

Mulighederne ved at kombinere registerdata med biologiske data

Preben Bo Mortensen, professor, PI iPSYCH
Jeppe Klok Due, specialkonsulent, KOR

Udfordringer for registerforskning

1. Kobling af registre og store datamængder
2. Mangel på gensidig tillid til analysemiljøer
3. Decentral infrastruktur
4. Data spredt hos mange myndigheder

Samarbejde med centrale aktører

KOR har initieret et samarbejde, der giver eksempler på en løsning sammen med

- ▶ Danmarks Statistik
- ▶ Computerome Center, DTU
- ▶ DeiC

Finansiering: DeiC og Styrelsen for Forskning og Uddannelse

Dogmer for samarbejdet

Forskere skal kunne kombinere registre med store datamængder *samtidig* med at dataejer har fuld kontrol med data og brugere

Løsningen tilpasses dataejerens behov, så dataejer kan være tryk ved systemet

Maksimere de forskningsmæssige fordele, og minimere risikoen for de behandlede

Proof of Concept

- ▶ Sikker forbindelse mellem DST-Computerome
- ▶ DST virtuel maskine på Computerome
 - Kun adgang til at læse data via DST
 - Brugerstyring
 - Logning
 - Kontrol af download
 - Egne programmer
- ▶ Adgang til store beregnings- og lagerfaciliteter på Computerome

Use case med iPSYCH

iPSYCH har konkret behov for analyse af registre og store data mængder

- ▶ Teste om DSTs kan være trygge
- ▶ Teste om forskere får de muligheder, de efterspørger
- ▶ Indhente konkrete erfaringer

iPSYCH: verdens største studie af årsager til psykisk sygdom

- ▶ Genetiske og miljømæssige årsager til Skizofreni, Bipolar sygdom, Depression, Autisme og ADHD
- ▶ Samarbejde mellem 6 centre på AU, Region Midt, KU, Region H og SSI
- ▶ Mere end 150 forskere, mange udenlandske samarbejdspartnere
- ▶ Baseret på de mange danske registre samt hælblodprøver fra danske nyfødte siden 1981
- ▶ Omfatter nu mere end 80,000 personer, forventes udvidet til det dobbelte
- ▶ Finansieret af Lundbeck Fonden, institutionerne og andre store danske og udenlandske fonde og samarbejdspartnere

iPSYCH: verdens største studie af årsager til psykisk sygdom

Behov:

- ▶ Sikkerhed
- ▶ Fleksibel adgang til analyse fra hele verden
- ▶ Tilstrækkelig Beregningskapacitet
 - Udfordring. Mange vigtige registerdata kan kun tilgås hos Danmarks Statistik
 - Genetiske data fylder MEGET mere end registerdata
 - Selv med en (efter registermålestok) relativt stor dedikeret server hos DST (144 CPU'er, 512GB hukommelse og 34TB diskplads) er det kun muligt for 1-2 forskere samtidigt at regne på mindre end 5% af de genetiske data ad gangen.

Figure 1. Proportion of cohort members developing ADHD over time stratified by RIA-score

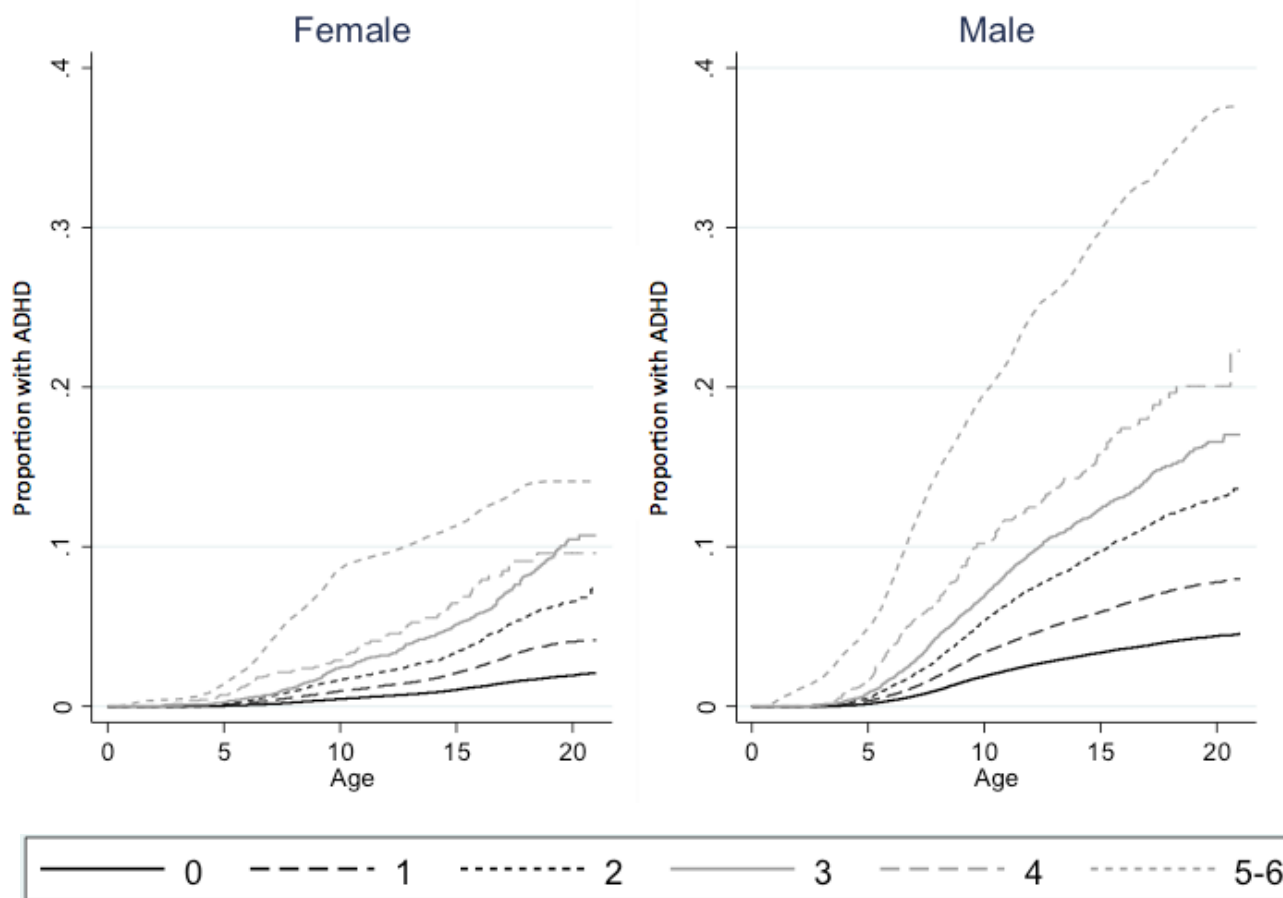
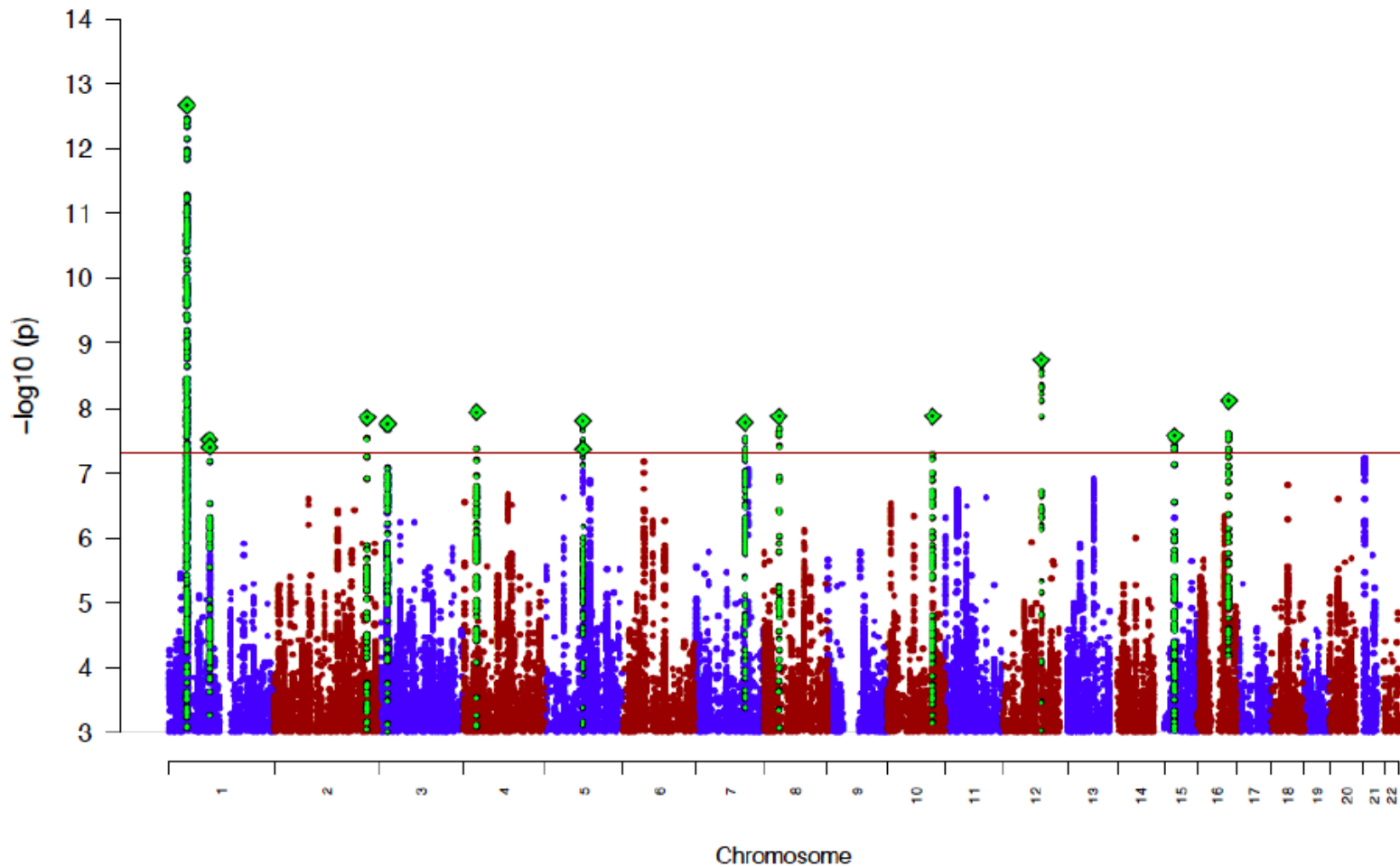
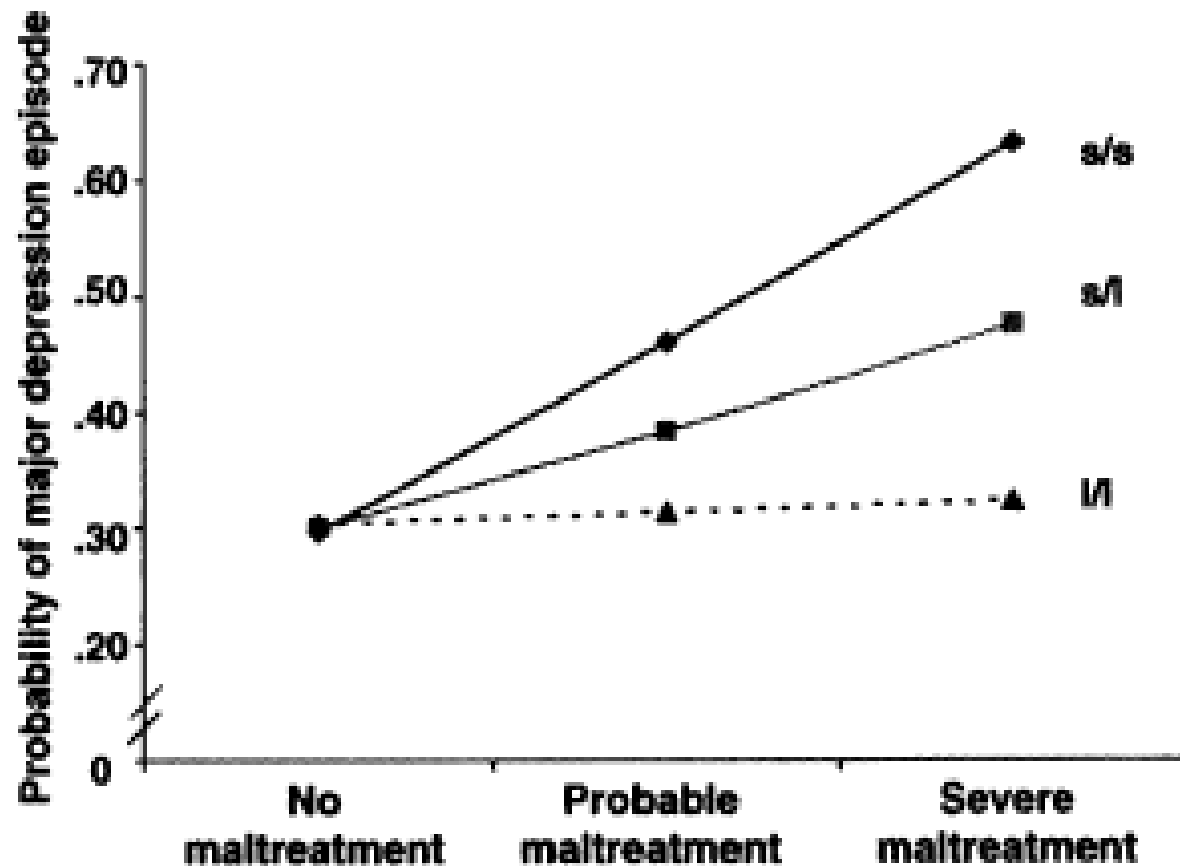


Figure 1. Fra Demontis et al. Nature Genetics 2018



Caspi et al. Science 2003



Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression

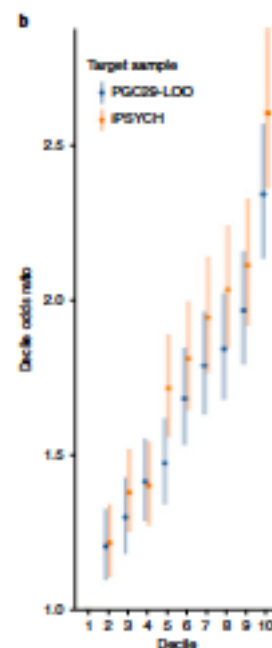
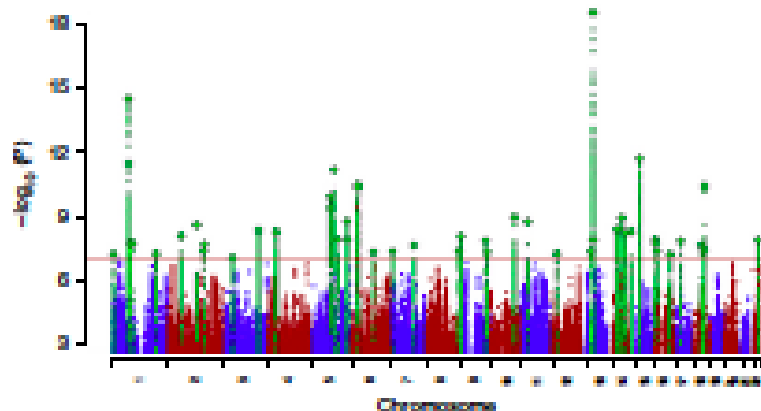
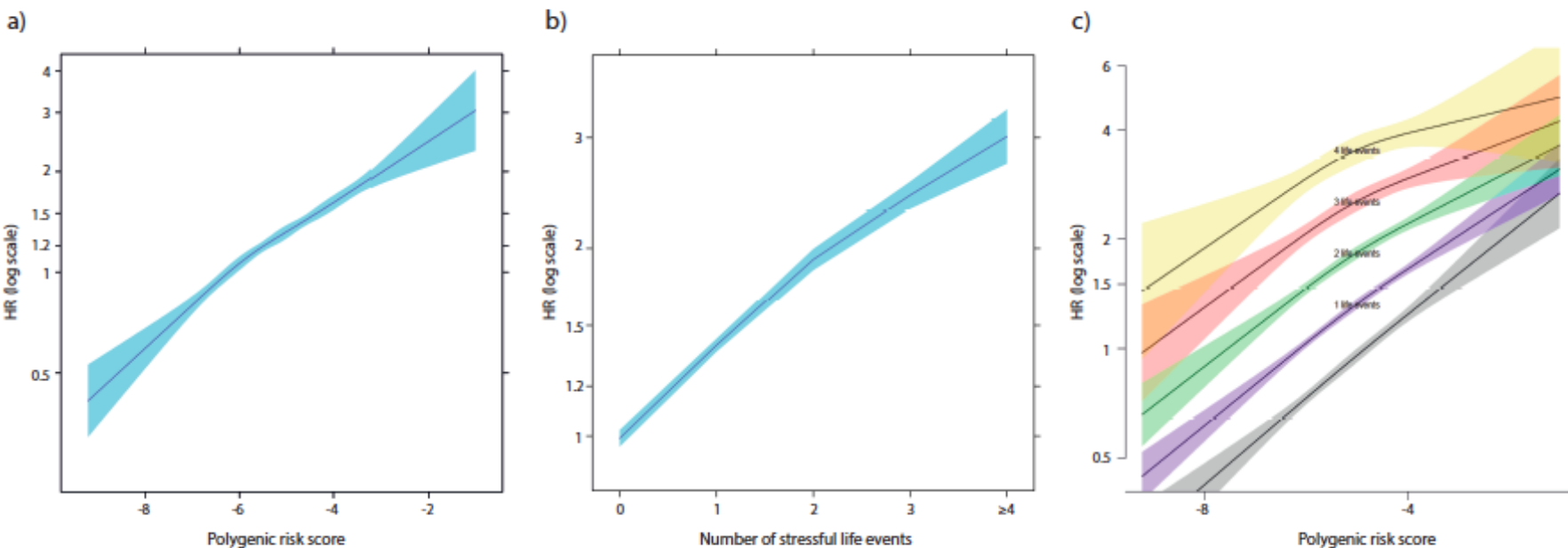


Figure 3. Main effects of polygenic risk score (PRS) and stressful life events (SLEs) as well as the interaction effect of PRS and SLEs on risk for depression among 18,533 depression cases and 20,569 random subcohort members.



All analysis were conducted as Cox regressions and results are presented as hazard ratios (HR). a) Association between polygenic risk score and risk for depression. b) Association between number of stressful life events and risk for depression. c) Association between polygenic risk score and risk for depression by number of stressful life events. Test for interaction was evaluated with Wald statistics, p-value = 0.10.

iPSYCH: verdens største studie af årsager til psykisk sygdom

- ▶ Danmark byder på unikke muligheder for studier af samspillet mellem arv og miljø som årsag til sygdom (og mange andre ting) i store, repræsentative populationer
- ▶ En væsentlig flaskehals er fleksibel adgang til data med høj sikkerhed og tilstrækkelig kapacitet

iPSYCH: verdens største studie af årsager til psykisk sygdom

- ▶ Etablering af sikker on-line adgang for forskere til Danmarks Statistiks data har indenfor de sidste 16 år revolutioneret vilkårene for dansk registerforskning
- ▶ Etablering af af muligheden for dette i en HPC setting vil være afgørende for at Danmark kan forløse sit unikke potentiale til at finde årsager til, og behandlingsmuligheder for en lang række sygdomme.

Udfordringer for registerforskning

1. Kobling af registre og store datamængder
2. Ingen gensidig tillid til analysemiljøer
3. Decentrale infrastrukturer
4. Data spredt hos mange myndigheder

Udfordringer for registerforskning

1. Kobling af registre og store datamængder
 - Nu muligt i regi af DST og Computerome
2. Ingen gensidig tillid til analysemiljøer
3. Decentral finansiering af infrastrukturer
4. Data spredt hos mange myndigheder

Udfordringer for registerforskning

1. Kobling af registre og store datamængder
2. Ingen gensidig tillid til analysemiljøer
 - Dataejer kan etablere eget sikkerhedsregime
 - DST stoler på løsningen
 - Samarbejde med Sundhedsdatastyrelsen på vej
3. Decentrale infrastrukturer
4. Data spredt hos mange myndigheder

Udfordringer for registerforskning

1. Kobling af registre og store datamængder
2. Ingen gensidig tillid til analysemiljøer
3. Decentrale infrastrukturer
 - Hostede servere kan erstattes af én fælles
 - Ressourcer og omkostninger kan deles
 - Kompetencer kan samles
4. Data spredt hos mange myndigheder

Udfordringer for registerforskning

1. Kobling af registre og store datamængder
2. Ingen gensidig tillid til analysemiljøer
3. Decentrale infrastrukturer
4. Data spredt hos mange myndigheder
 - AI analyse af sensitive data samme sted
 - Data skal ikke flyttes rundt mellem faciliteter
 - Samme høje sikkerhedsniveau

Erfaringer

- ▶ Eksisterende infrastrukturer kan samles
- ▶ Fokus på dataejerers behov
- ▶ Dataejere skal være trygge infrastrukturen
- ▶ Forskere kan tilfredsstilles
- ▶ Fælles koordinering og finansiering
- ▶ Tekniske retningslinjer for hvad et sikkert analysemiljø er



KOR

Det Koordinerende Organ
for Registerforskning

Tak

Preben Bo Mortensen
Professor NCRR,
PI iPSYCH

pbm@econ.au.dk

8716 5359

ipsych.au.dk

Jeppe Klok Due
Specialkonsulent,
sekretariatsleder KOR

jkd@sa.dk

4171 7211

registerforskning.dk