

Registerforskningens brug af og behov for digitale arkiver

Carsten B. Pedersen
Professor, dr.med.

**Center for Interdisciplinær Registerforskning
Center for Registerforskning
Aarhus University, Denmark
E-mail: cbp@ncrr.dk**

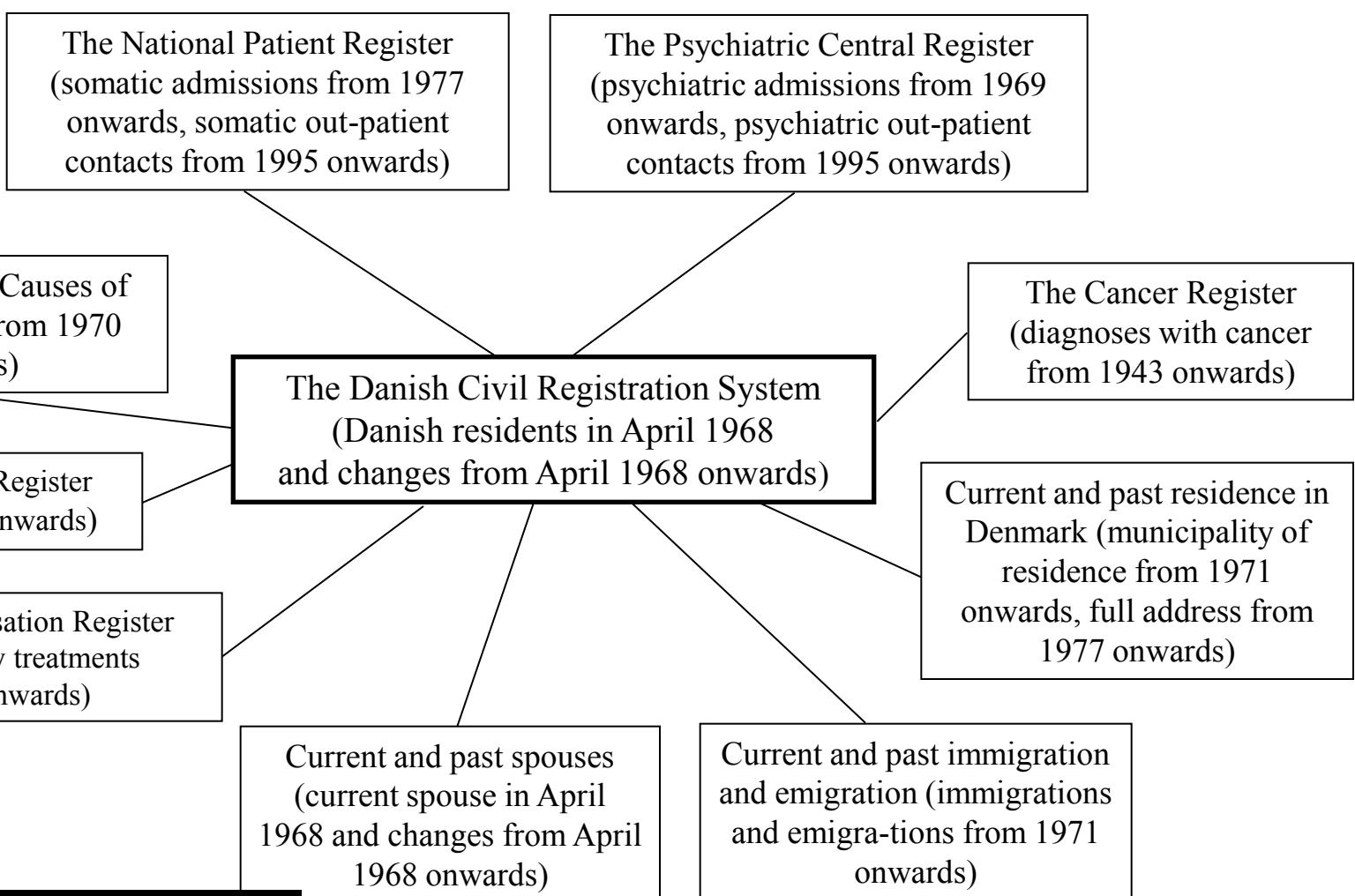
Finansiering / interessekonflikter

- Center for Interdisciplinær Registerforskning
 - Finansieret af Aarhus Universitet 2012-2017
- Center for Registerforskning
 - Finansieret af Aarhus Universitet 2005-
 - Danmarks Grundforskningsfond 2000-2005
- Psykiatrisk epidemiologisk forskning
 - Stanley Medical Research Institute 1993-
 - European Research Council (2012-2017)
 - Lundbeckfonden (2010-2015)

Familiedatabasen – samarbejdsaftale med SSI

- Sikker adgang til forskningsdata
- Baseret på SSI's helbredsrelaterede data
- Oprensning af data, der er målrettet longitudinelle populations-baserede studier
- NCRR er databehandler for SSI

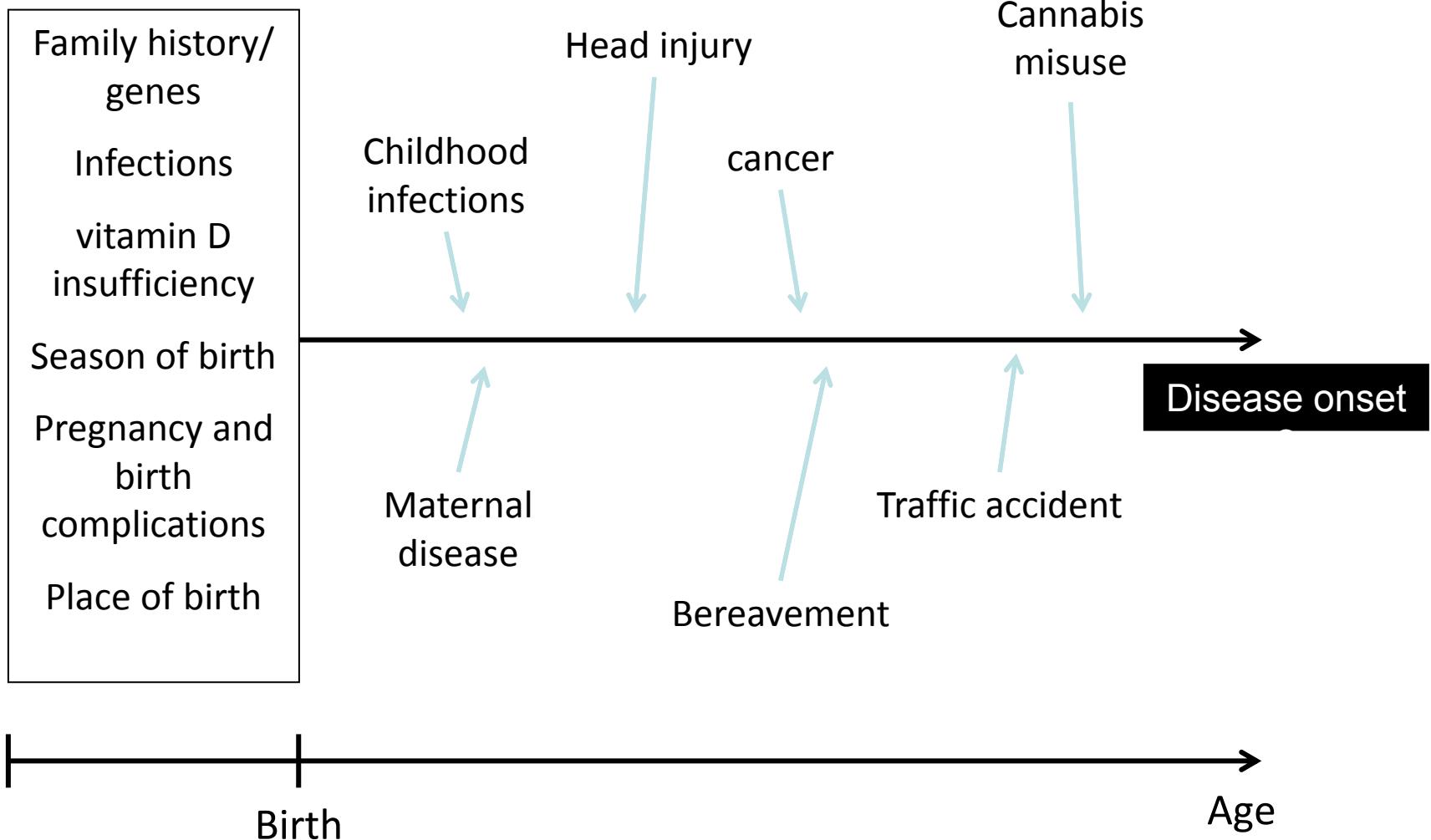
Familiedatabasen



Forskningsmæssig baggrund

- Register-baseret forskning
 - Forskning baseret på historiske og aktuelle oplysninger fra registre – primært administrative
 - I nogle tilfælde suppleret med oplysninger fra patientjournaler, biobanker mv.
 - Min baggrund er primært inden for den psykiatriske epidemiologi, dvs søgen efter årsager til psykisk sygelighed

What is life course epidemiology?



Registerbaseret forskning: Fordele

- Repræsentativ undersøgelse
- Longitudinel (fra vugge til grav)
- Stor population
- Billig undersøgelse
- Hurtig dataindsamling
- High impact journals
- Tiltrække forskningsmidler

Registerbaseret forskning: Udfordringer

- Data er indsamlet til et andet formål
- Ændring i betydningen af variabler over tid
- Typisk mangelfuld dokumentation

Disposition

- Den ideelle situation for register-baseret forskning
- Eksempler på register-baseret epidemiologisk forskning
- Eksempler på ændringer i variabler over tid

Den ideelle situation for register-baseret forskning

- Longitudinelle data med mange komplette og konsistente oplysninger for alle personer over en lang årrække
 - Det gode eksempel
 - Fødselsdato fra CPR-registret
 - (perfekt bortset fra nogle immigranter)
 - Det dårlige eksempel
 - Gestationsalder fra Medicinsk Fødselsregister

Potentielle studier af sygdomsrisiko

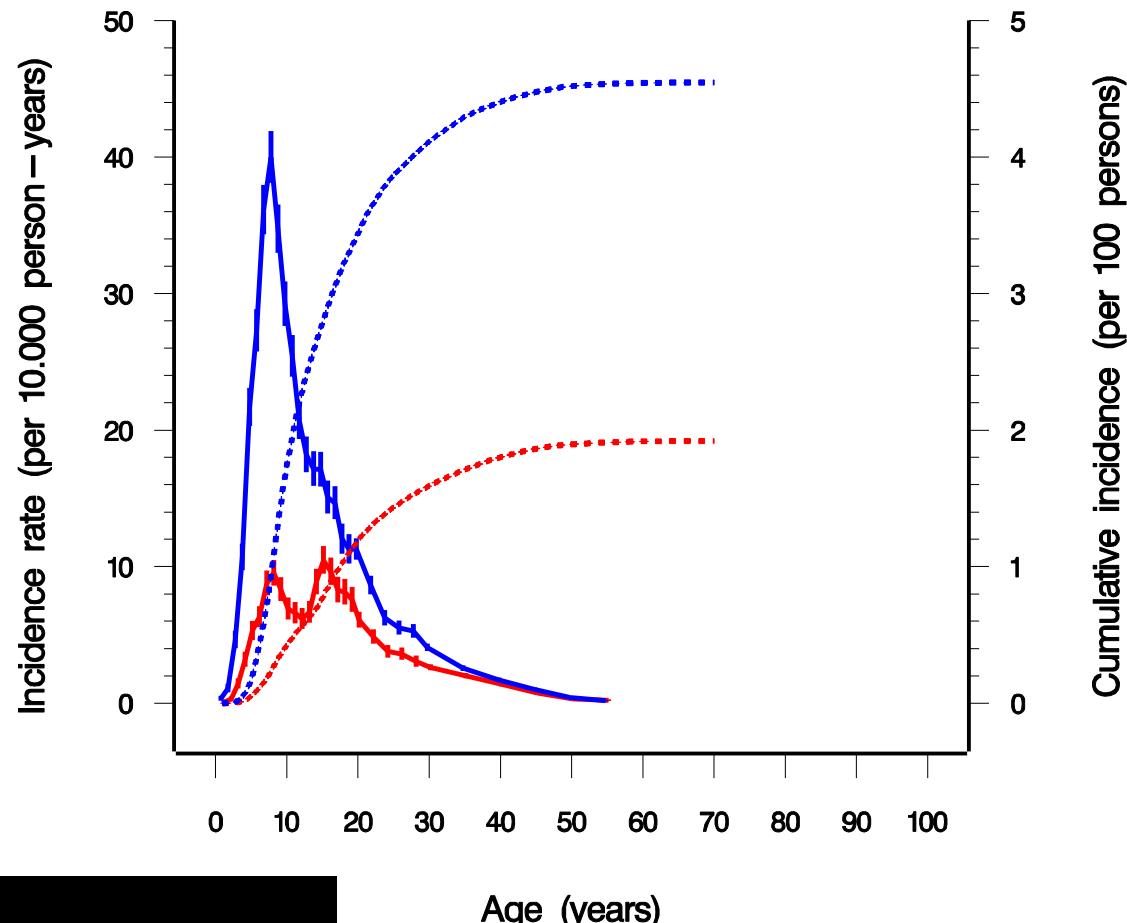
- Forekomst af sygelighed
- Familiær ophobning af sygdom
- Fødselskomplikationer
- Fødested og bopæl
- Immigration og emigration
- Adoptioner
- Miljømæssige eksponeringer
- Gener og biomarkører

Potentielle studier af sygdomsrisiko

- Forekomst af sygelighed
- Familiær ophobning af sygdom
- Fødselskomplikationer
- Fødested og bopæl
- Immigration og emigration
- Adoptioner
- Miljømæssige eksponeringer
- Gener og biomarkører

F90: Hyperkinetic disorder

Figure 3D: F90: Hyperkinetic disorder



Udfordringer

- Forskellige diagnostiske klassifikationer
- Oversættelse mellem klassifikationer er ikke altid oplagt
 - ICD7: 1943 – 1977 (kun Cancerregistret)
 - ICD8: 1968 – 1993
 - ICD10: 1994 – 2015?
 - ICD11: 2015? –
- Viden går tabt når der ændres klassifikation

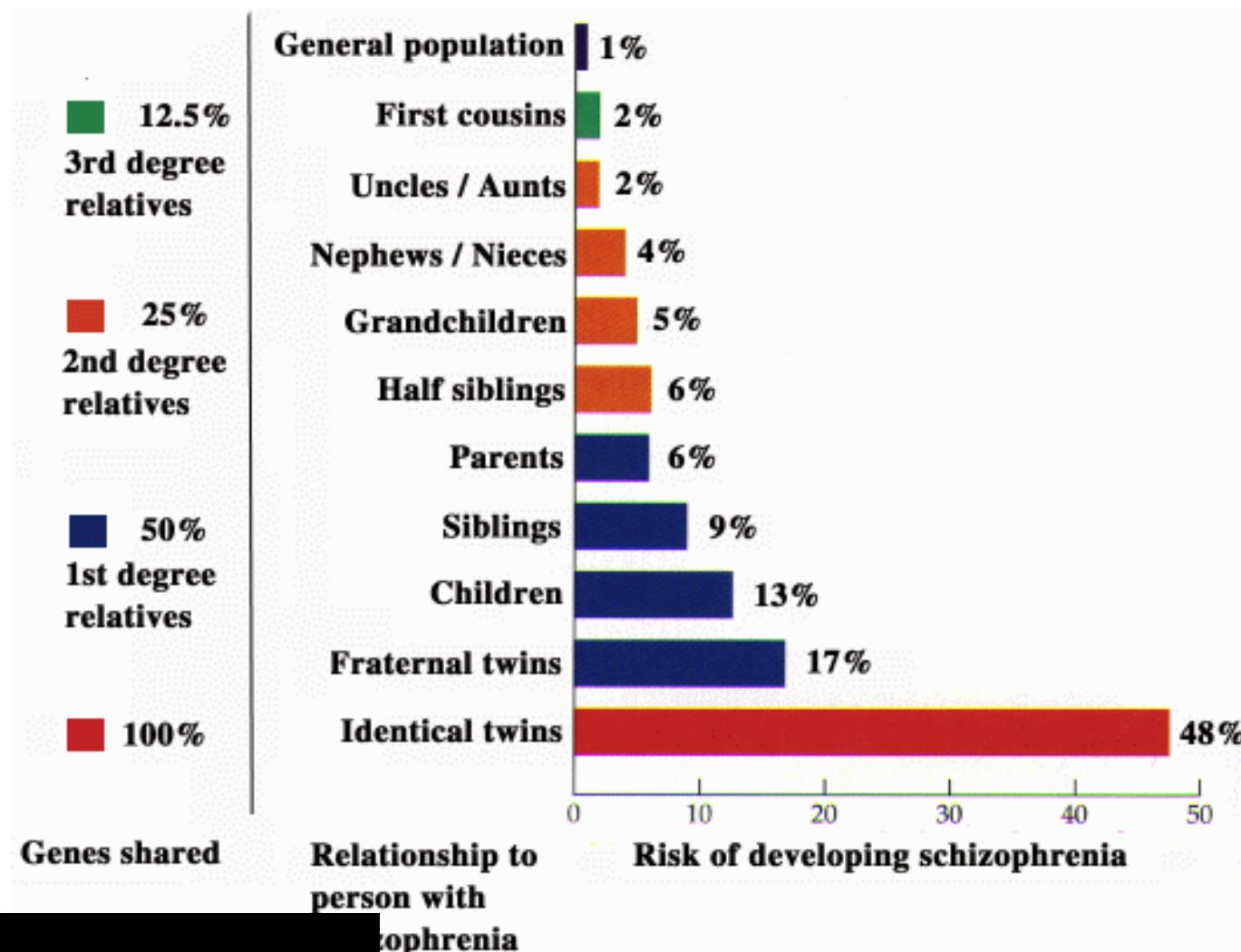
Potentielle studier af sygdomsrisiko

- Forekomst af sygelighed
- **Familiær ophobning af sygdom**
- Fødselskomplikationer
- Fødested og bopæl
- Immigration og emigration
- Adoptioner
- Miljømæssige eksponeringer
- Gener og biomarkører

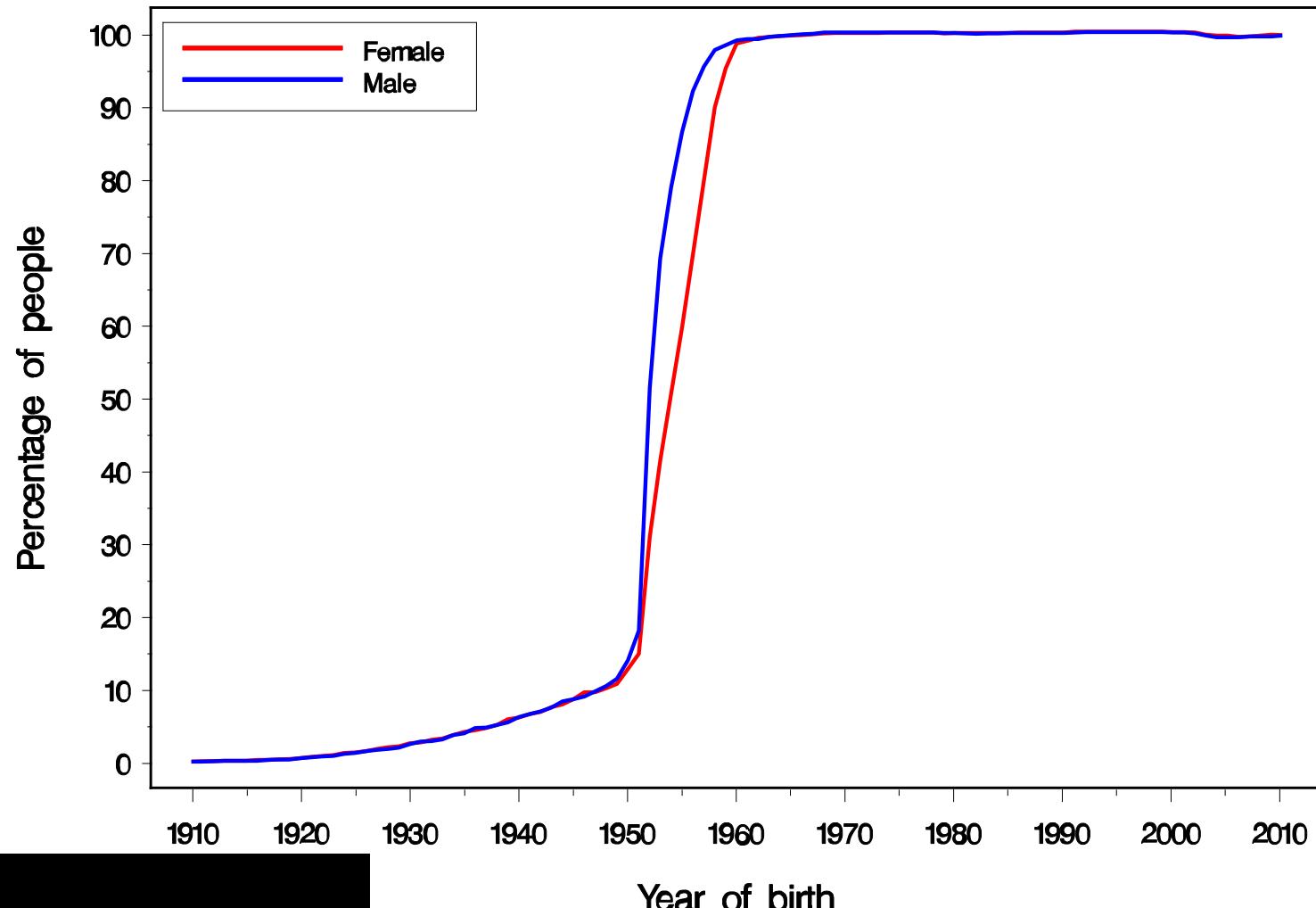
Skizofrenirisiko ift mors sygelighed

	RR	CI
Mor skizofren	7.10	6.28-8.01
Mor skizofreni lignende psykose	3.68	3.28-4.13
Mor andre psykiske lidelser	1.95	1.85-2.06
Mor ikke psykisk syg	1.00	

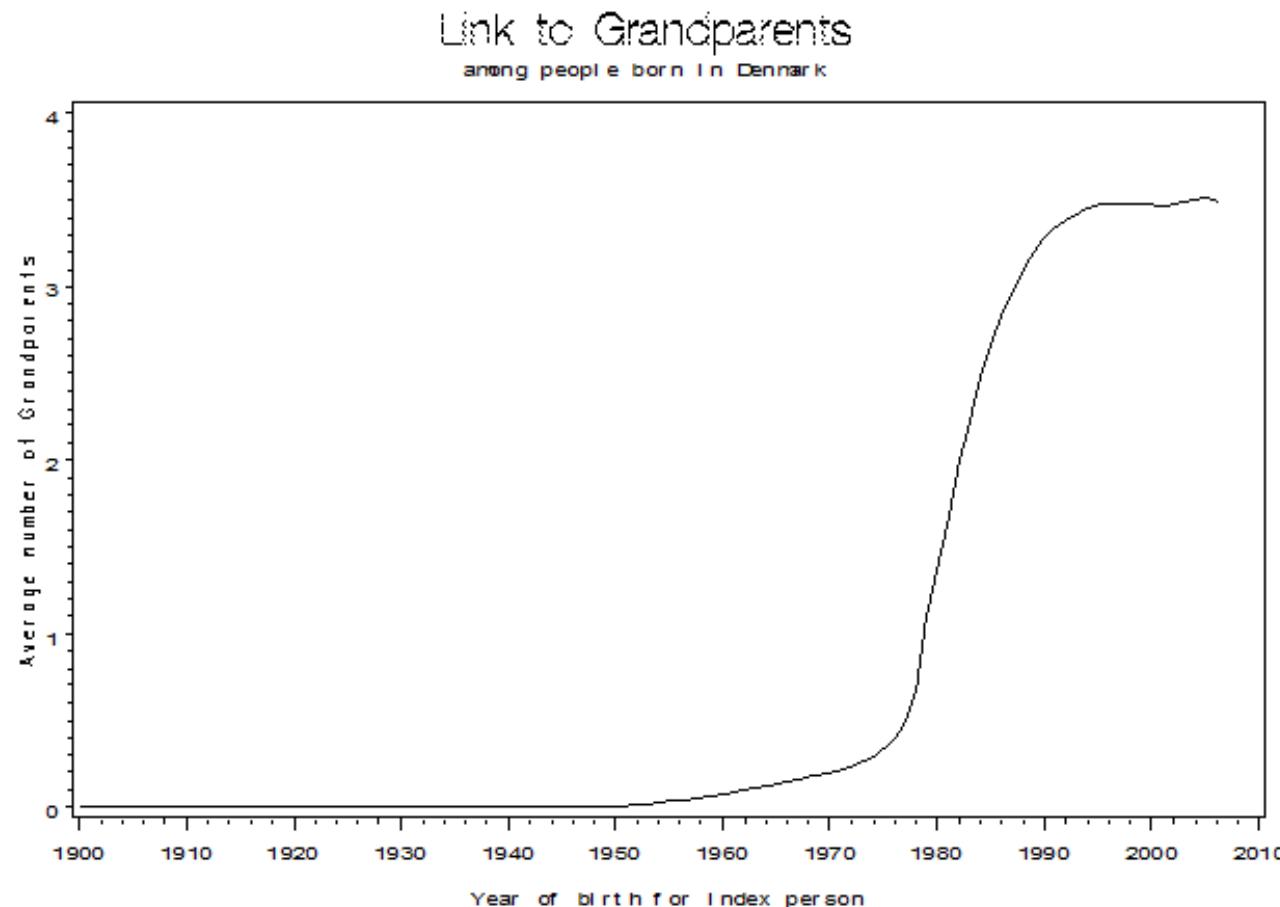
Familiær ophobning af skizofreni



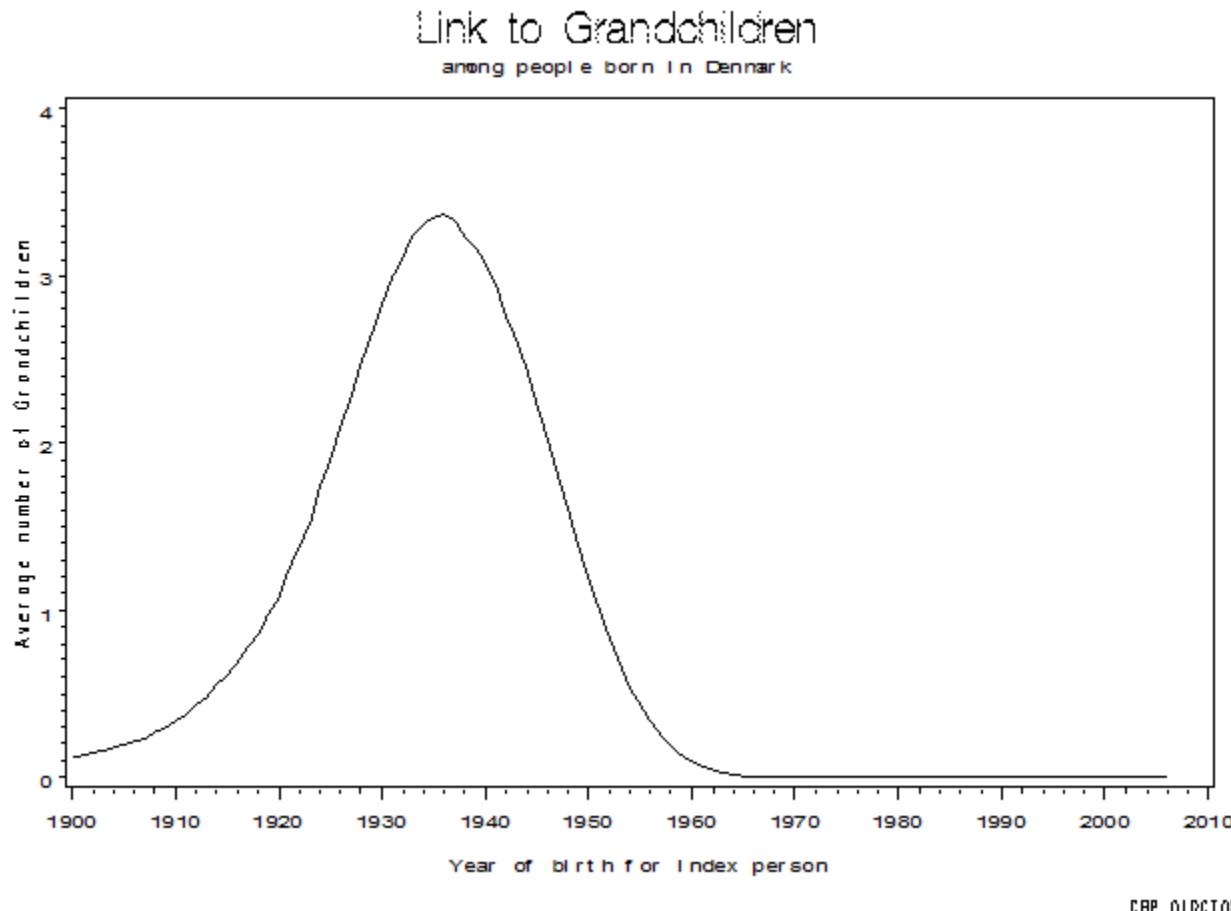
Oplysning om mors identitet i CPR-registret



Komplethed af bedsteforældre



Komplethed af børnebørn



Udfordringer

- Kun oplysning om 1-2 generationer afhængigt af fødeår
- Oplysninger baseret på de juridiske forældre registreret i CPR
 - Oplysninger om adoptioner mistes
 - Biologiske forældre versus juridiske forældre (validitet af far)

Komplette familier

- Etablering af mere komplette familier er muligt
- Indtastning af fødsler fra alle Kirkebøger
- Projekt iværksat af Statens Arkiver
- Potentiale
 - Bedre mulighed for familiebaserede studier
 - Identificere gener, der øger tilbøjeligheden til sygdomme

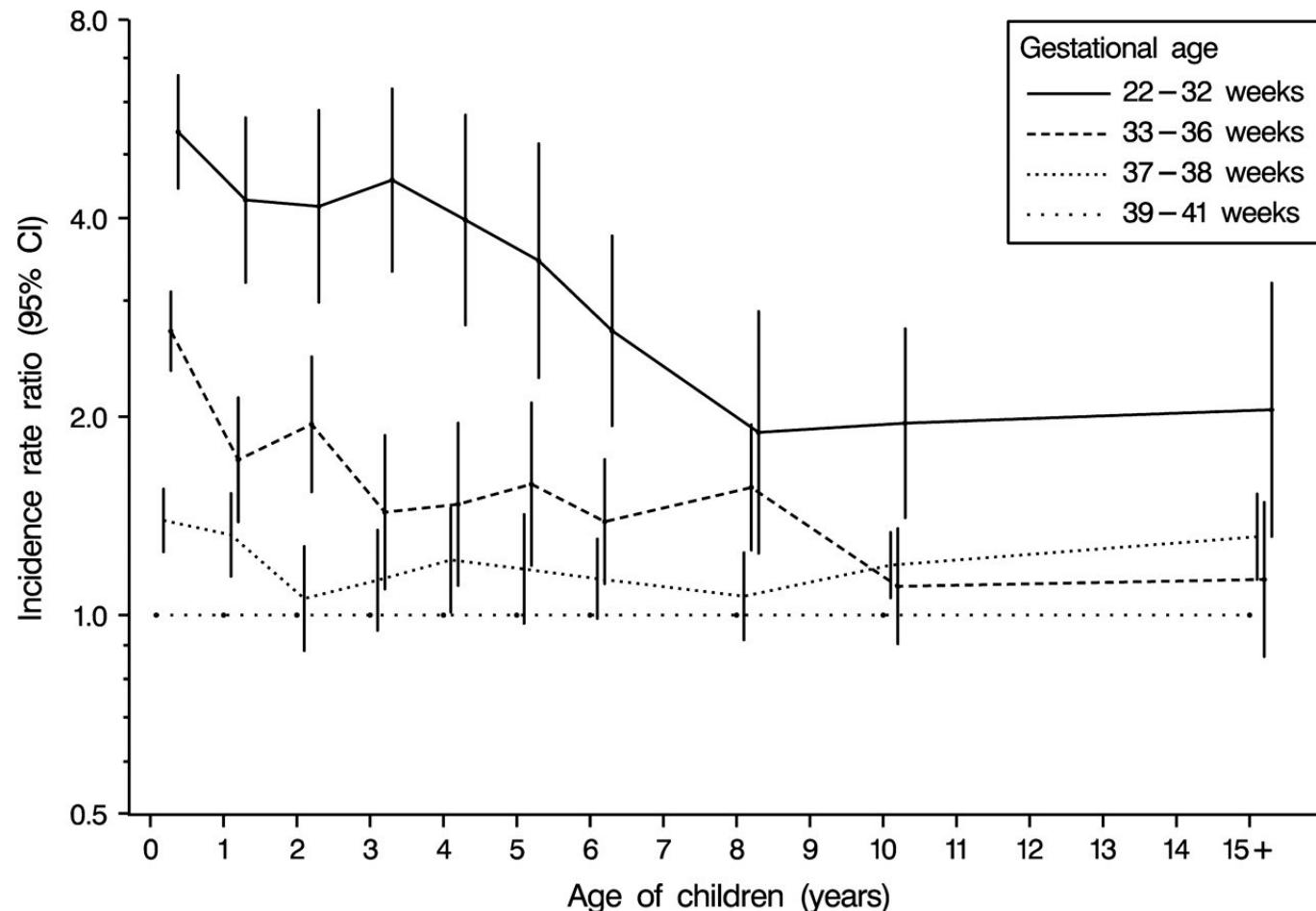
Komplette familier

Persons fødeår	Personer i CPR		Mors forventede fødeår	Mors mindste levealder for inklusion i CPR	Forventet andel af personer hvis mor kan genfindes i CPR
1920-24	Alle	Uden mor link	1894	≥ 74 år	52 %
1925-29	280234	276732	1899	≥ 69 år	66 %
1930-34	273710	267652	1904	≥ 64 år	77 %
1935-39	294652	284592	1909	≥ 59 år	84 %
1940-44	358335	340860	1914	≥ 54 år	89 %
1945-49	416211	388344	1919	≥ 49 år	93 %
1950-54	368758	237843	1924	≥ 44 år	96 %
1955-59	362606	51991	1929	≥ 39 år	97 %
I alt	2657994	2149509			80 %

Potentielle studier af sygdomsrisiko

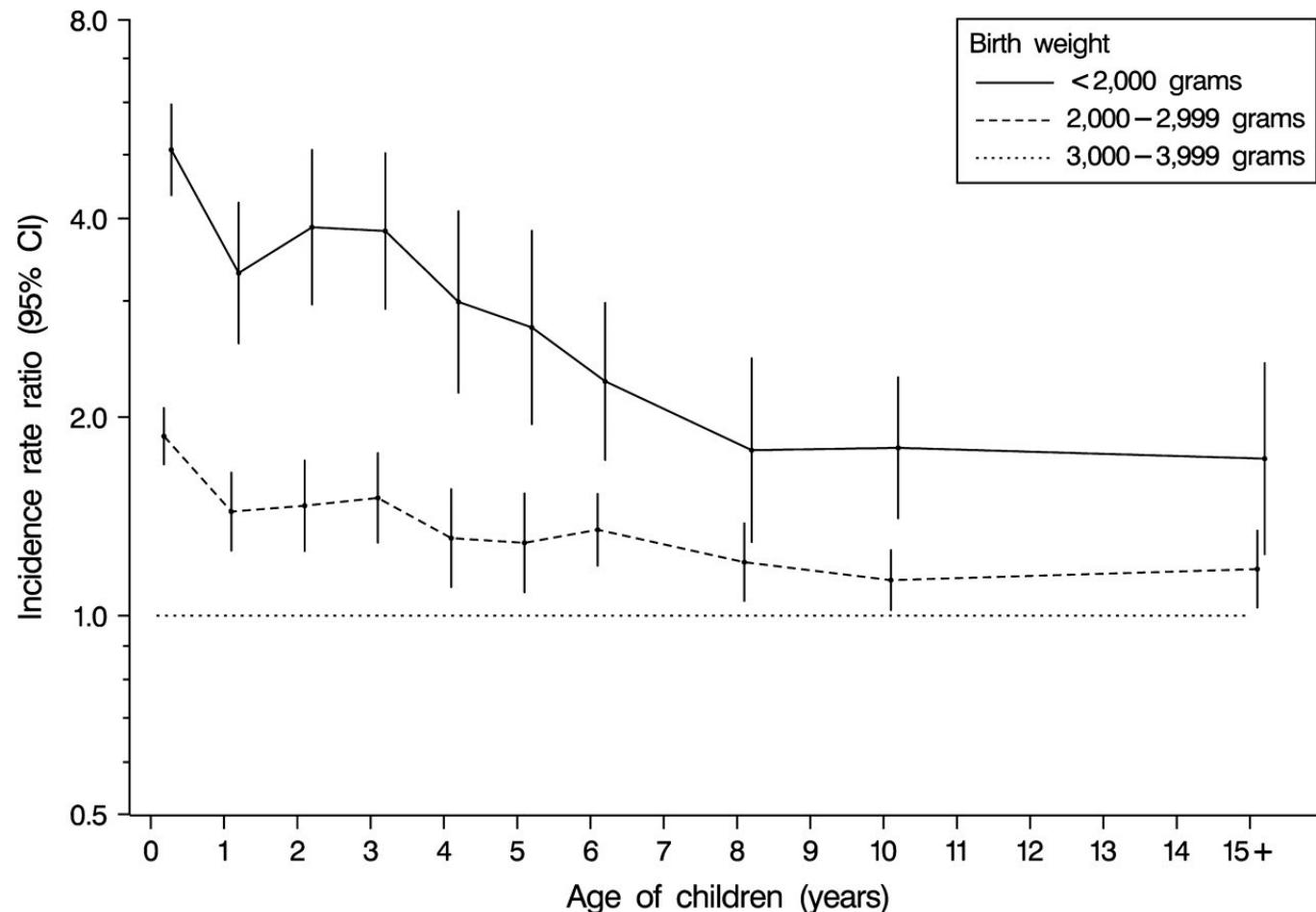
- Forekomst af sygelighed
- Familiær ophobning af sygdom
- **Fødselskomplikationer**
- Fødested og bopæl
- Immigration og emigration
- Adoptioner
- Miljømæssige eksponeringer
- Gener og biomarkører

Incidence rate ratio of epilepsy according to gestational age and age at first diagnosis of epilepsy, Denmark, 1979–2002



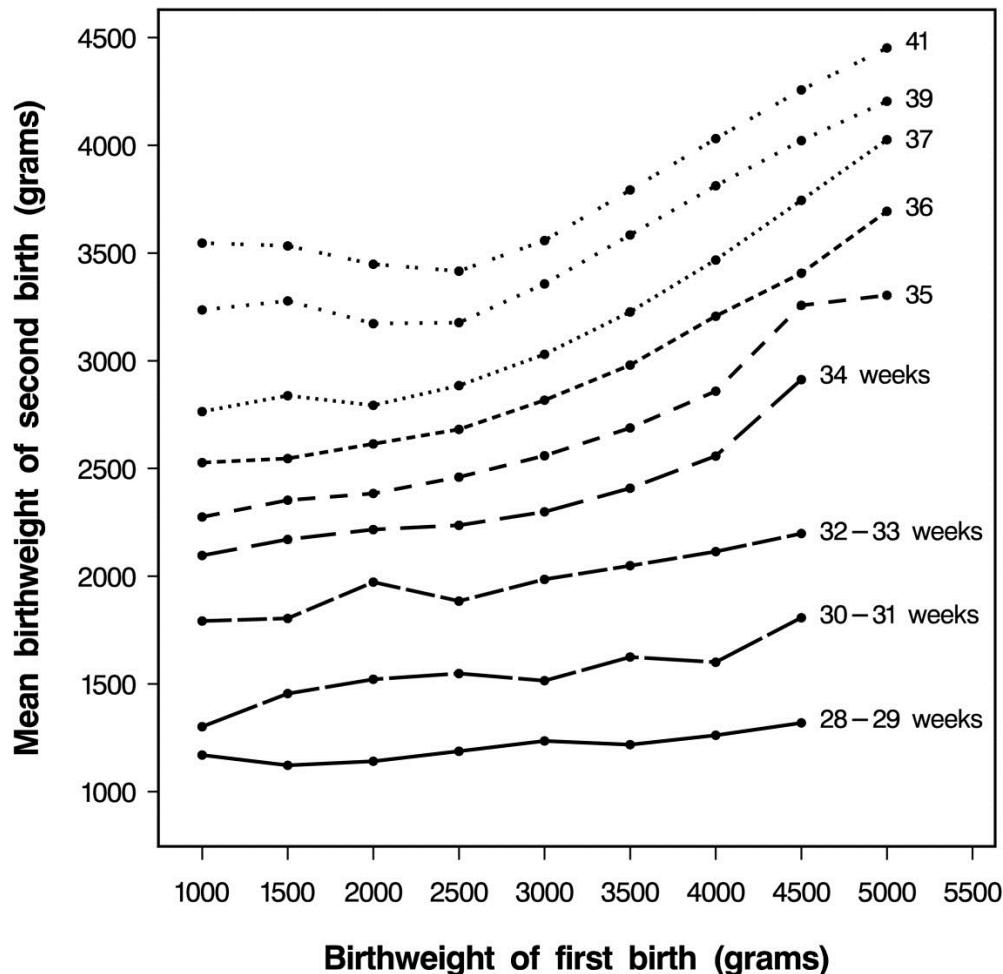
Sun Y et al. Am. J. Epidemiol. 2007;167:262-270

Incidence rate ratio of epilepsy according to birth weight and age at first diagnosis of epilepsy, Denmark, 1979–2002.



Sun Y et al. Am. J. Epidemiol. 2007;167:262-270

Forventet fødselsvægt



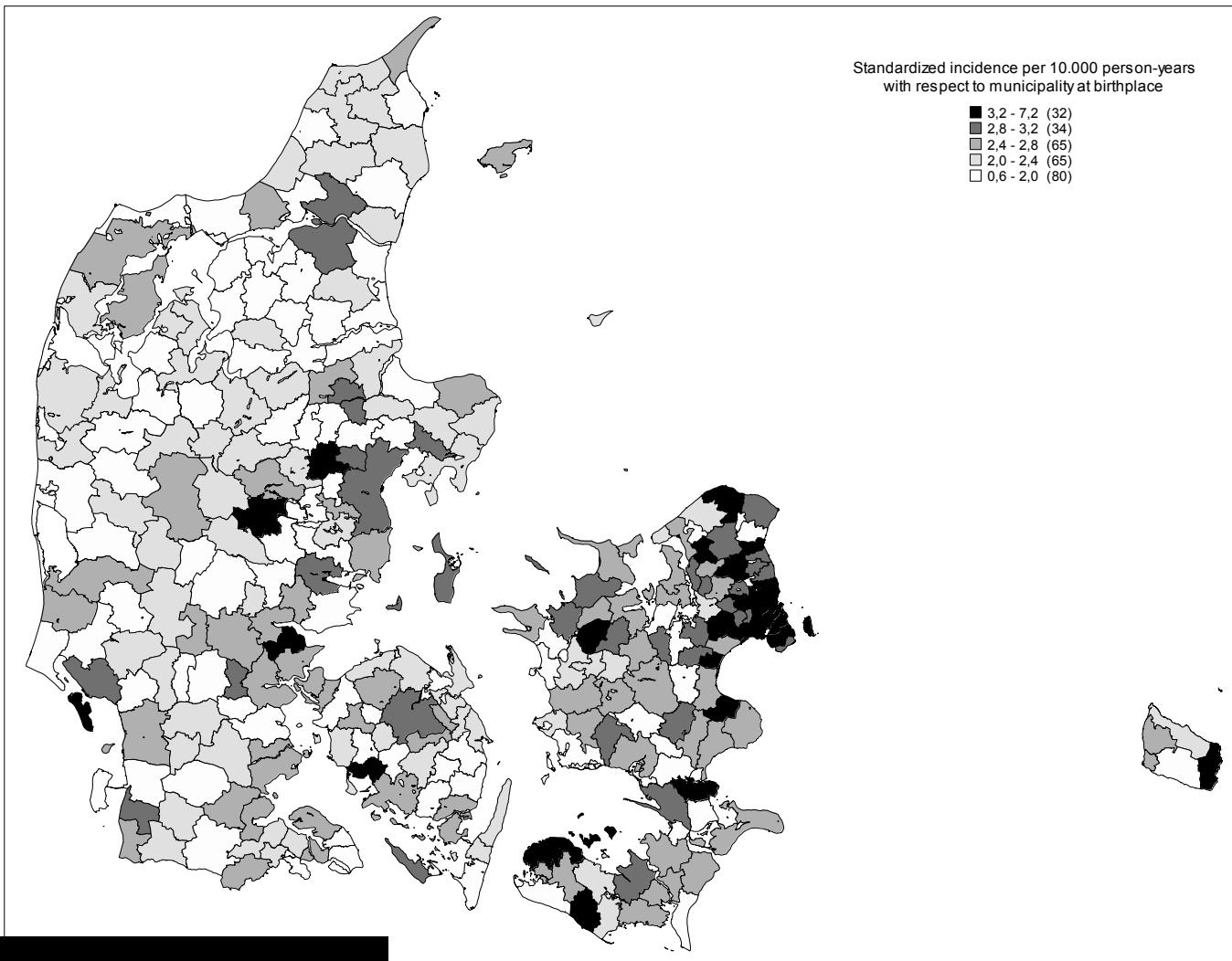
Udfordringer

- Information fra Medicinsk Fødselsregister
 - Stor variation i indholdet af de ca. 132 forskellige variabler
 - Gestationsalder fra Medicinsk Fødselsregister
 - 1973-1977: Antal uger født for tidligt
 - 1978-1996: Antal komplette uger, suppleret med ultralyd (1978-1961 ca 25% ukendt)
 - 1996-: Antal dage, suppleret med ultralyd
 - Rygning under graviditet, etc

Potentielle studier af sygdomsrisiko

- Forekomst af sygelighed
- Familiær ophobning af sygdom
- Fødselskomplikationer
- **Fødested og bopæl**
- Immigration og emigration
- Adoptioner
- Miljømæssige eksponeringer
- Gener og biomarkører

Forekomst af skizofreni



Relative risici for skizofreni

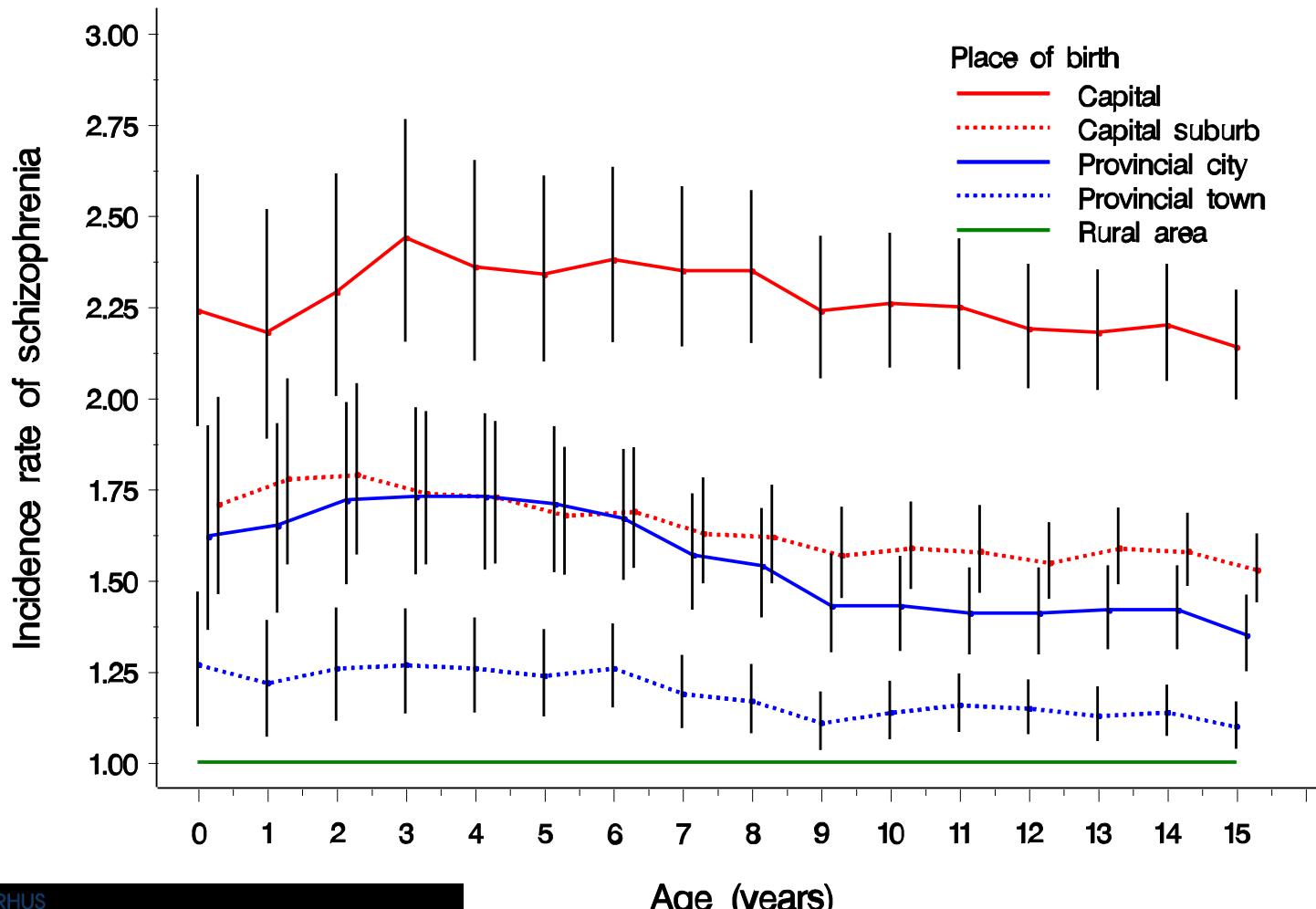
Fødested	RR1 (95% CI)	RR2 (95% CI)
Capital	2.49 (2.20-2.80)	2.40 (2.13-2.70)
Capital suburb	1.64 (1.40-1.93)	1.62 (1.37-1.90)
Provincial city	1.57 (1.36-1.81)	1.57 (1.36-1.81)
Provincial town	1.24 (1.10-1.41)	1.24 (1.10-1.41)
Rural area	1.00 (ref)	1.00 (ref)

RR1: Estimater var justeret for køn, alder, kalenderår, fødemåned og forældres alder ved barns fødsel

RR2: Estimater var yderligere justeret for familiehistorie med skizofreni

Effects of family history and place and season of birth on the risk of schizophrenia. N Engl J Med. 1999

Urbanization by age at residence



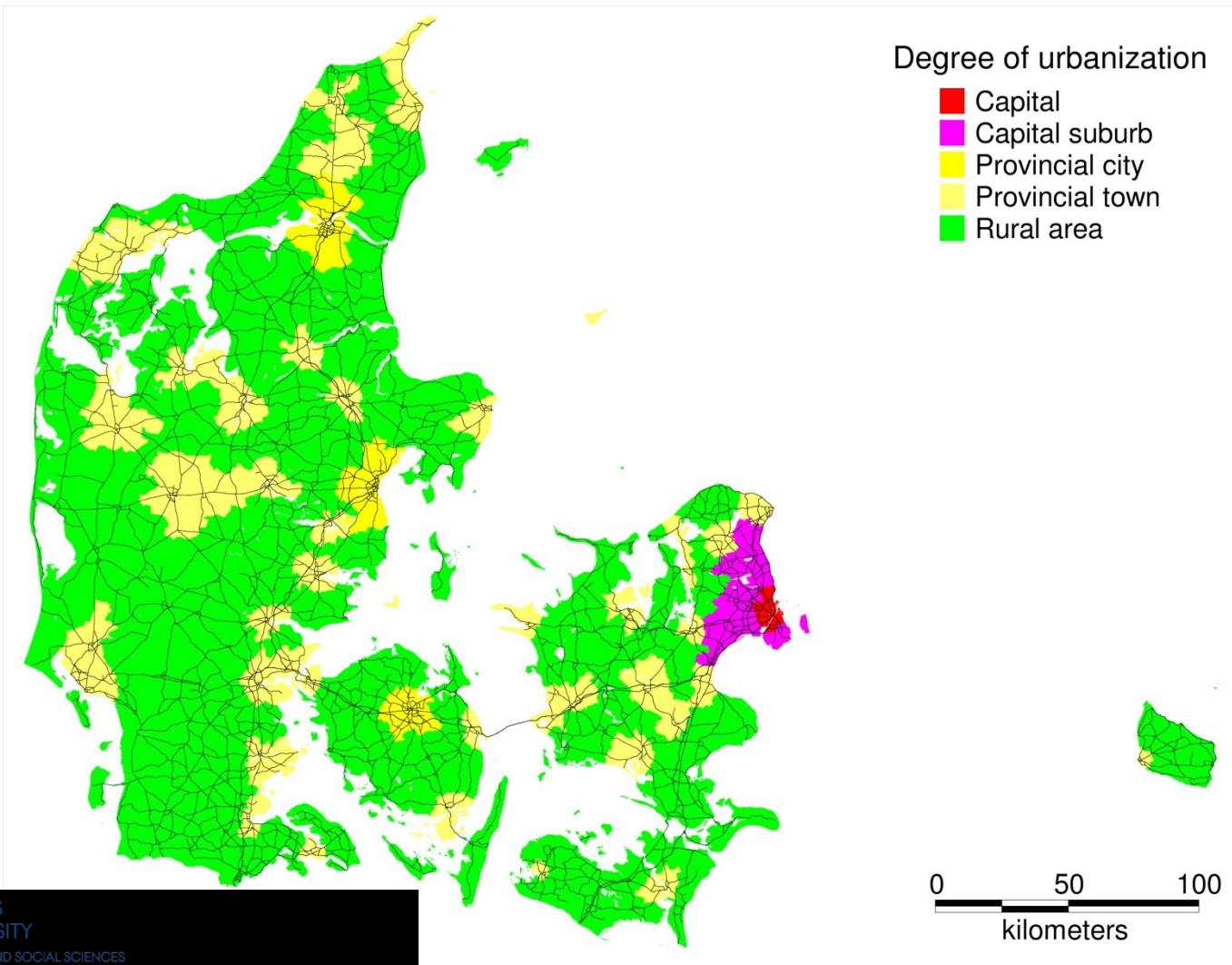
Udfordringer

- Gode oplysninger om fødested (sogn)
- Klassifikation baseret på antal indbyggere i største by i kommunen
(før kommunalreformen i 2007)
- Klassifikation af fødested for personer født efter kommunalreformen er besværlig

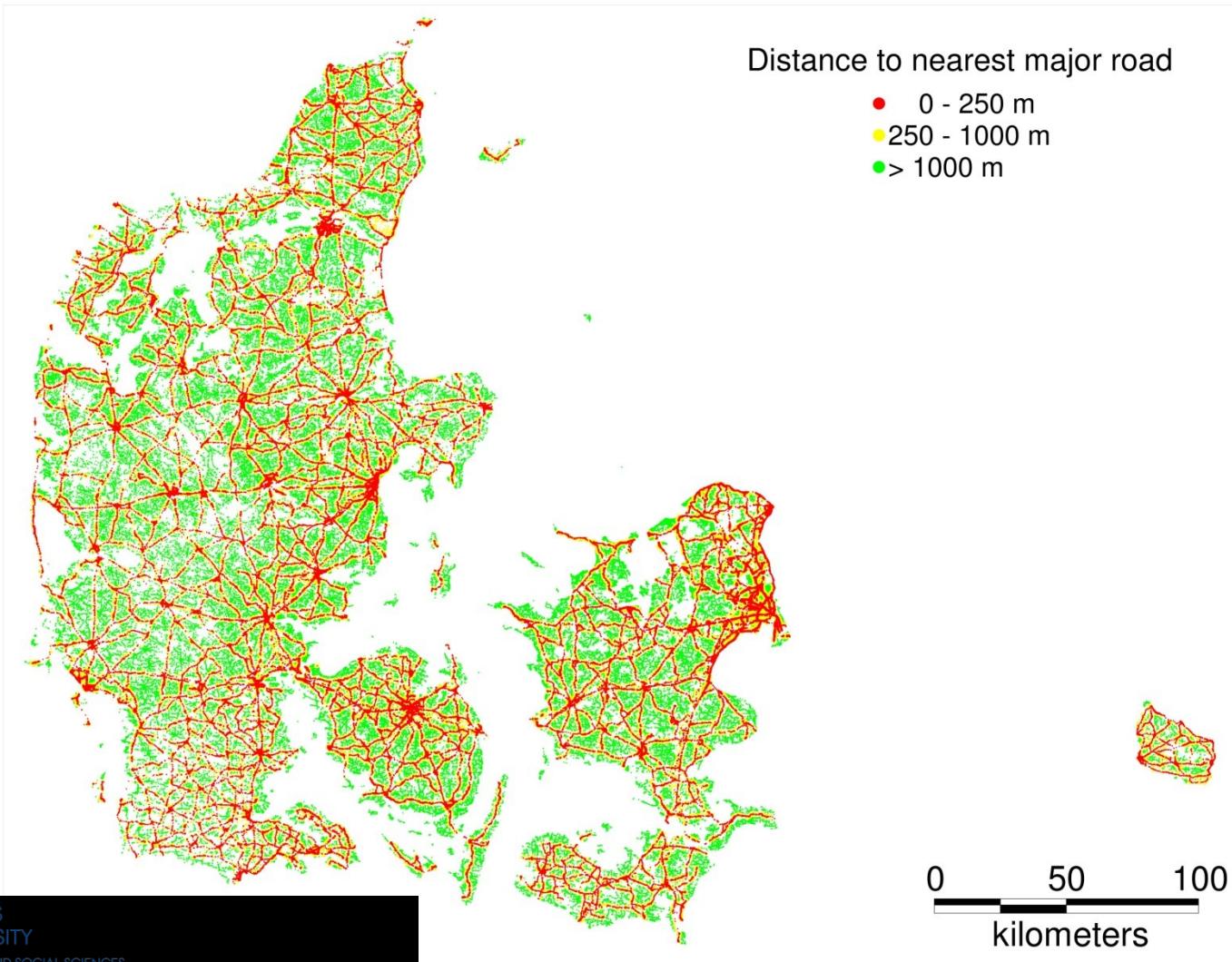
Potentielle studier af sygdomsrisiko

- Forekomst af sygelighed
- Familiær ophobning af sygdom
- Fødselskomplikationer
- Fødested og bopæl
- Immigration og emigration
- Adoptioner
- Miljømæssige eksponeringer
- Gener og biomarkører

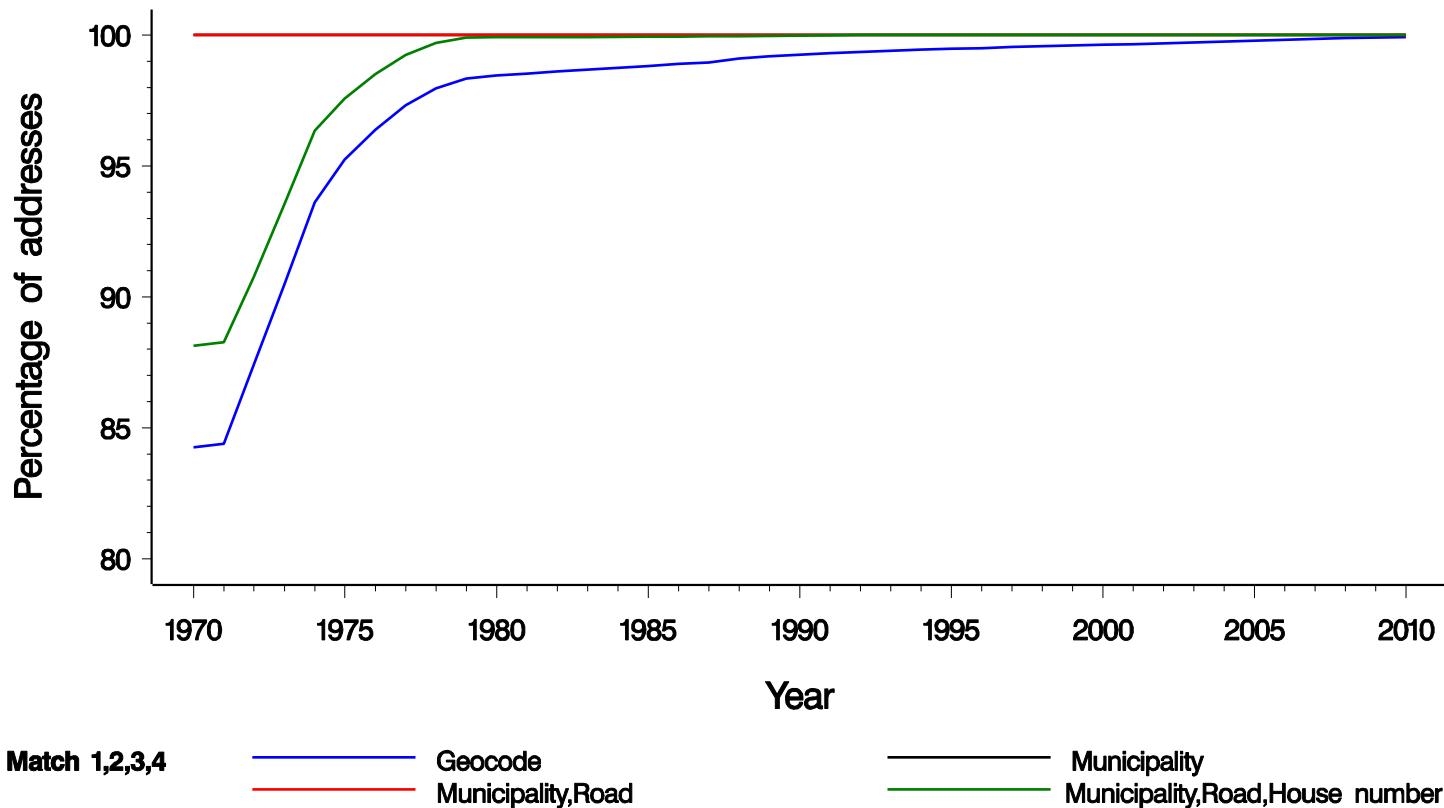
Urbanicity and major roads



Adresser i Danmark



Completeness of information on permanent address in Denmark



Udfordringer

- Gode oplysninger om befolkningens bopæl
- Kommunalreformer
 - 1970-1978, ændring af mange vejkoder??
 - 2003-2006: Bornholm, Ærø, Langeland
 - 2007: Landsdækkende
 - Mangler link mellem historiske og aktuelle
vejkoder for de ældste data
 - Nedlagte veje

Potentielle studier af sygdomsrisiko

- Forekomst af sygelighed
- Familiær ophobning af sygdom
- Fødselskomplikationer
- Fødested og bopæl
- Immigration og emigration
- Adoptioner
- Eksponering for miljømæssige eksponeringer
- **Gener og biomarkører**

iPsych (from ultimo 2014)

- Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH)
- Schizophrenia, Bipolar affective disorder, Affective disorder, Autism, ADHD, Anorexia
- 55,000 cases and 30,000 shared controls
 - GWAS (500,000 variables), NGS (approx 20 million variables), neonatal infections, food related antigens, vitamin D, cytokines

PKU biobanken

BLODPROV FRA NYFØDT

Moder CPR-nr:	<input type="text"/> <input type="text"/>
Efternavn	<input type="text"/>
Fornavn	<input type="text"/>
Gentaget prøve <input type="checkbox"/>	Årsag til gentagelse:
Prøvetagningsdato <input type="text"/> <input type="text"/>	
Prøven taget af [initialer] <input type="text"/> <input type="text"/>	
CPR-nr. [eller fødselsdag] Barnet: <input type="text"/> <input type="text"/> Fødelevvægt: <input type="text"/> Fulde graviditetsuger ved fødselsstidspunkt: <input type="text"/> Køn <input type="checkbox"/> Pige <input type="checkbox"/> Dreng	

OBS!

Etiket bedes anbragt vendret op mod det optiske venstre hjørne (af hensyn til optisk læsning)



AARHUS
UNIVERSITY

BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

Scanning af arvemassen hos skizofrene i DK Manhattan plot

OPEN

Molecular Psychiatry (2013), 1–9
© 2013 Macmillan Publishers Limited. All rights reserved 1369-6184/13
www.nature.com/mp

ORIGINAL ARTICLE

Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci

AD Berglum^{1,2}, D Demontis^{1,3}, J Grove^{1,3,4}, J Palleisen^{1,2}, MV Hollegaard⁵, CB Pedersen^{1,6}, A Hedemand^{1,7}, M Mattiasson^{7,8,9}, GROUP Investigators^{1,9}, A Utterlinden¹, M Nyergaard^{1,10}, T Ørntholm¹, C Wu^{1,11}, M Didriksen¹, M Nordenstoft^{1,13}, MM Nothen^{1,14,15}, M Reetzschel¹⁰, RA Ophoff¹⁰, S Schou^{11,12}, RH Yolken¹⁰, DM Hougaard¹, PB Mortensen^{1,6} and O Mors^{1,3}

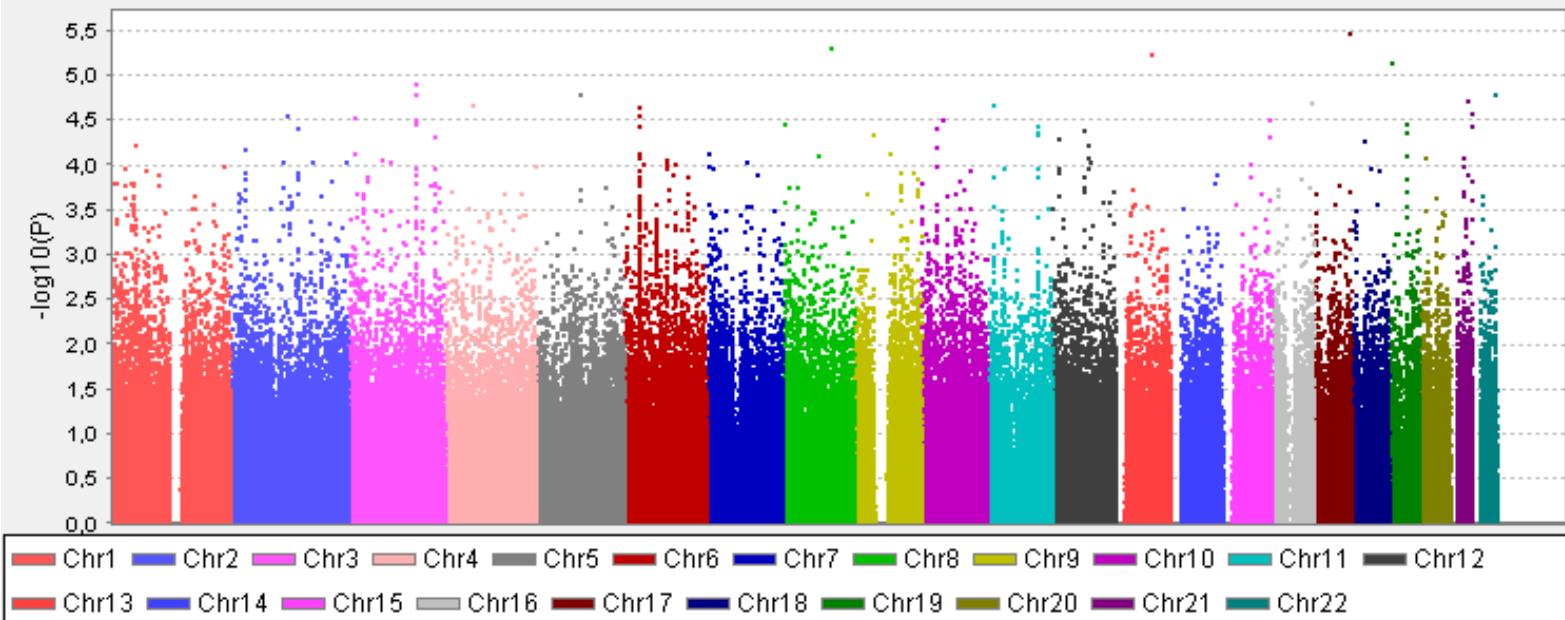
Genetic and environmental components as well as their interaction contribute to the risk of schizophrenia, making it highly relevant to include environmental factors in genetic studies of schizophrenia. This study comprises genome-wide association (GWA) and follow-up analyses of all individuals born in Denmark since 1981 and diagnosed with schizophrenia as well as controls from the same birth cohort. Furthermore, we present the first genome-wide interaction survey of single nucleotide polymorphisms (SNPs) associated with schizophrenia. The study includes 888 cases and 882 controls, and the follow-up analysis included 888 cases and 882 controls. The results show that the combined Danish GWA analysis identified 15 SNPs associated with schizophrenia, including 11 novel loci. One SNP, rs11038, located in the ZEB1 gene, was significantly associated with schizophrenia in the GWA analysis ($P = 7.0 \times 10^{-8}$). This SNP was also significantly associated with schizophrenia in the follow-up analysis ($P = 1.39 \times 10^{-8}$). Both genes have previously been suggested as being involved in schizophrenia. In addition, we found significant associations between the risk of schizophrenia and the presence of antibodies against cytomegalovirus (CMV). We performed a genome-wide interaction analysis summarizing independent signals e-wide significant locus overlapping the gene ZEB1 ($P = 7.0 \times 10^{-8}$). Significant interaction with maternal CMV infection was found to previously implicated in schizophrenia, stressing the importance of environmental factors in schizophrenia.

2013; doi:10.1038/mp.2013.2

IWS: region-wide analysis; ZEB1

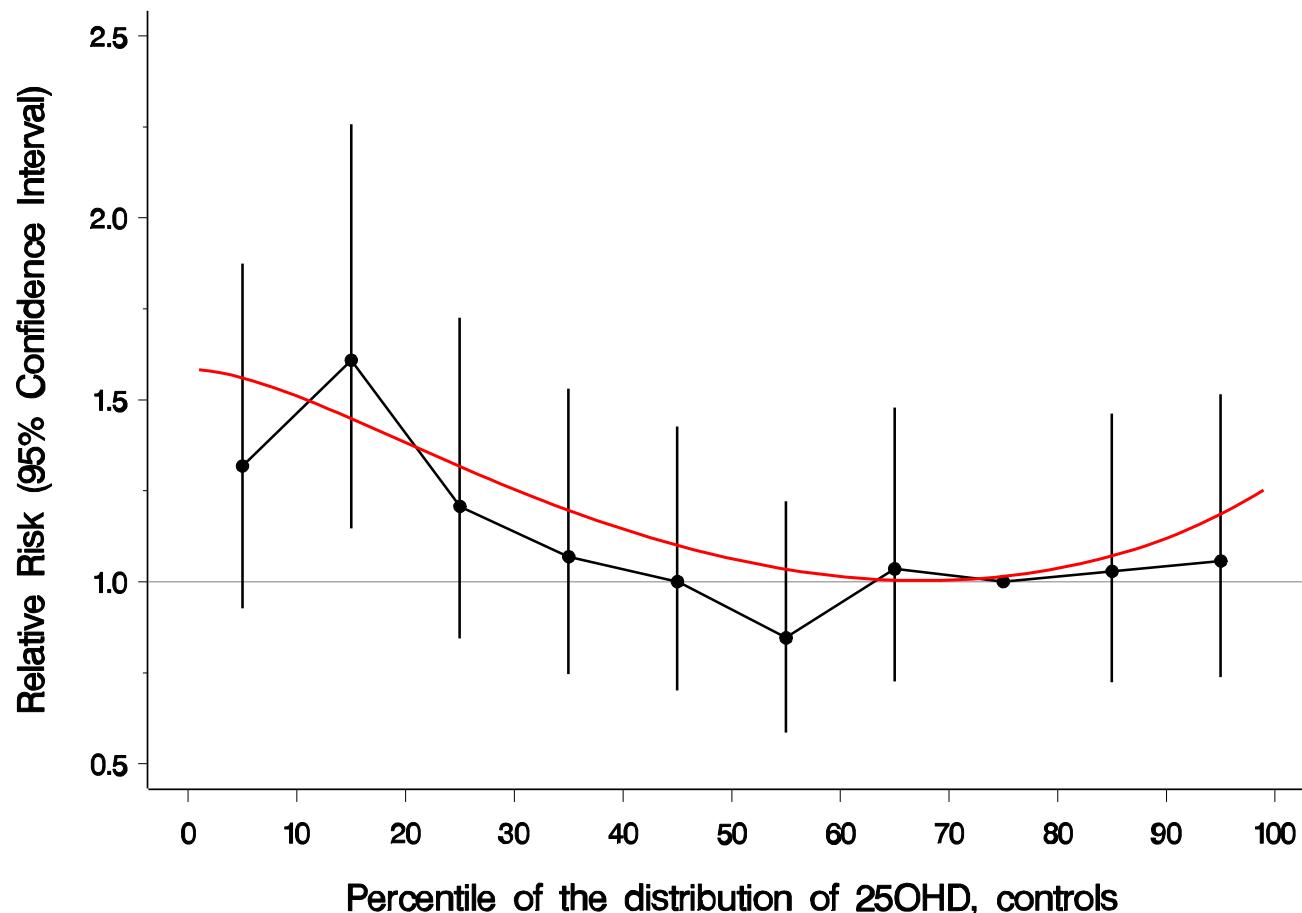
common susceptibility variants of small effect are involved, collectively capturing around 30% of the genetic risk.^{1,2} The remaining genetic risk is due to rare mutations, variants and gene-environment interactions.^{3,4,10} Due to the low-effect sizes of the common risk variants, the heritability can be between 1% and 20%.^{1,2} It is well documented that the environment has an important role in the development of schizophrenia.¹¹ Especially in the last years, the role of environmental risk factors may be increased, supported by several studies demonstrating an association of maternal infection with increased risk of the child developing schizophrenia.^{12,13} In addition, evidence for a strong interaction between genetic variation in the offspring and markers of maternal infection (maternal antibodies) may influence the risk of schizophrenia,¹⁴ stressing the importance of taking environmental factors into account in genetic studies.

¹ University, Aarhus, Denmark; ²Centre for Psychiatric Research, Aarhus University Hospital, research, ³Psych, Aarhus and Esbjerg, Denmark; ⁴Neuroimaging Research Centre, Aarhus



- ca. ~ 1800 personer (900 syge, 900 raske)
- ca. 550.000 genvarianteundersøgelse hos hver person.

Neonatal Vitamin D og skizofreni



Udfordringer

- PKU biobanken
 - Fremfinding af PKU prøver
 - Identifiers er indskannet - hyppige fejl (2-3%)
 - Ikke alle oplysninger fra kartotekskortet er indtastet/skannet
 - Prøvesamplingsdato
 - Kombination af biobank, register og kartotekskort

Afslutning

- Dansk register-baseret forskning i verdensklasse
 - Repræsentativ undersøgelse
 - Longitudinel (fra vugge til grav)
 - Stor population
 - Billig undersøgelse
 - Hurtig dataindsamling
 - High impact journals
 - Tiltrække forskningsmidler

Afslutning

- Udfordringer inkluderer
 - Konsistens af oplysninger over tid
 - Manglende dokumentation
 - Ringe incitament for arkivering af data