

Infrastruktur for analyse af registre og store datamængder

Preben Bo Mortensen, professor, iPsych
Jeppe Klok Due, specialkonsulent, KOR

KORs formål

Er et rådgivende organ under Uddannelses- og Forskningsministeriet, formålet er at stimulere og styrke dansk registerforskning.

KORs medlemmer

- ▶ Professor, overlæge Henrik Toft Sørensen, Aarhus Universitet (formand)
- ▶ Forskningschef Hans Hummelgaard, VIVE
- ▶ Professor Preben Bo Mortensen, Aarhus Universitet
- ▶ Professor Lisbeth B. Knudsen, Aalborg Universitet
- ▶ Professor Marianne Simonsen, Aarhus Universitet
- ▶ Klinisk professor, overlæge Merete Osler, Københavns Universitet
- ▶ Institutleder, professor Kirsten Ohm Kyvik, Syddansk Universitet
- ▶ Klinisk professor, overlæge Niels Obel, Københavns Universitet
- ▶ Professor Lotte Bøgh Andersen, Aarhus Universitet
- ▶ Professor Morten Grønbæk, Syddansk Universitet

KORs opgaver

- ▶ Rådgive Forskningsministeriet
- ▶ Skabe sammenhæng og koordination i miljøer
- ▶ Finansiere infrastruktur for registerforskning
- ▶ Projekter
 - Register over statslige registre
 - Pilot for multigenerationsregister
 - Højkvalitetsdokumentation af centrale variabler
 - Referencegruppe for registerforskning
 - Kortlægning af datawarehouses

Udfordringer for registerforskning

- ▶ Unikke danske registre udfordres af int. datasamlinger, i kraft af deres størrelse
- ▶ Vanskeligt at kombinere data
- ▶ Kobling af registre og store datamængder
- ▶ Forskerservice er dyr og mangler kompetencer
- ▶ Datasikkerhed er uoverskuelig med mange aktører

KORs løsningsforslag

- ▶ One-Stop-Shop or tilladelser
- ▶ Kombinere data fra flere data ansvarlige
- ▶ Få certificerede beregningsfaciliteter
- ▶ Gensidig tillid til sikkerhed, men mulighed for at definere eget sikkerhedsregime
- ▶ Kombinere flere datatyper
- ▶ Fælles finansiering af beregningsfaciliteter

Proof of Concept

Formål: At give forskere mulighed for at kombinere DST-registre med store datamængder *samtidig* med at dataansvarlig har fuld kontrol med data og brugere

Deltagere: KOR, Danmarks Statistik, DeiC og Computerome

Finansiering: DeiC og KOR

Use case: iPsych

Proof of Concept

- ▶ Dedikeret forbindelse DST-Computerome
- ▶ DST virtuel maskine på Computerome
- ▶ DST registre i krypteret kopi
- ▶ DST sikkerhedsregime implementeres
- ▶ DST kontrollerer data
- ▶ Kobling af socioøkonomiske registre med genomiske data

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

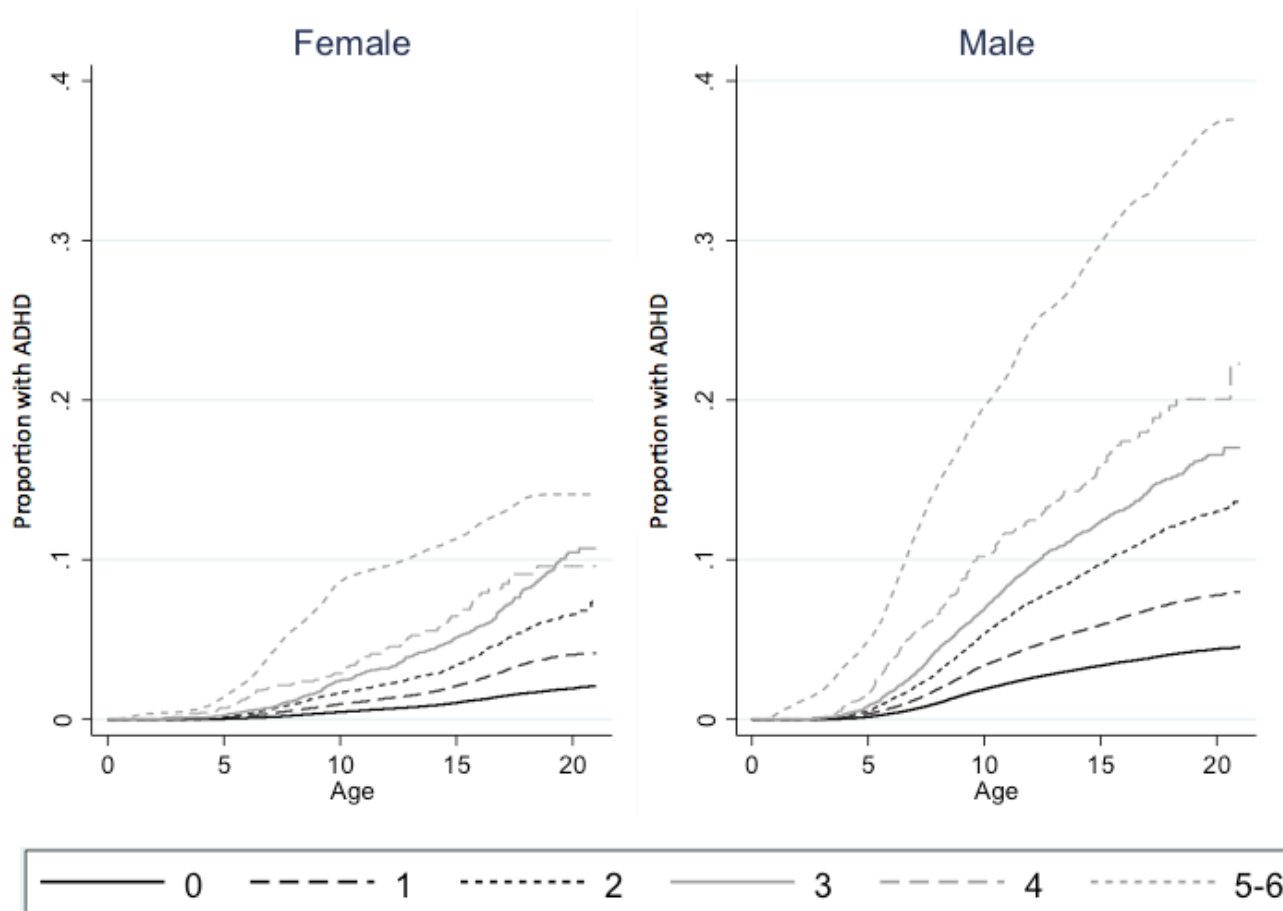
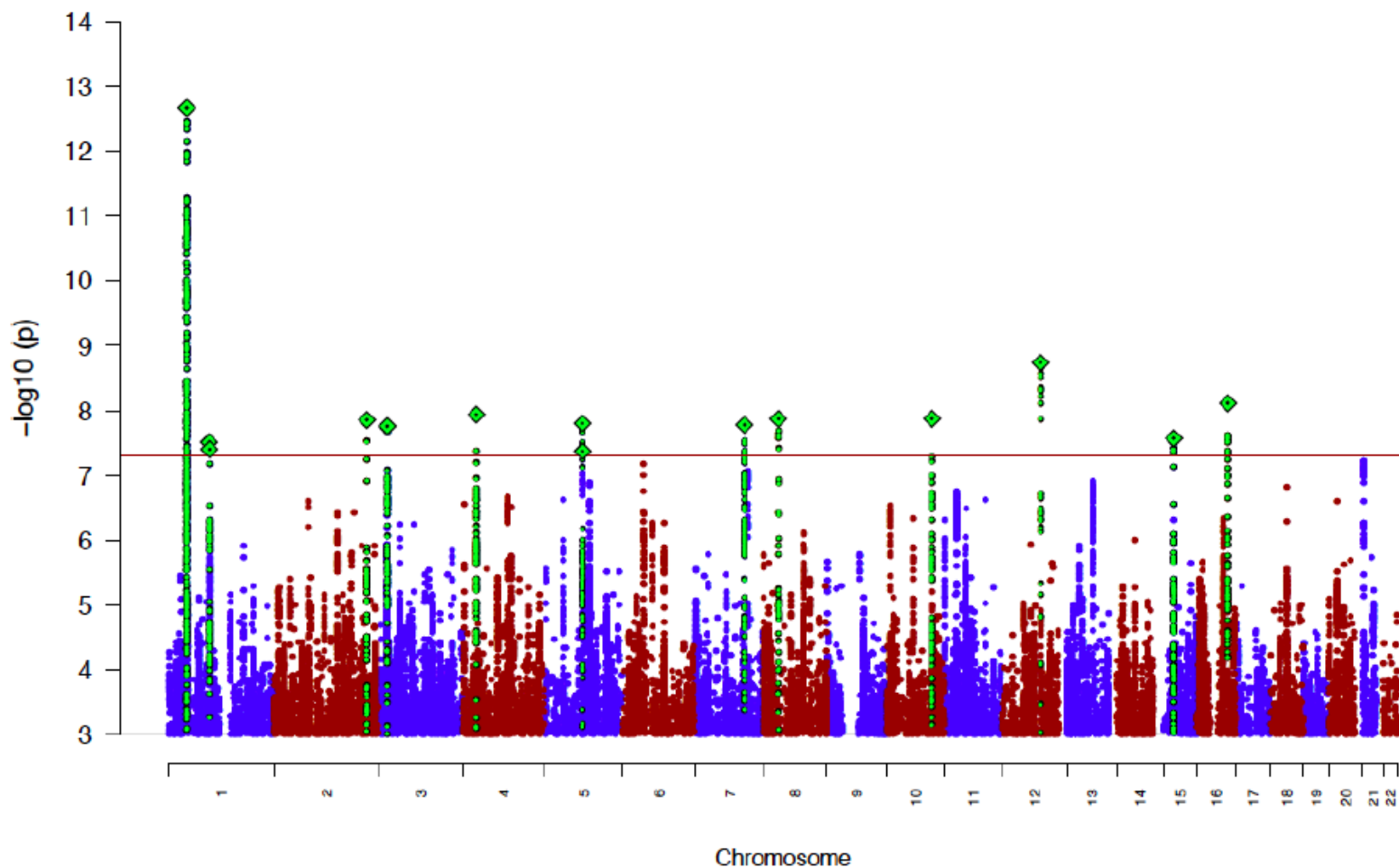


Figure 1. Fra Demontis et al. BioRxiv <https://doi.org/10.1101/145581>



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 15 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.

Perspektiver for fremtiden

En teknologisk løsning, der

- ▶ Giver forskerne de bedste muligheder
- ▶ Giver dataansvarlige kontrol over data
- ▶ Samler investeringer og kompetencer
- ▶ Højeste sikkerhed

Spørgsmål

Preben Bo Mortensen
Professor NCRR,
PI iPsych

pbm@econ.au.dk

8716 5359

ipsych.au.dk

Jeppé Klok Due
Specialkonsulent,
sekretariatsleder KOR

jkd@sa.dk

4171 7211

registerforskning.dk