.** |

|  |  |  |  |
| --- | --- | --- | --- |
| Table 5 – Testing Stability of FGRS Estimates and Area Under the Surve by Alterations of FGRS Algorithm | | | |
|  | MD FGRS | DUD FGRS | ADHD FGRS |
| **Correlation with the FGRS used in the manuscript** | | | |
| FGRS(a) - 1st degree relatives only | 0.760 | 0.763 | 0.731 |
| FGRS(b) - no age correction | 0.949 | 0.995 | 0.999 |
| FGRS(c) - no cohabitation correction | 0.961 | 0.980 | 0.991 |
| FGRS(d) - no weighting for # relatives | 0.912 | 0.971 | 0.974 |
| FGRS(e) -std by YoB only | 0.973 | 0.990 | 0.983 |
| FGRS(f) - std by geography only | 0.955 | 0.951 | 0.948 |
| FGRS(g) - std only by entire sample | 0.935 | 0.946 | 0.934 |
|  |  |  |  |
| **Area Under the Curve** | | | |
| FGRS used in the manuscript | 0.587 (0.586; 0.588) | 0.653 (0.652; 0.654) | 0.631 (0.629; 0.634) |
| FGRS(a) - 1st degree relatives | 0.575 (0.574; 0.575) | 0.608 (0.607; 0.610) | 0.570 (0.567; 0.572) |
| FGRS(b) - no age correction | 0.589 (0.588; 0.590) | 0.657 (0.656; 0.659) | 0.642 (0.640; 0.644) |
| FGRS(c) - no cohabitation correction | 0.588 (0.587; 0.590) | 0.657 (0.655; 0.658) | 0.641 (0.639; 0.643) |
| FGRS(d) - no weighting for # relatives | 0.585 (0.585; 0.586) | 0.648 (0.646; 0.649) | 0.621 (0.618; 0.623) |
| FGRS(e) -std by YoB only | 0.601 (0.600; 0.602) | 0.657 (0.656; 0.659) | 0.638 (0.636; 0.641) |
| FGRS(f) - std by geography only | 0.584 (0.583; 0.585) | 0.665 (0.664; 0.667) | 0.703 (0.701; 0.705) |
| FGRS(g) - std only by entire sample | 0.598 (0.597; 0.599) | 0.671 (0.670; 0.672) | 0.717 (0.715; 0.719) |

**Table 6 - 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: {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. , where 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 (),
      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 , where are independent standard normal (Gaussian) variates (j=1,…m),
         2. Simulate the common environment for all sibs within family as, , with a single Gaussian variant for entire sibship,
         3. Simulate the independent environment for each sib within family as, , where are independent standard normal (Gaussian) variates (j=1,…m),
         4. Compute liability for each sib as
         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).

Figure 1 – The Chance in Population Prevalence for DUD, ADHD and MD as a Function of the FGRS Score Ranging from 0 (population mean) to + 2 (2 SDs above the mean)

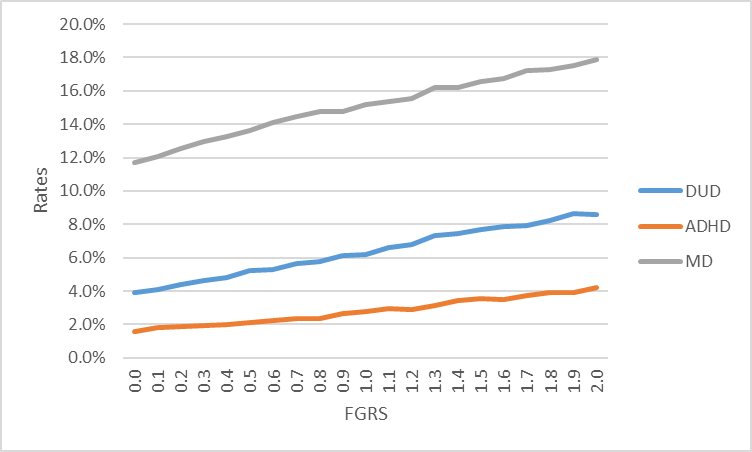


Figure 2 – Age at First Registration for ADHD

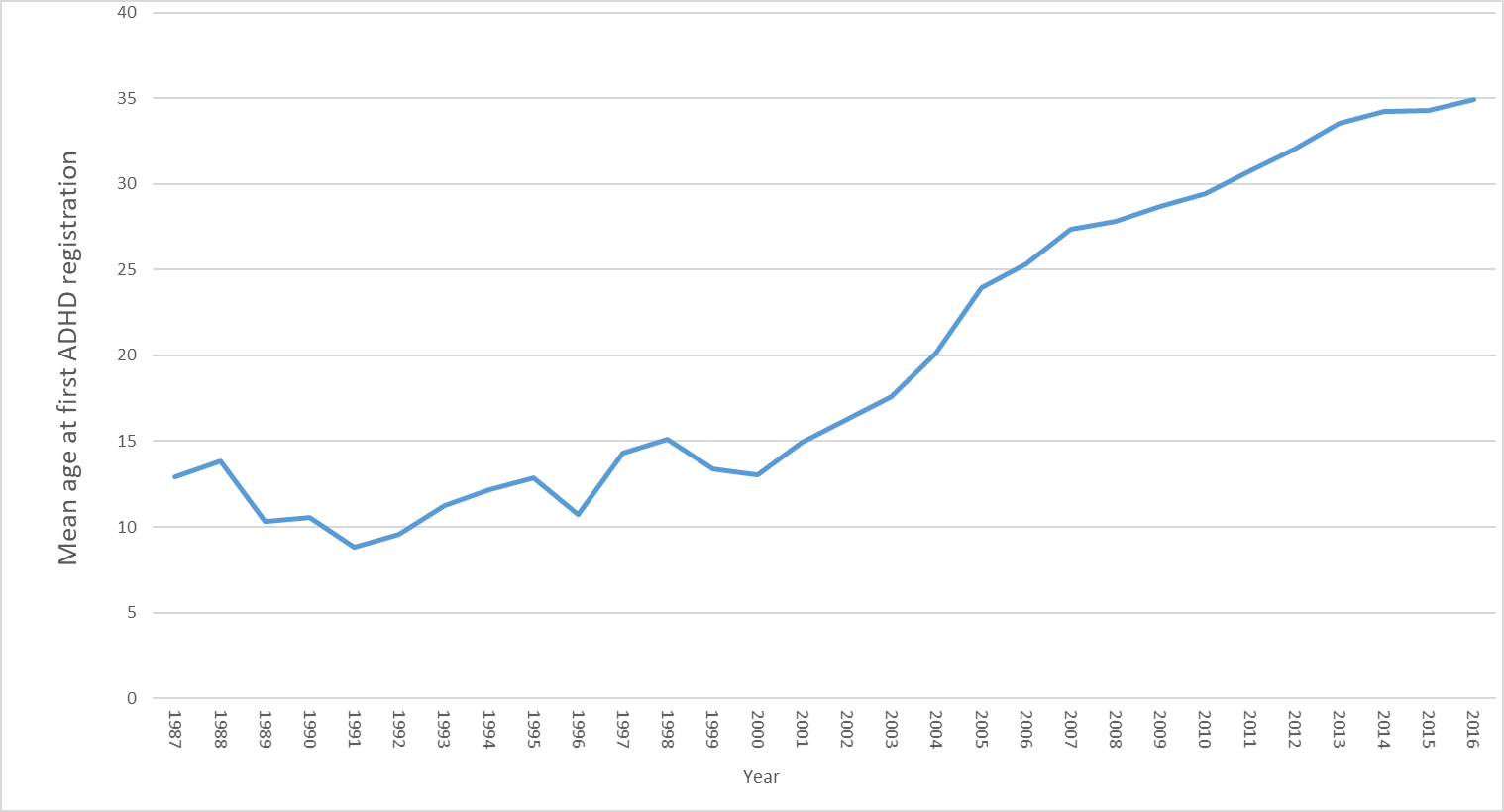


Figure 3- Stability of FGRS Scores for MD, DUD and ADHD by Median Splits for Cohort and Geographical Region within Sweden

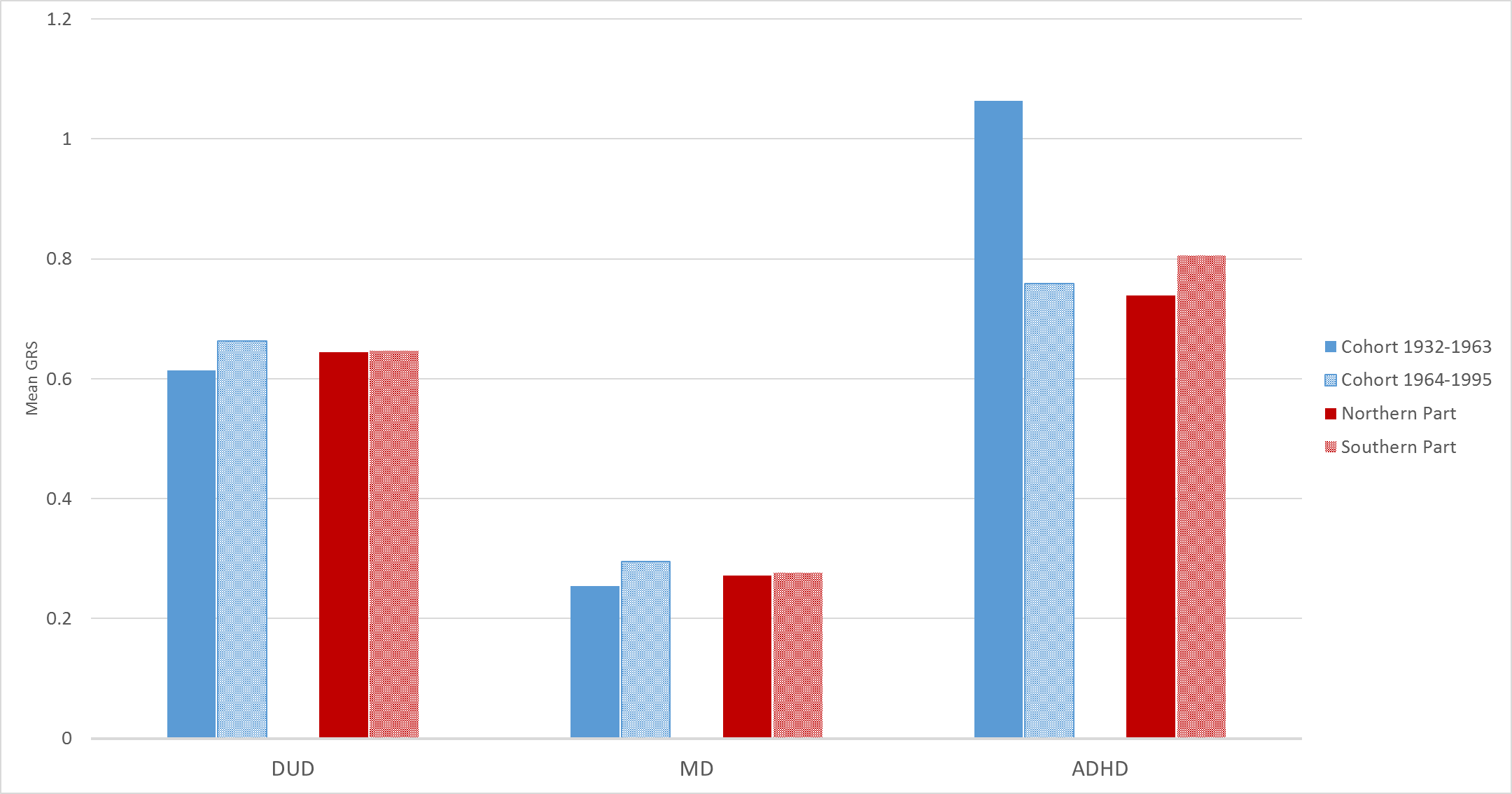


Figure 4a – Results of Simulations of Pedigrees Containing 1st-5th Degree Relatives Analyzed by FGRS as a Function of Heritability and Prevalence

For Figures 4a-4d – see above for simulation methods

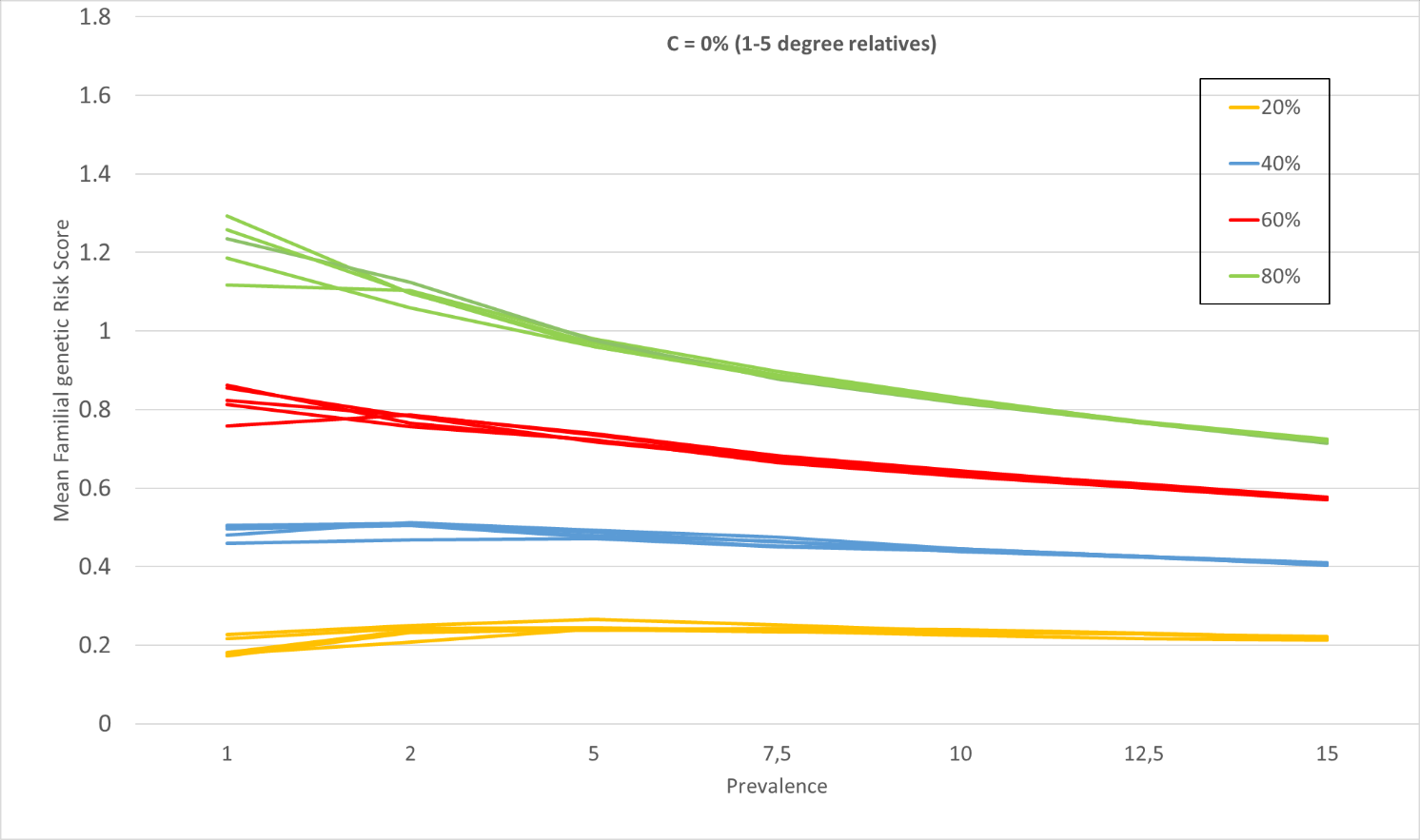


Figure 4b

For Figures 3s-e, 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. The down-weighting values used in this paper are seen above in table 6 step 3.

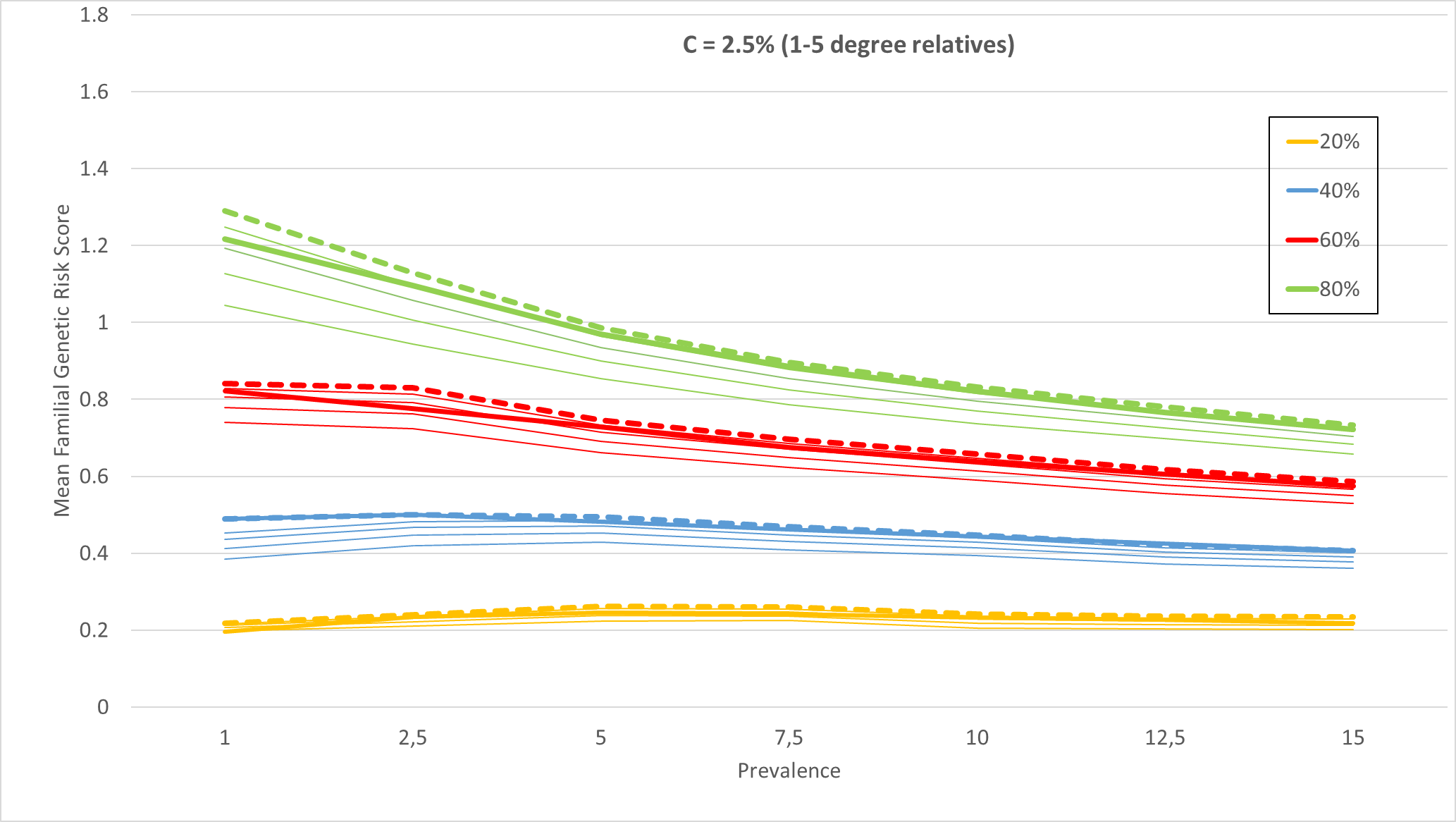
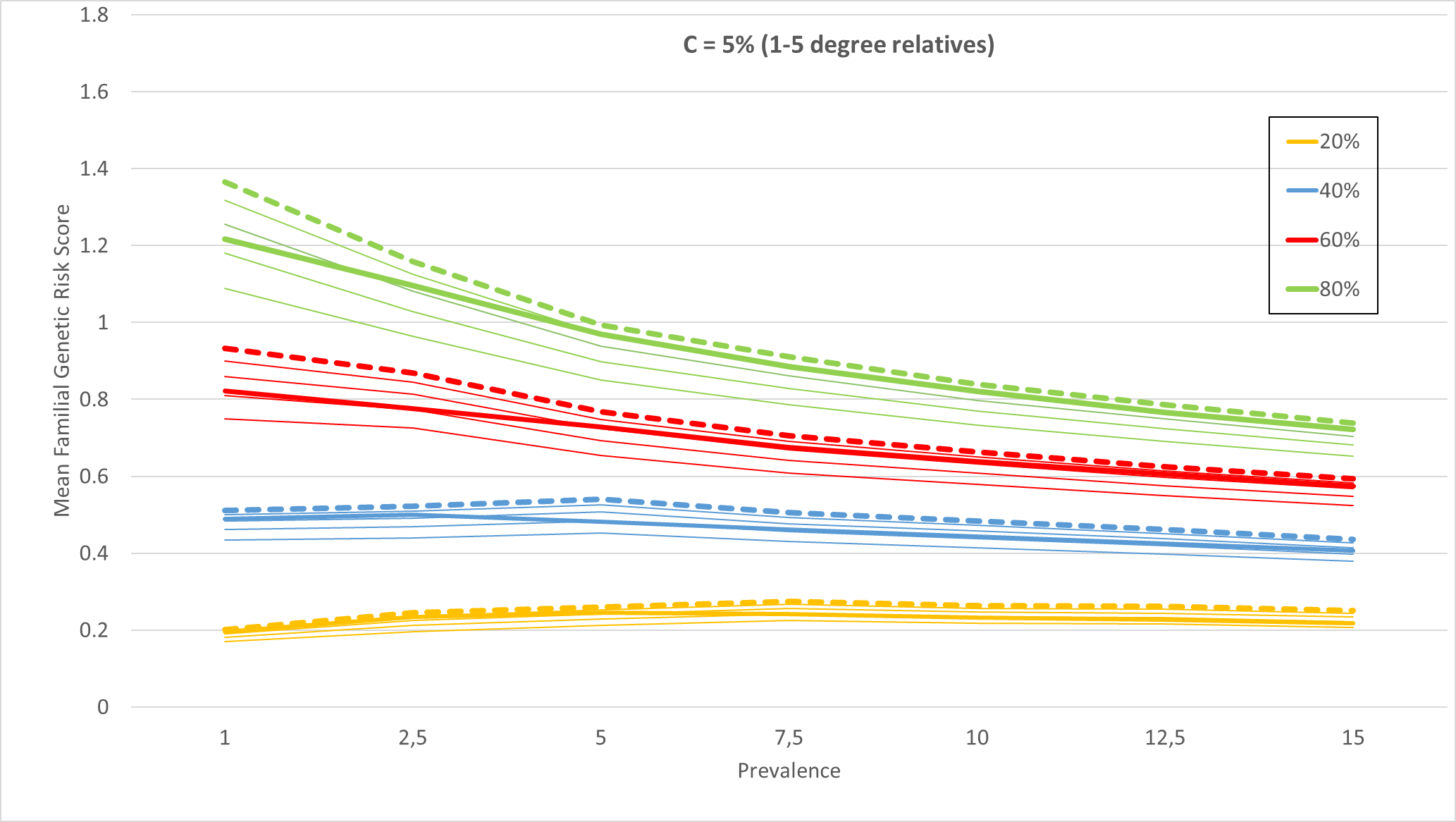


Figure 4c

Figure 4d

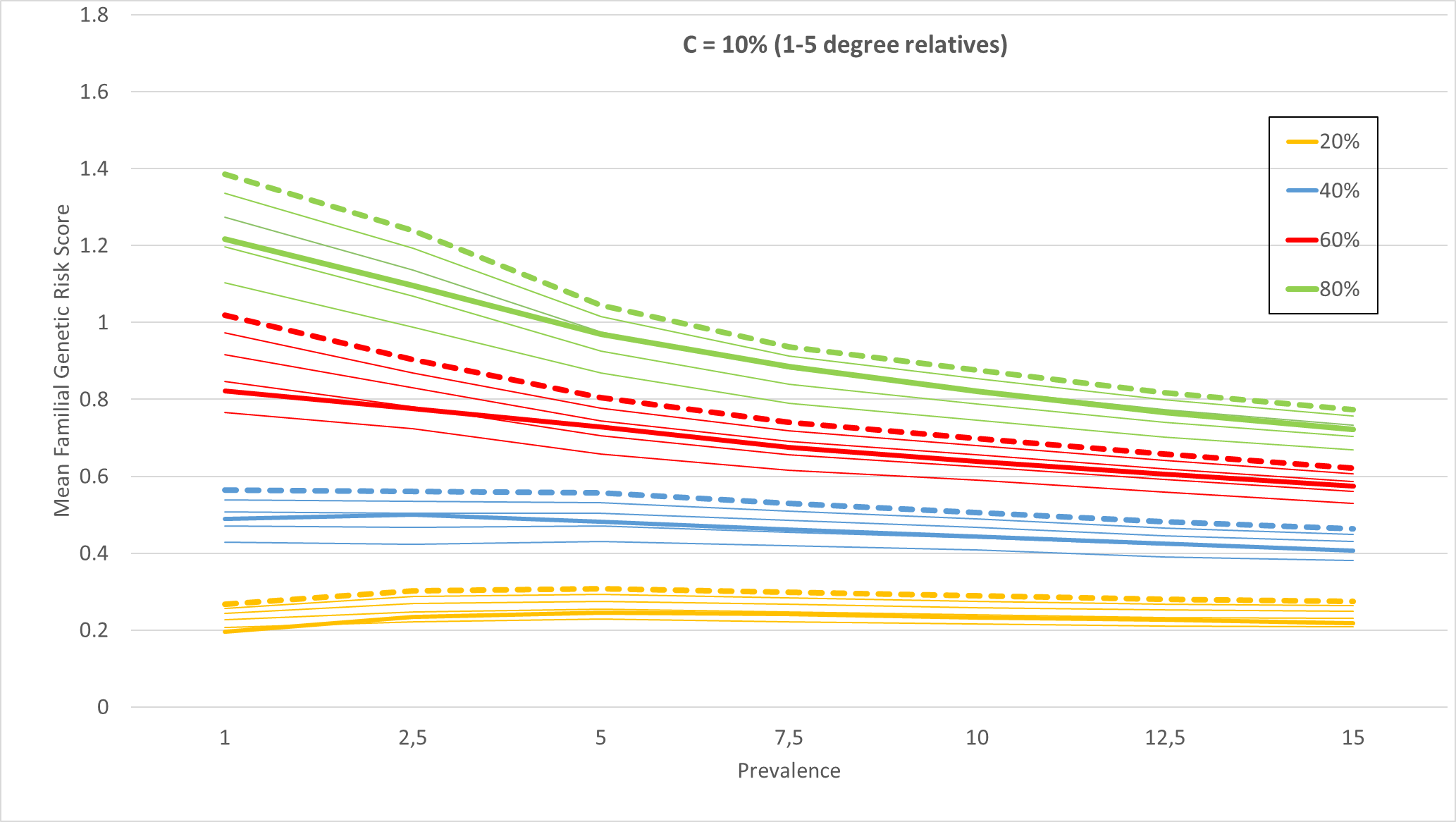


Figure 5-9 – Results depicted in figures 1-5 reanalyzed controlling for the FGRS of the Primary Disorder – that is, DUD, ADHD and MD.

Figure 5a Sex DUD

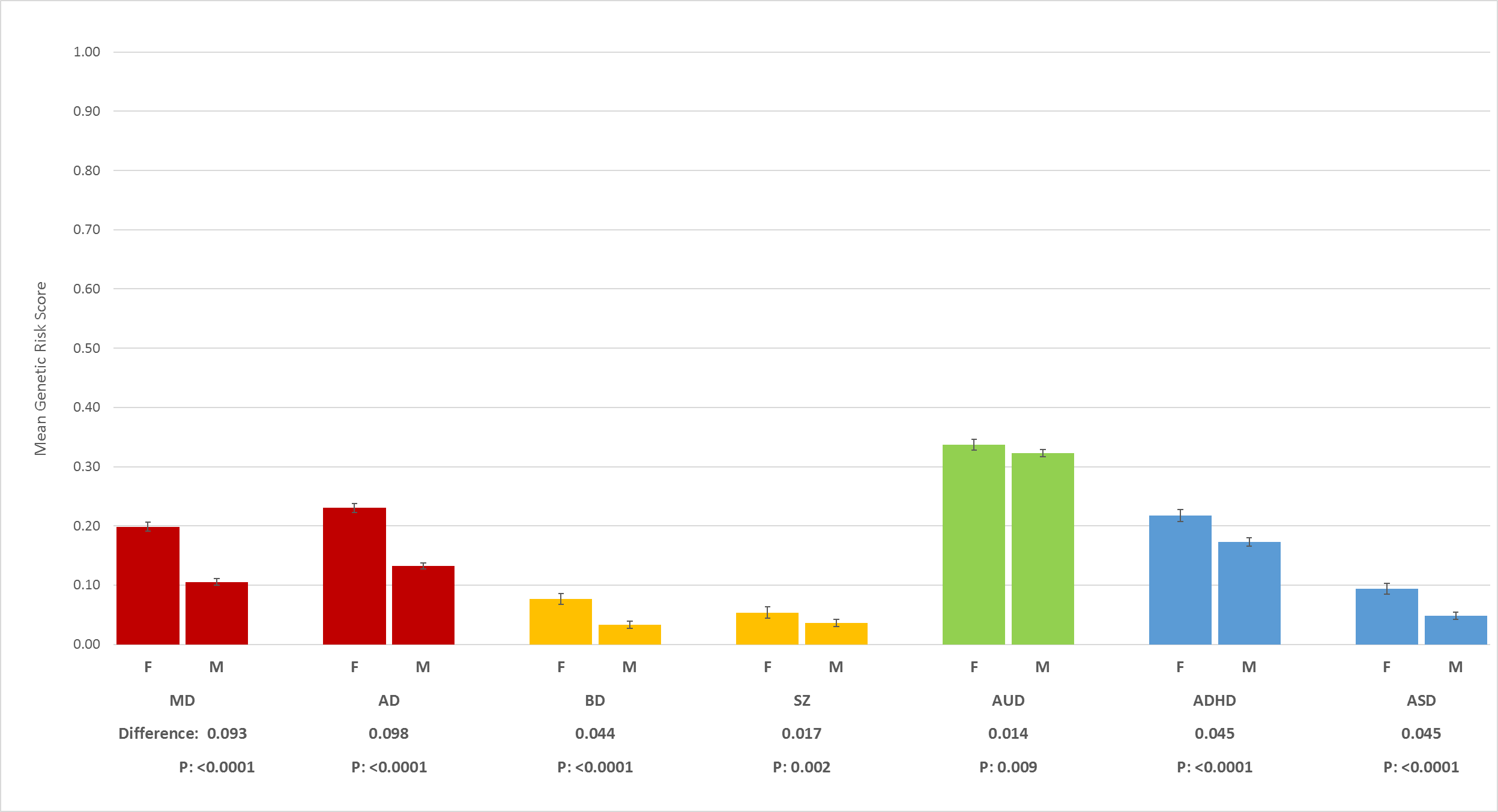


Figure 5b Sex MD

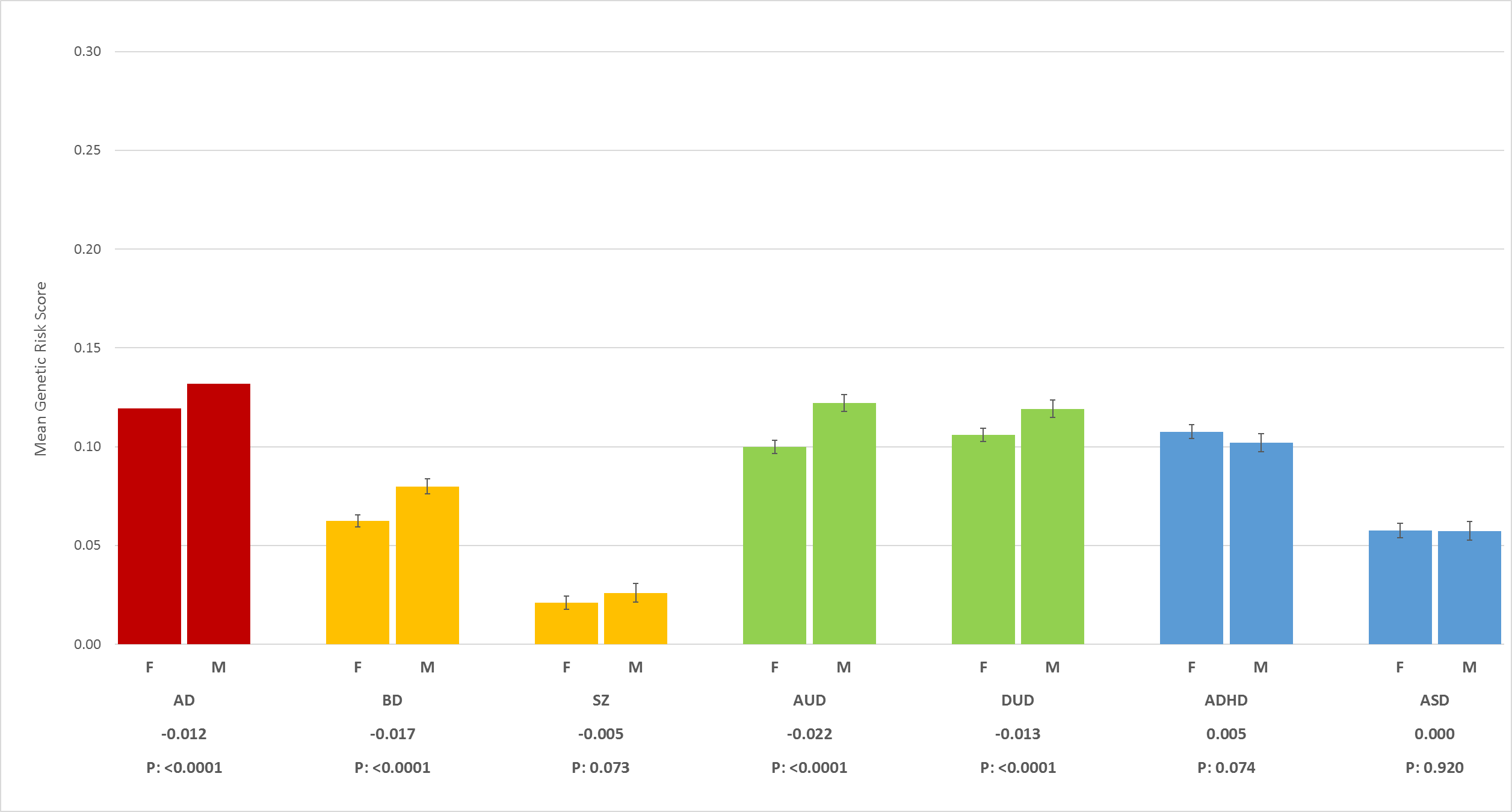


Figure 5c – Sex ADHD

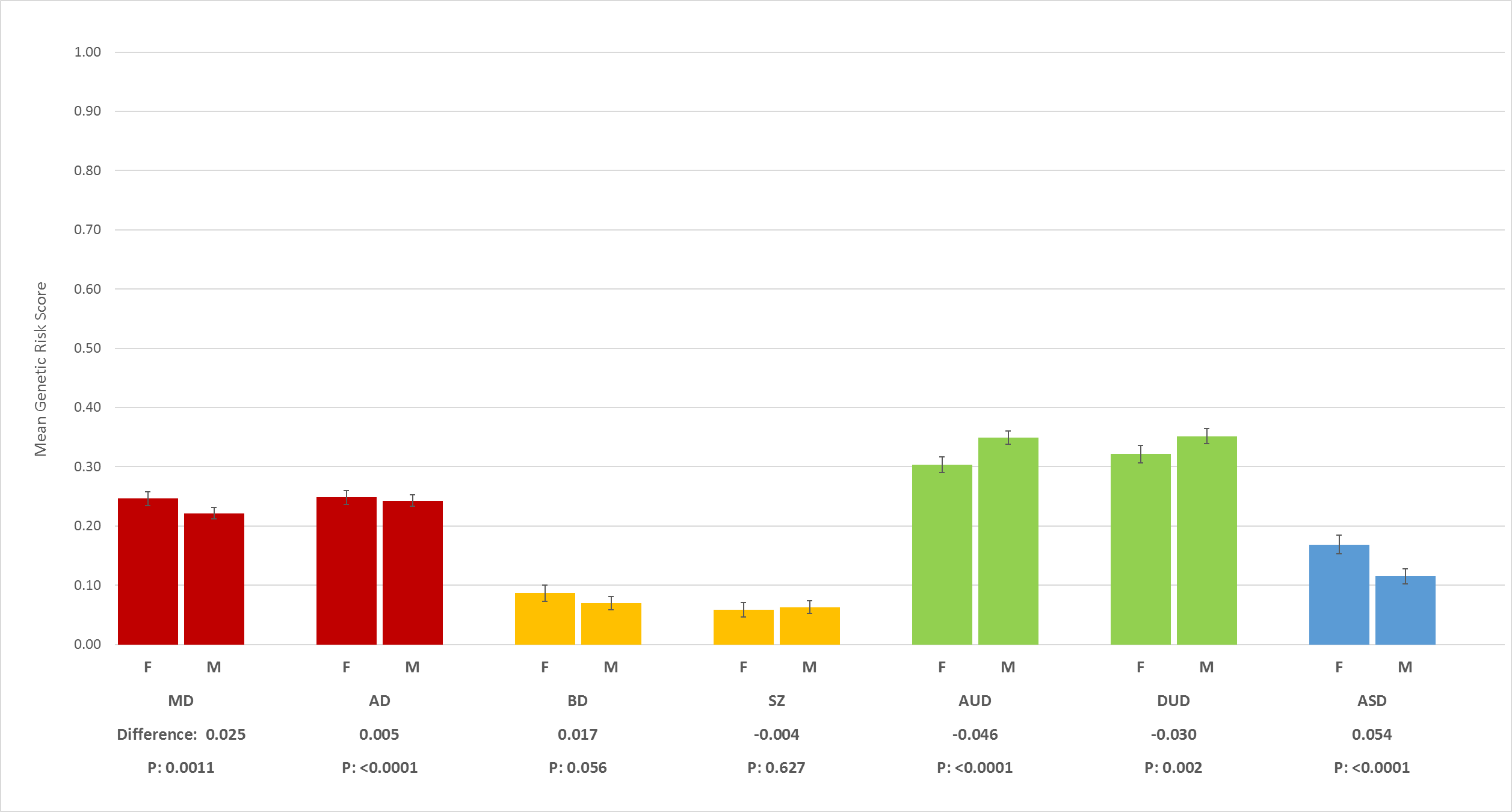


Figure 6a - Age at Onset DUD

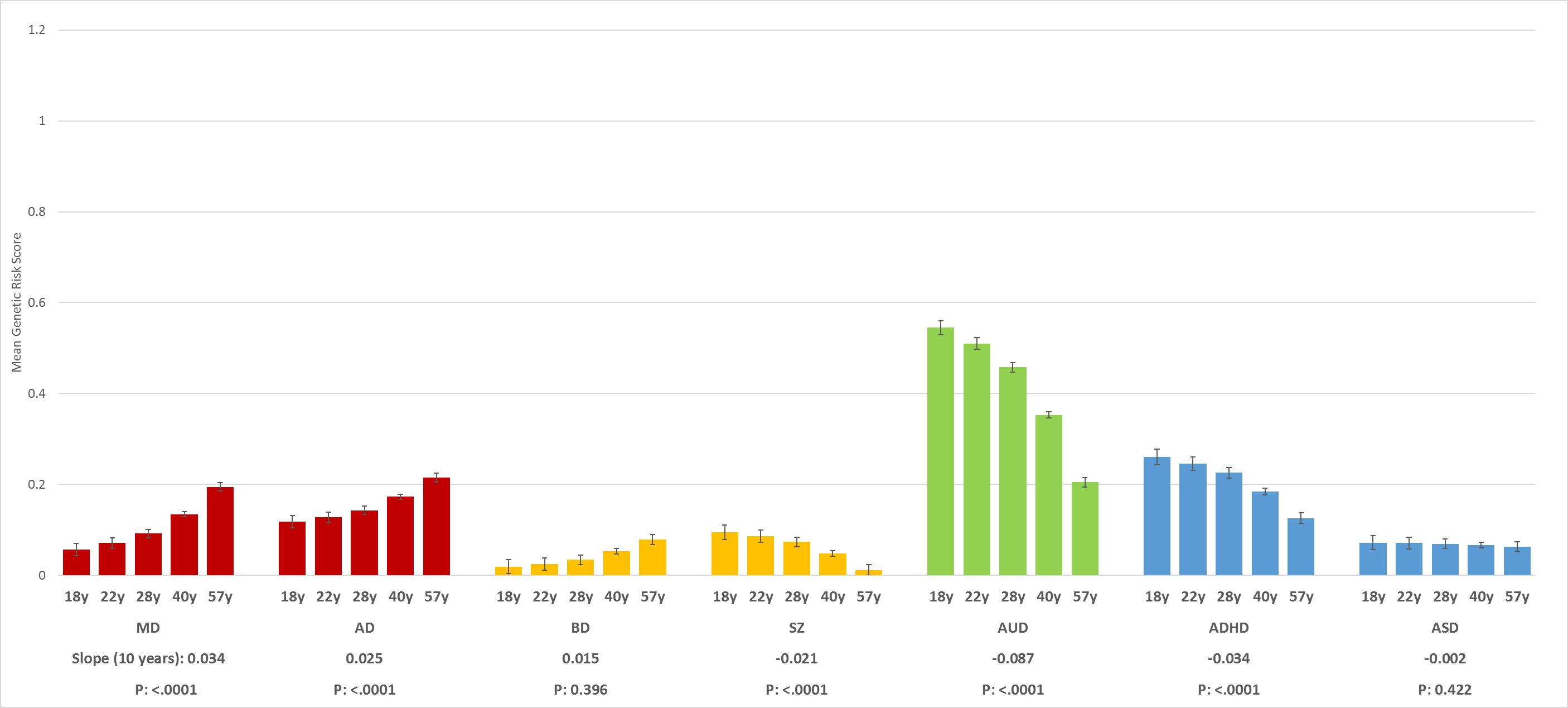


Figure 6b AAO MD

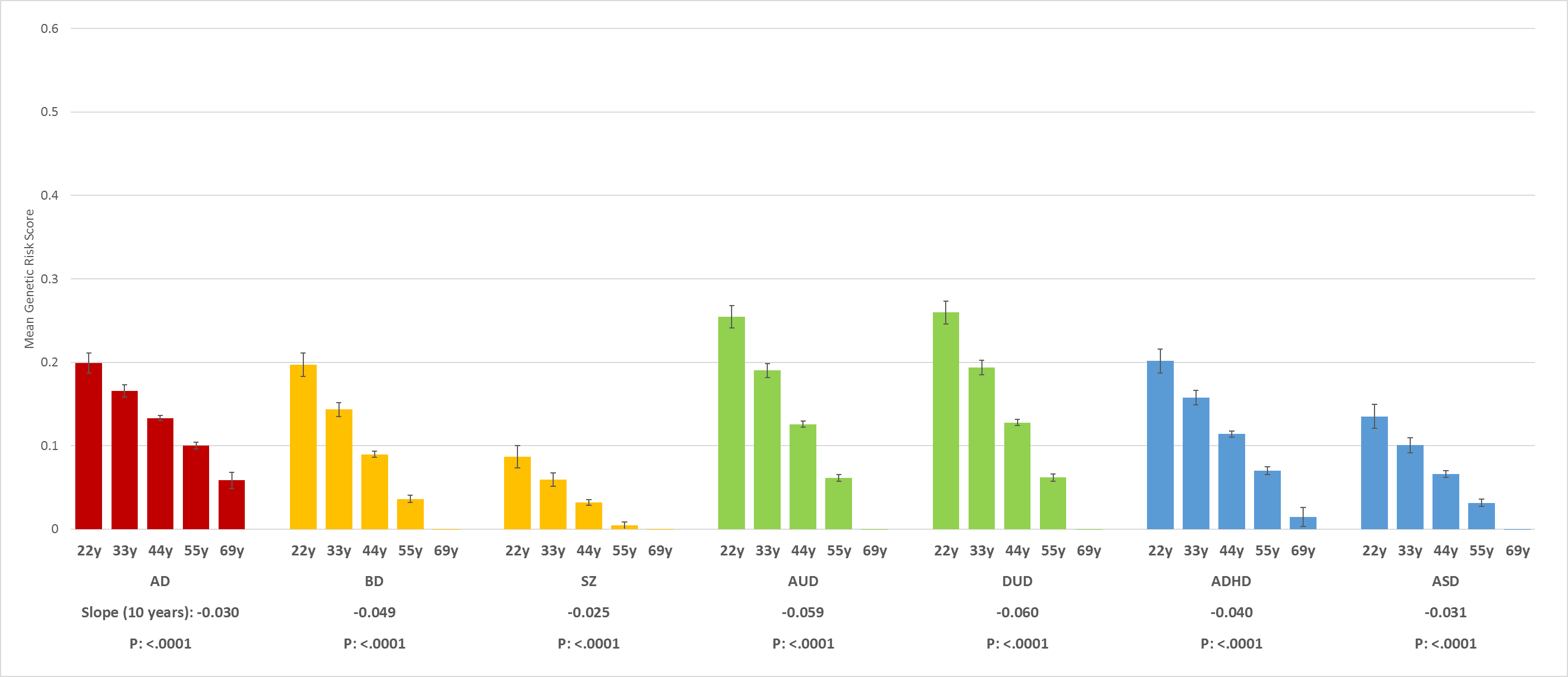


Figure 6c – AAO - ADHD

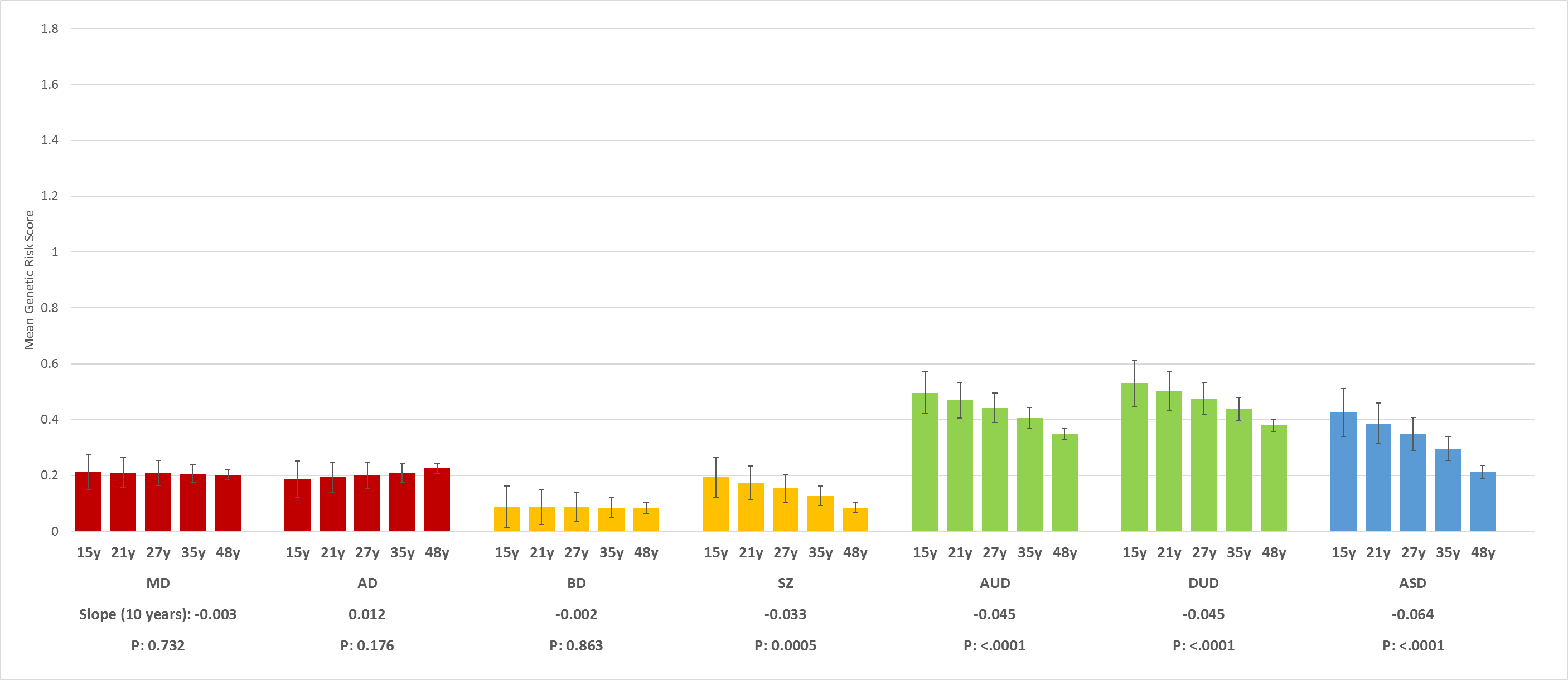


Figure 7a # Episodes DUD

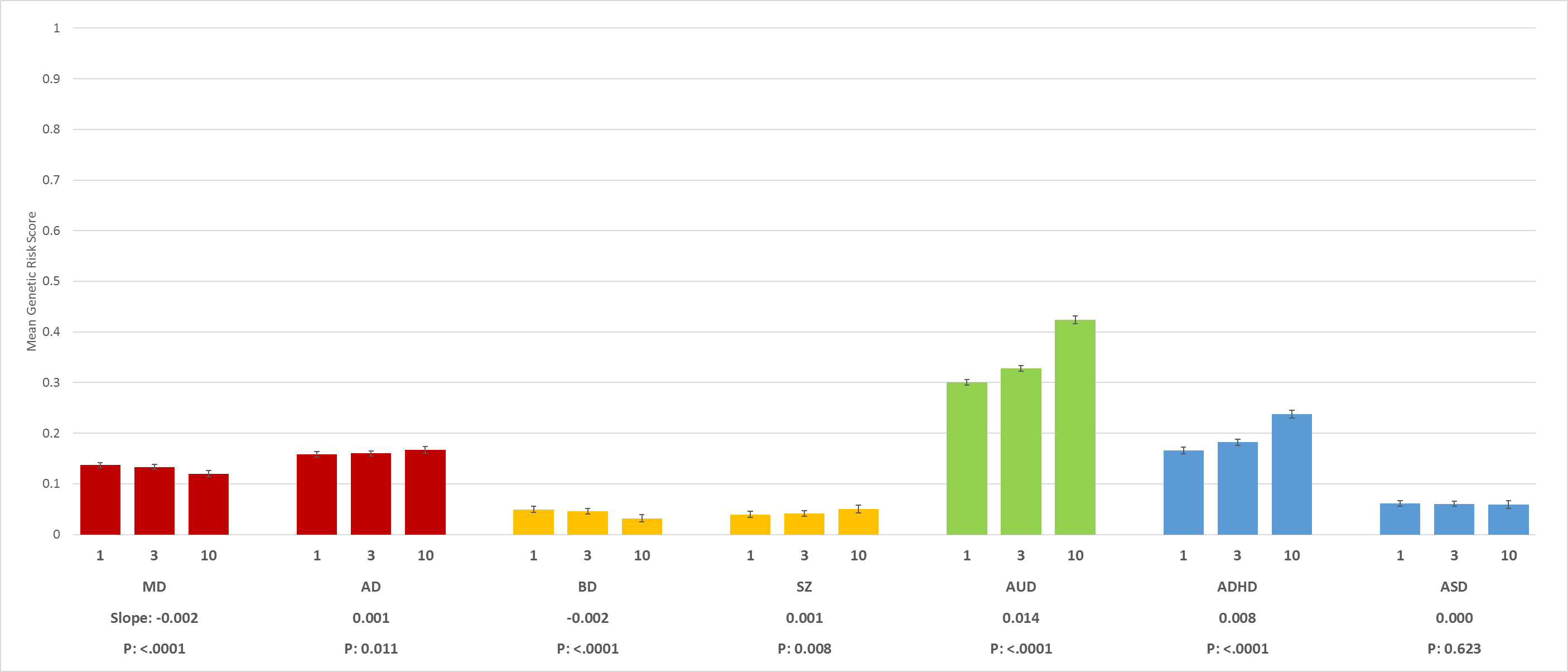


Figure 7b # Episodes MD

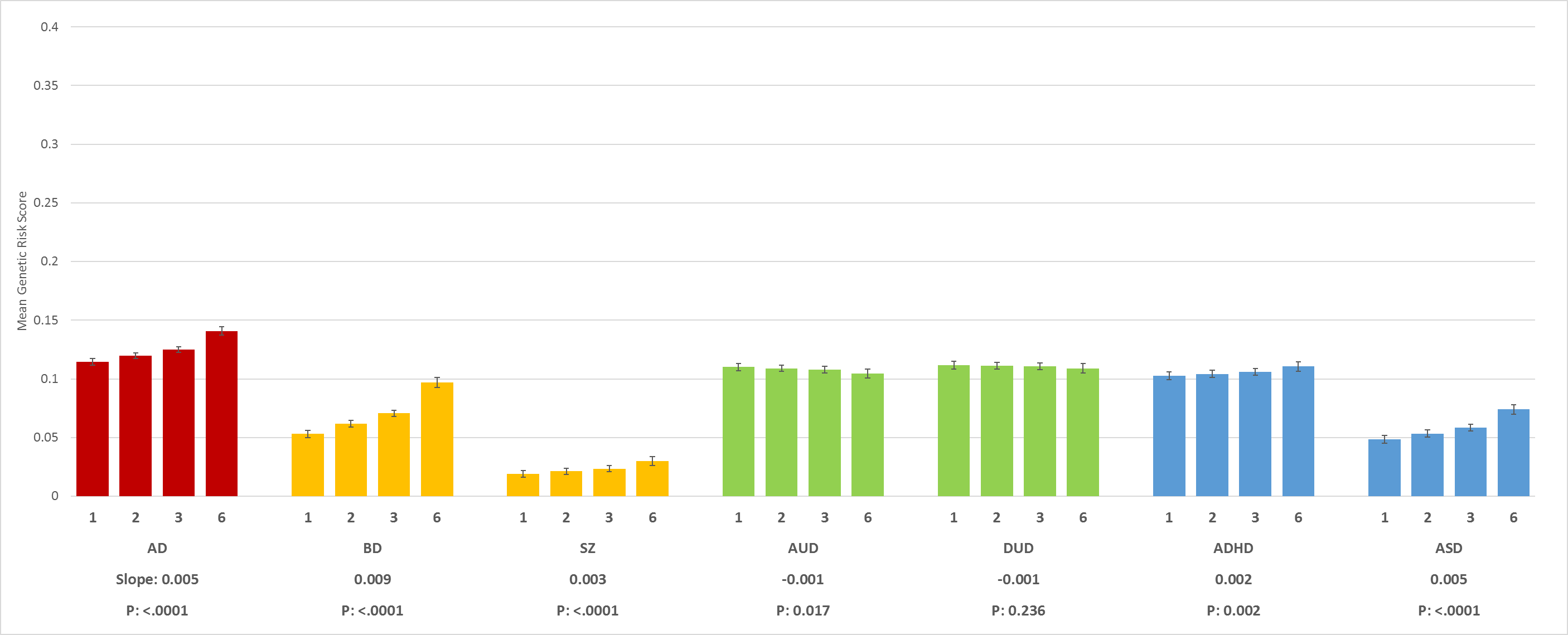
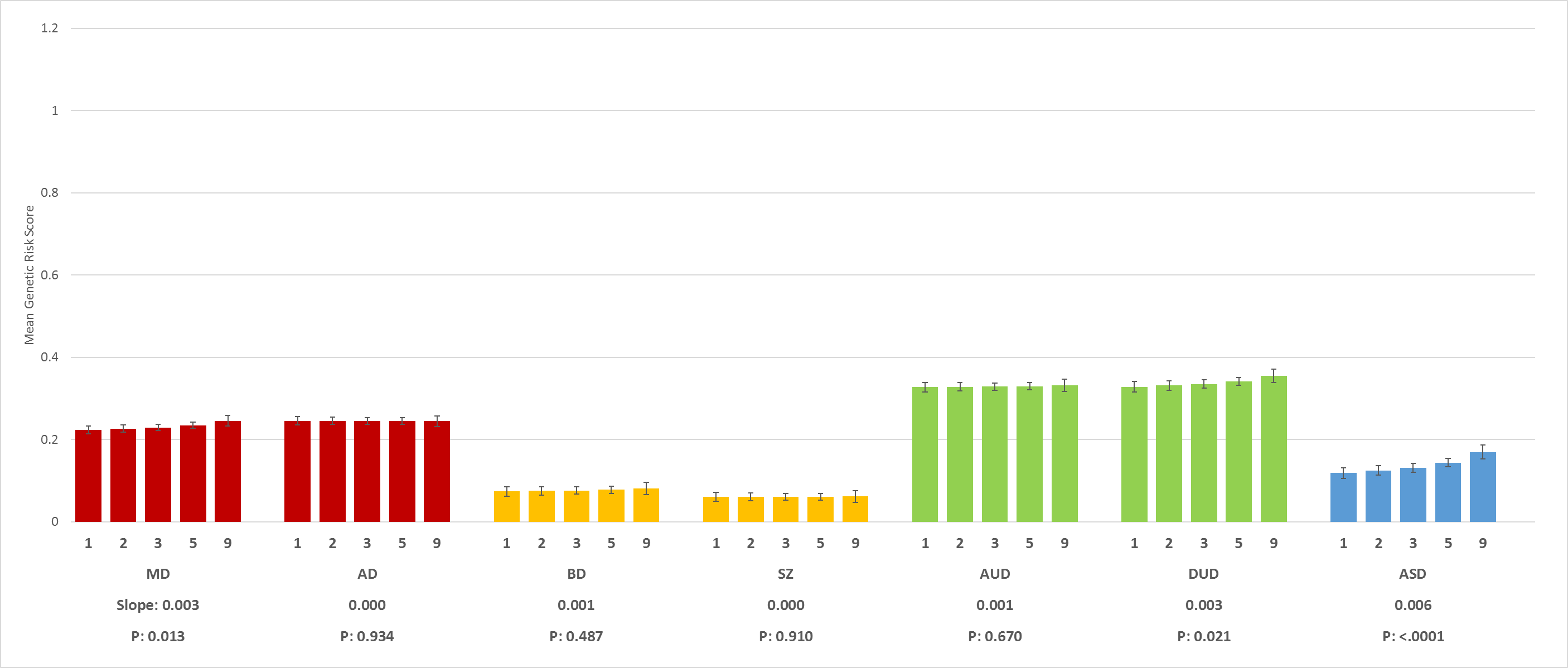
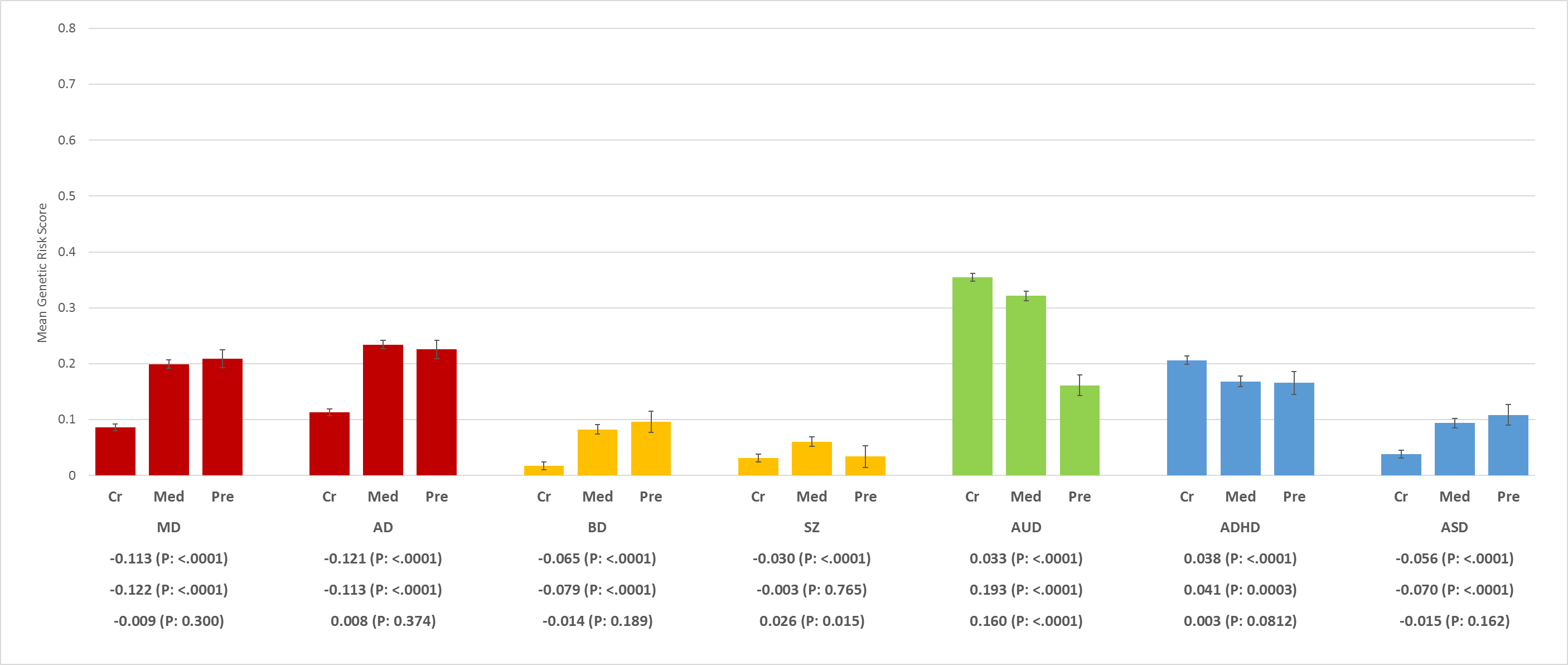


Figure 7c # Episodes ADHD



Figure 8a – Source of Ascertainment DUD By Register

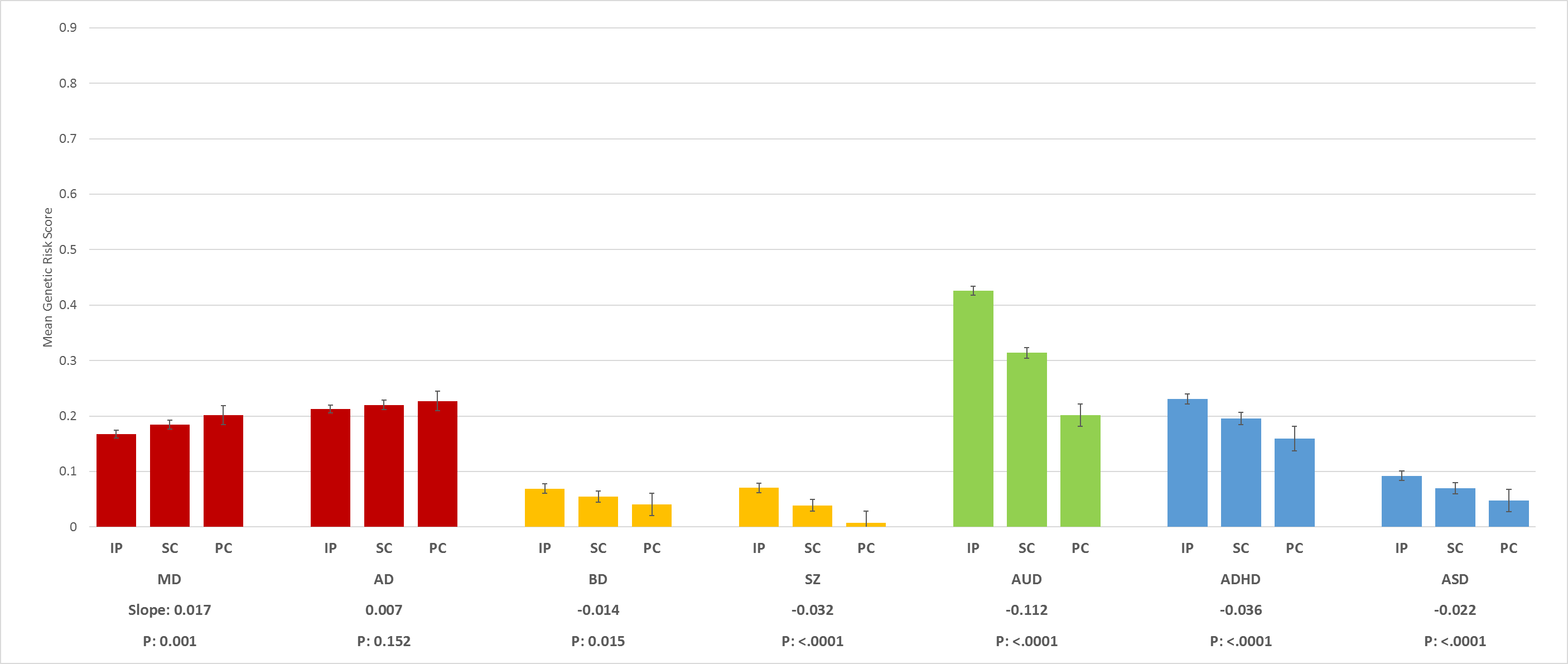
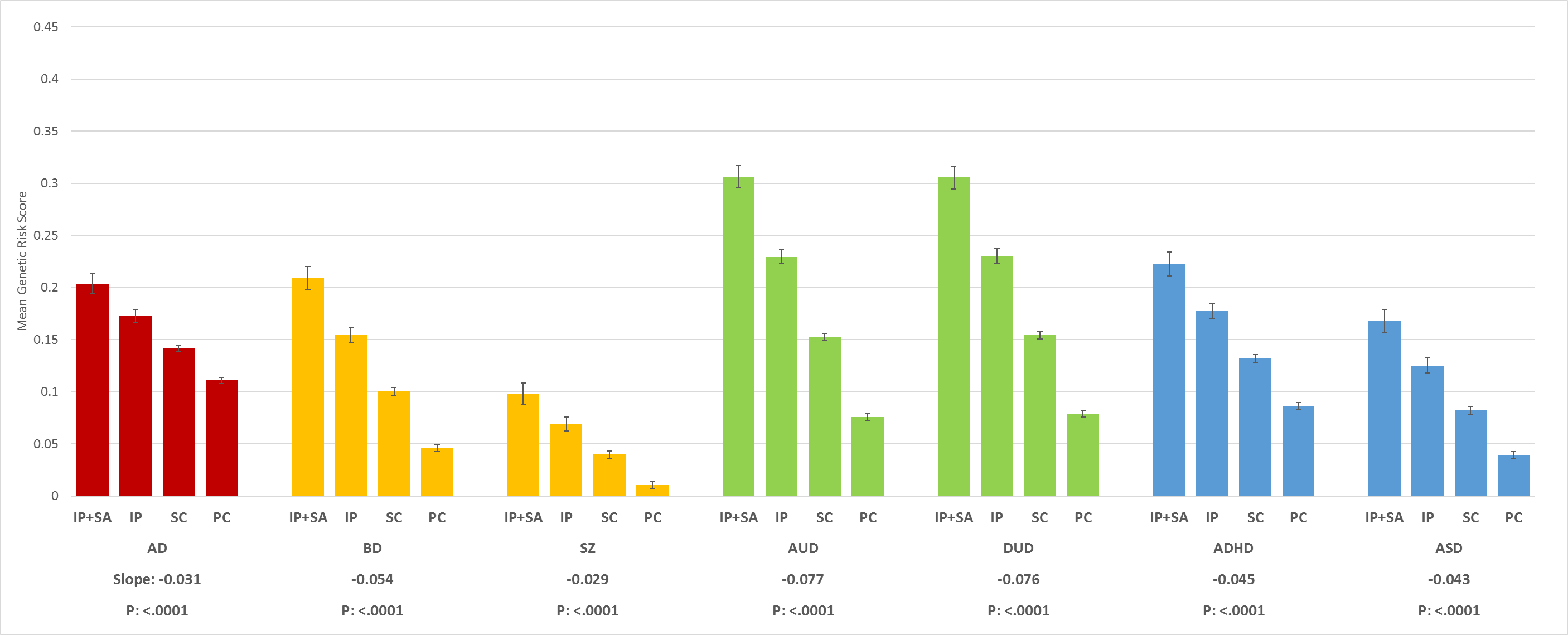
Figure 8b Source of Ascertainment for DUD within Medical Register

Figure 8c – Source of Ascertainment MD



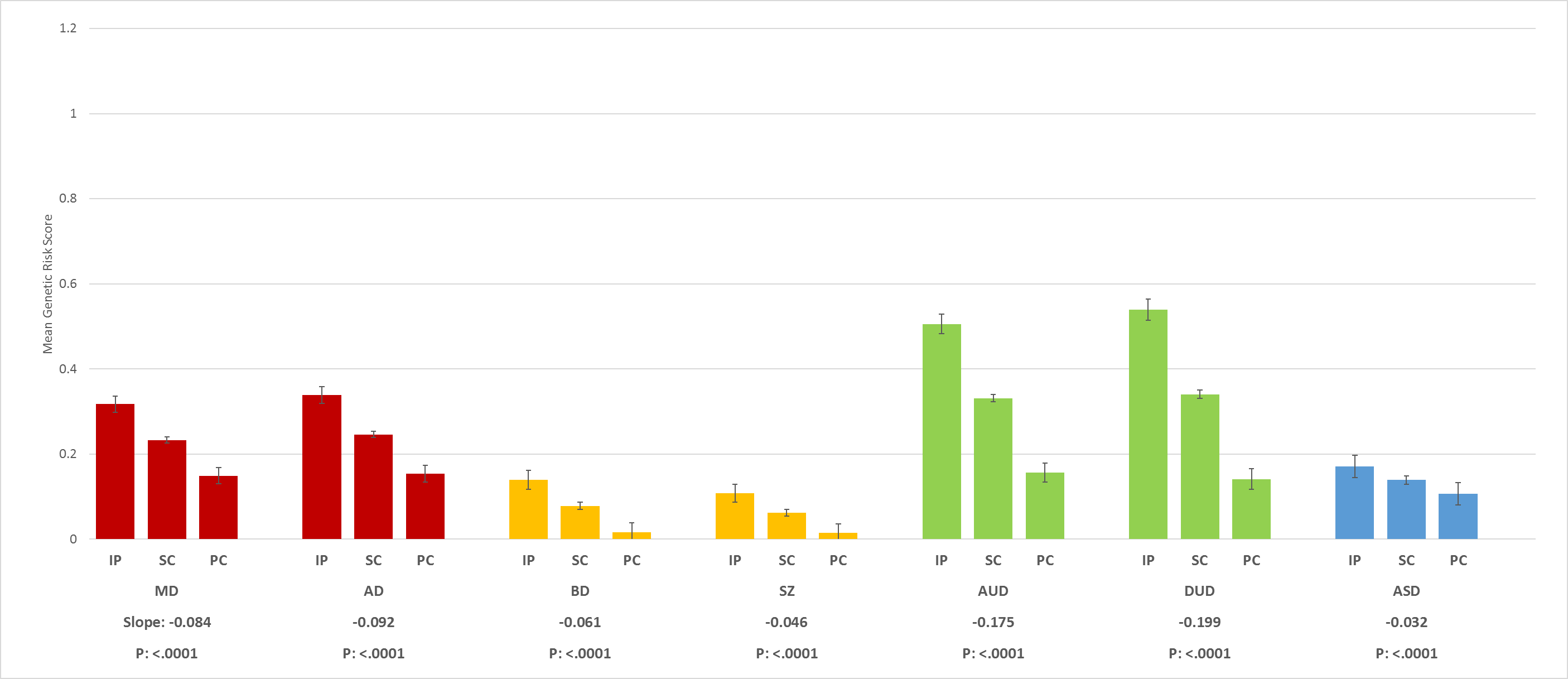
Figure 8d – ADHD

Figure 9a MD – Clinical Severity by ICD-10 Code

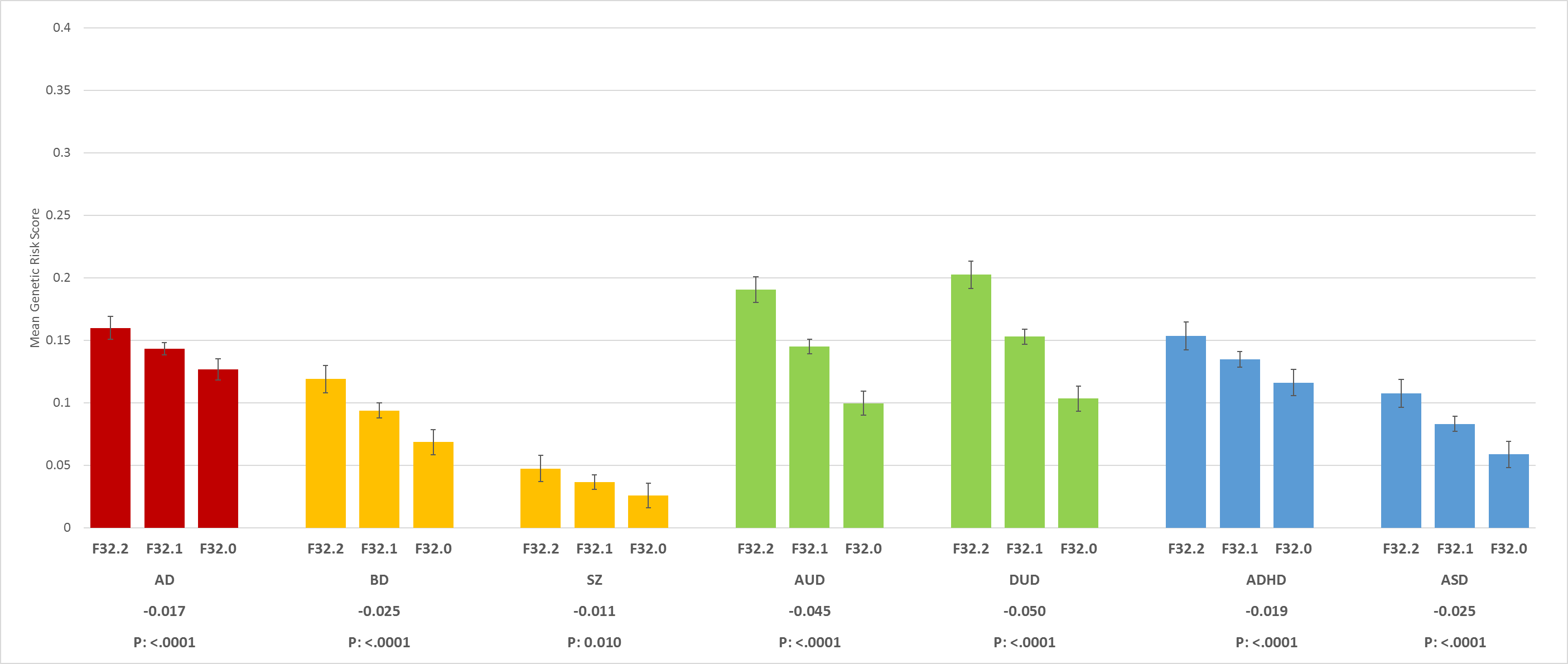
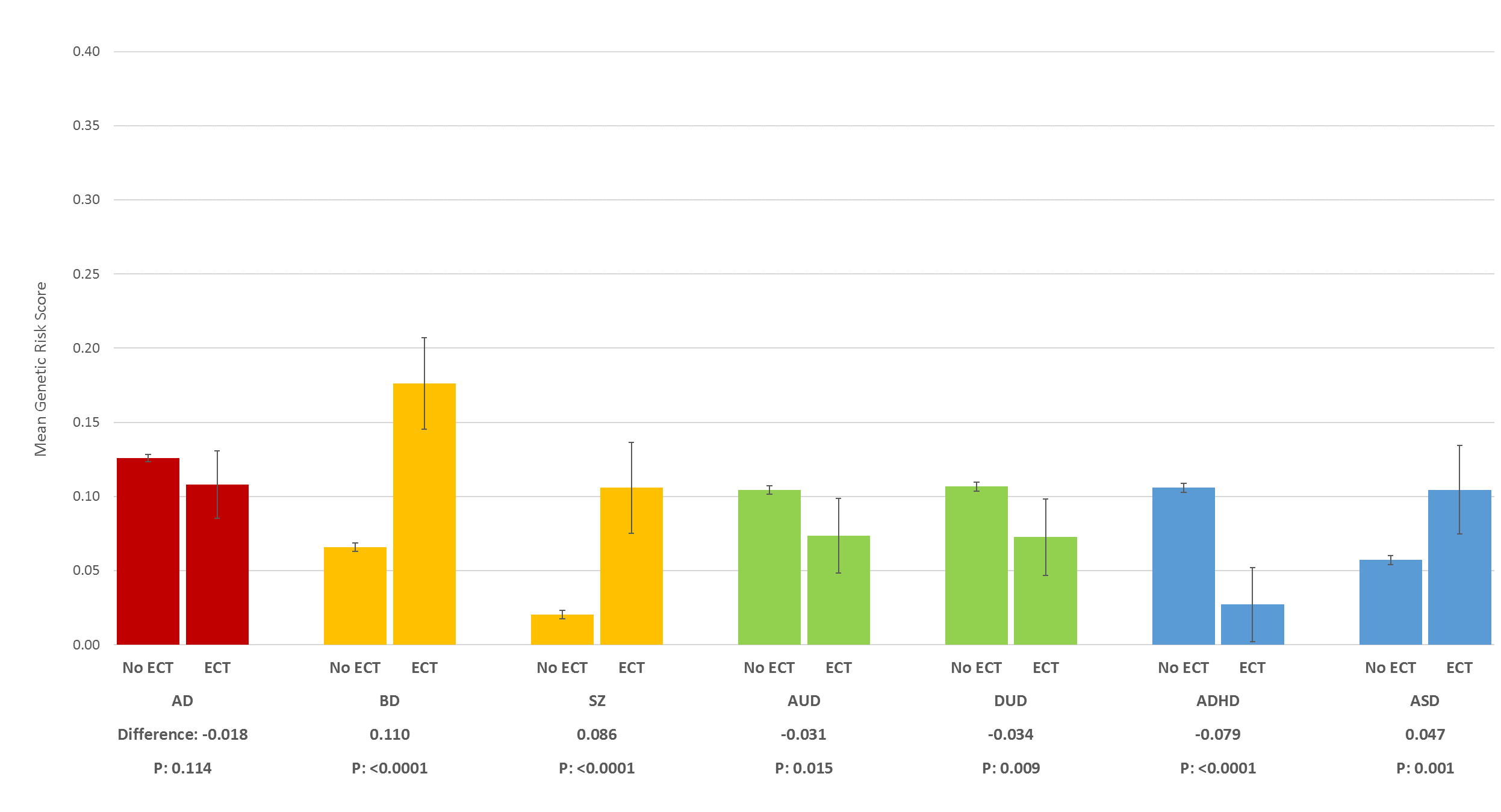
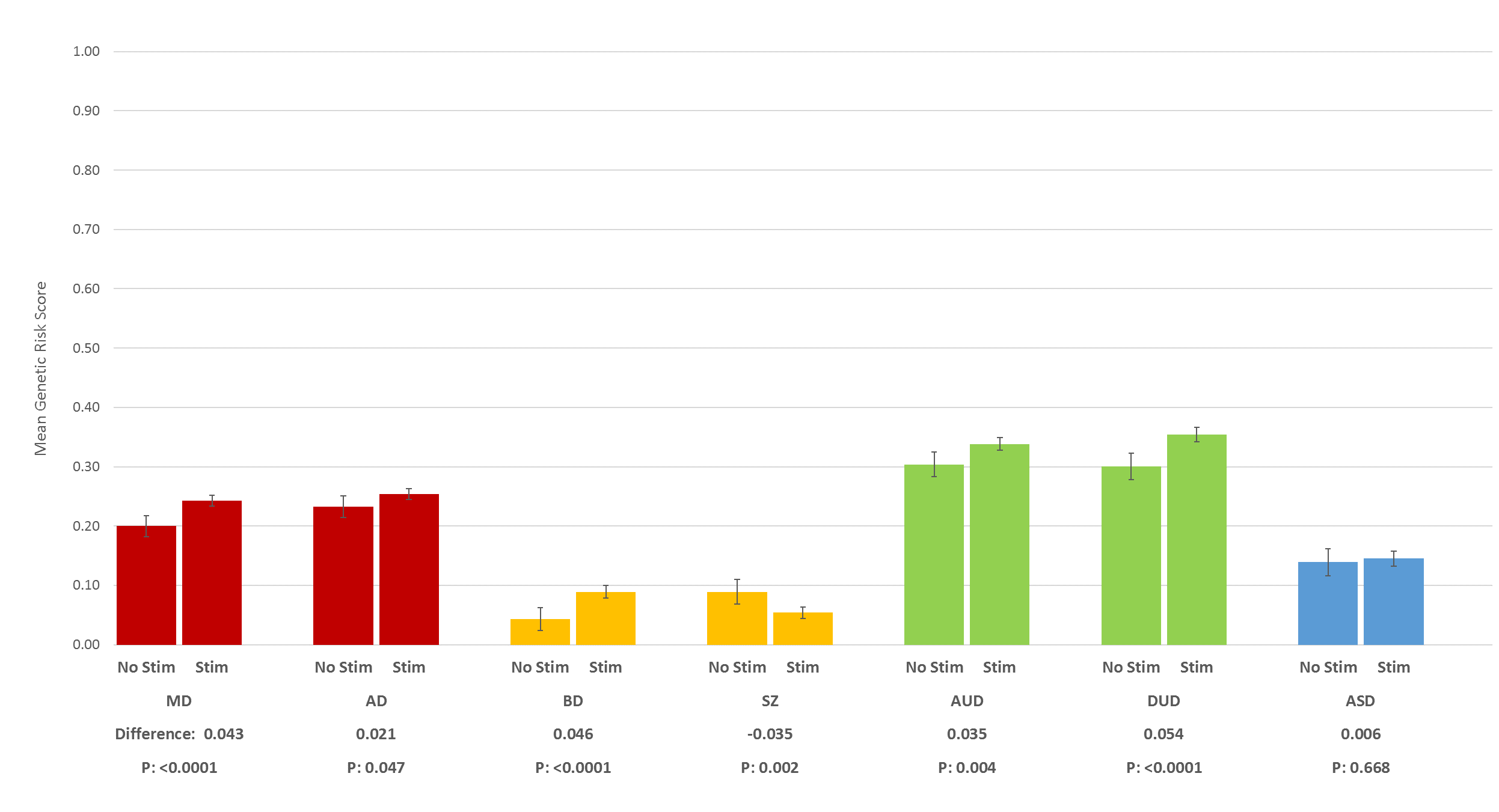


Figure 9b MD by ECT



Figure 9c – ADHD - Treatment with Stimulants