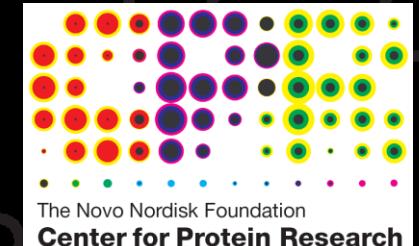


Nye muligheder med big data

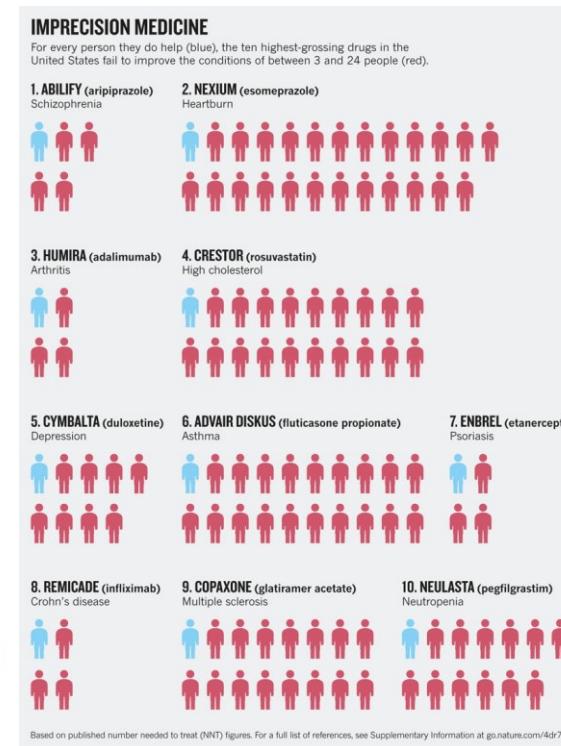
Søren Brunak, Novo Nordisk Foundation
Center for Protein Research

UNIVERSITY OF COPENHAGEN



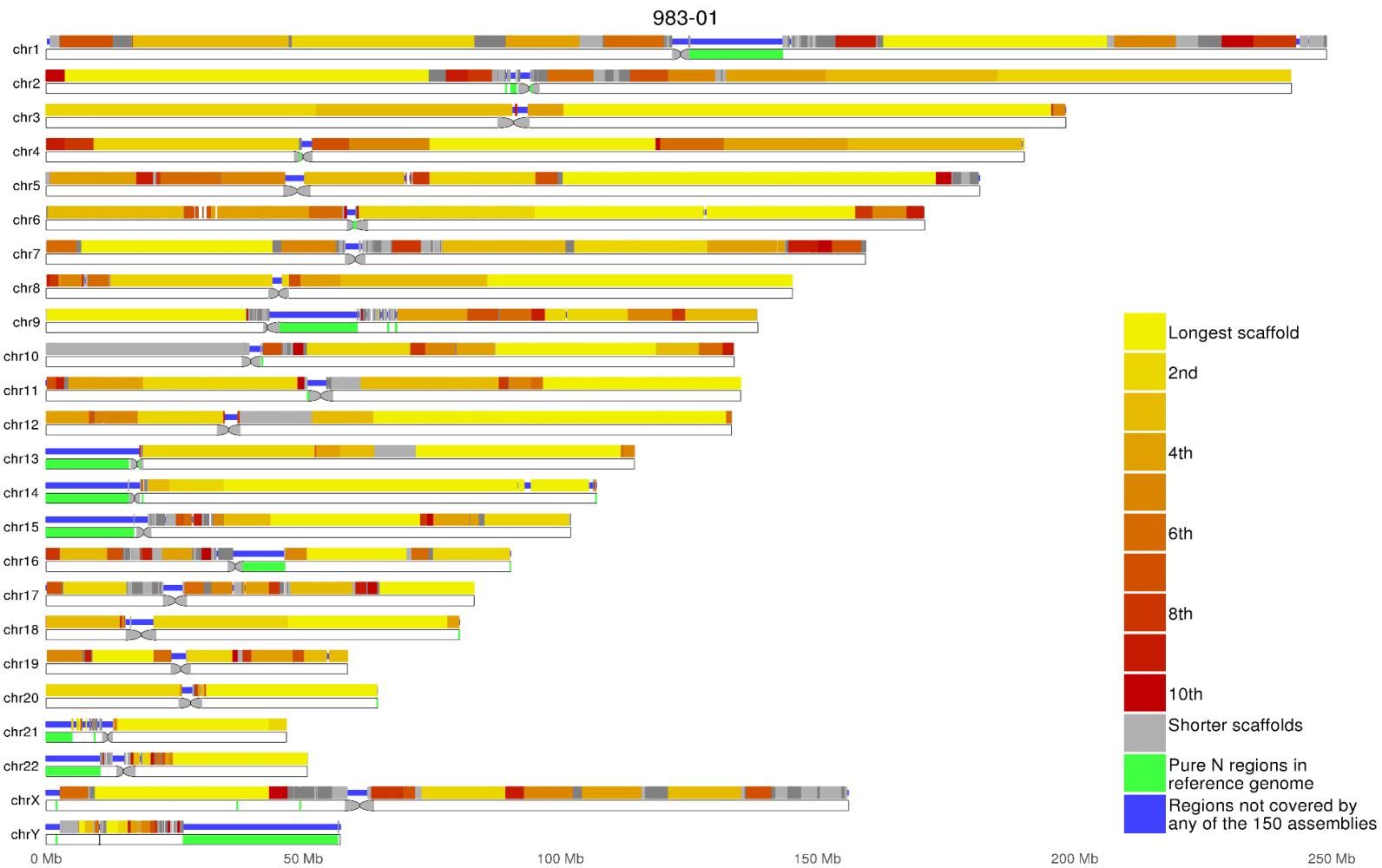


Imprecision medicine



Det "raske" danske referencegenom er færdigt

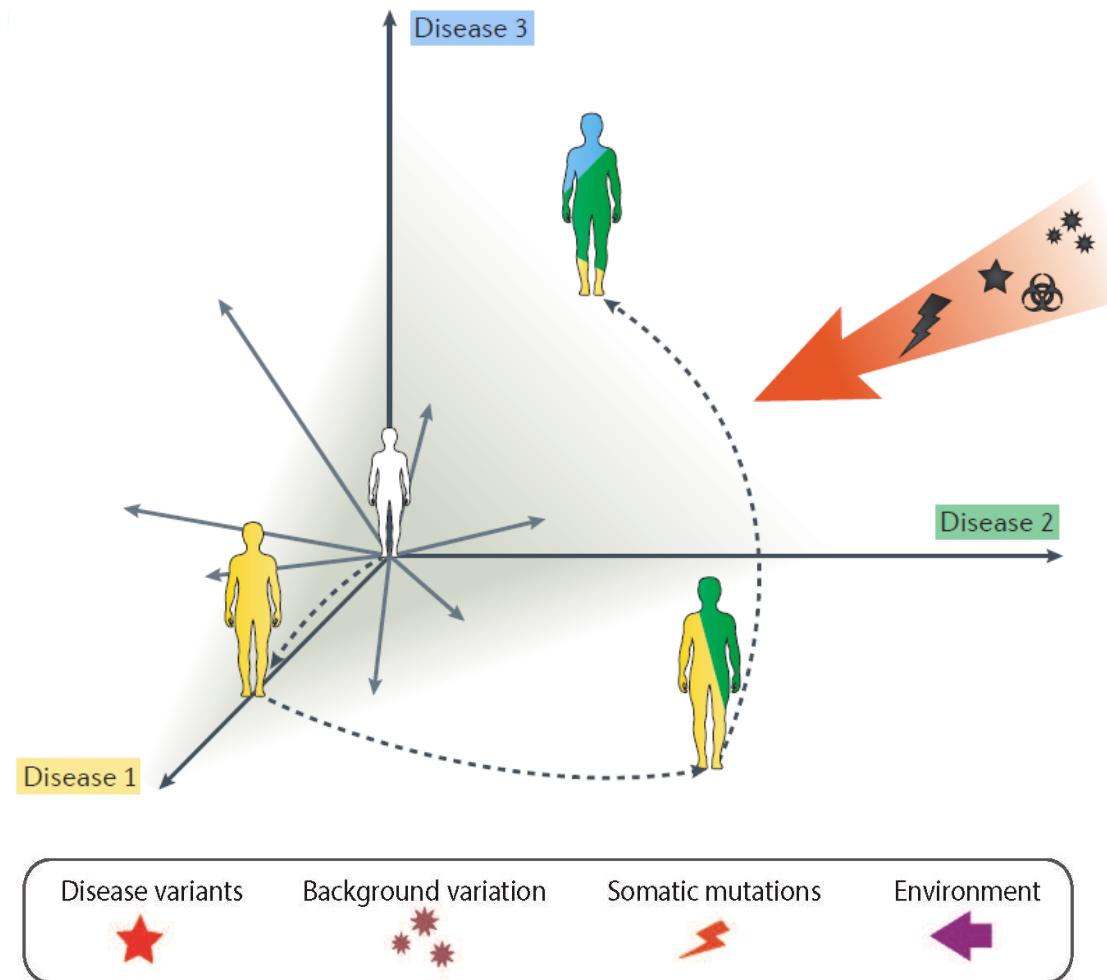
Et enkelt individuelt danske genom vist på den officielle "verdensreference"



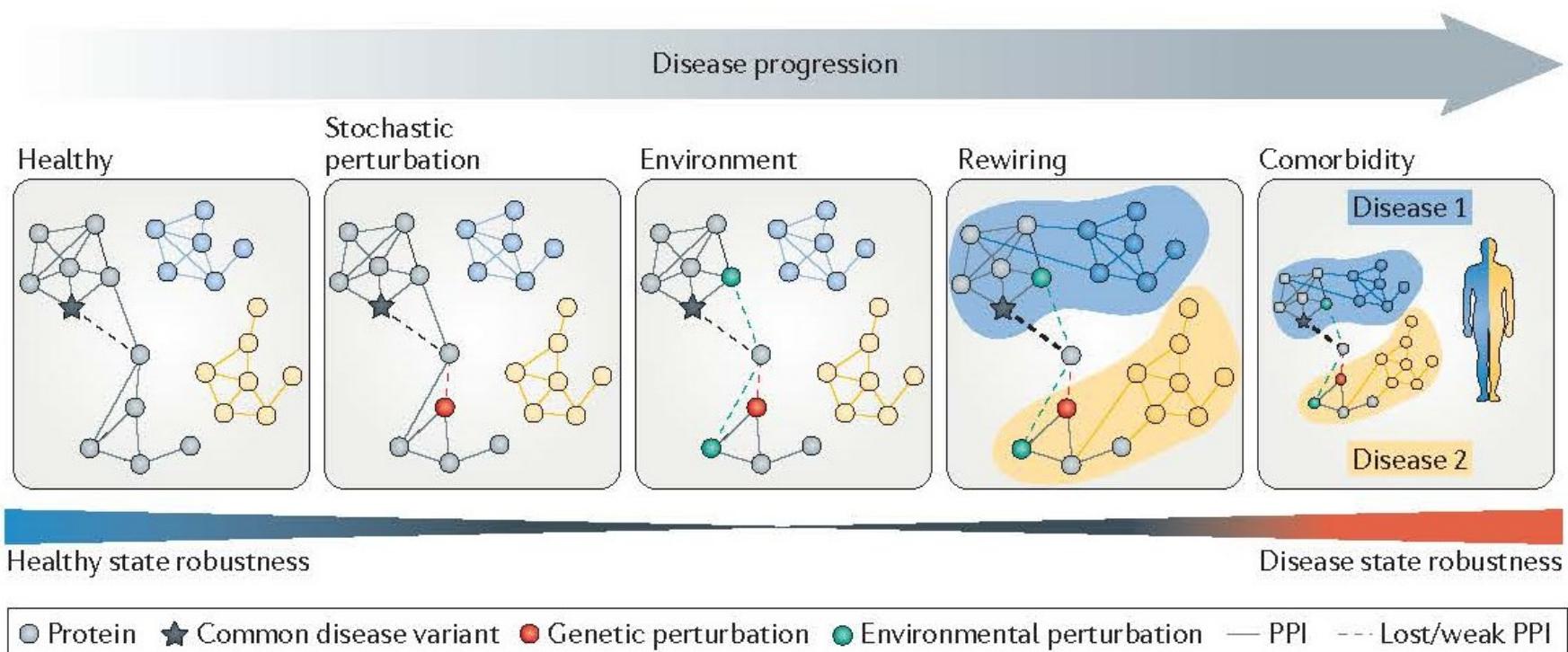
Sequencing and de novo assembly of 150 genomes from Denmark as a population reference.
Marety L, ..., Kristiansen K (*), Brunak S (*), Schierup MH (*). Nature 2017, to appear.

TCCAAACCAGGCTCTCCCCAACCGAGTTGCCAGATGCCAGTGAACCTCACTCTCATCAGTAAAAGGGGGCAGAGTGAGGTCTGAGAGCTAGTACAGGGACTGT
TGAAGTAGACAATGCCAGTGTAGCGTAAGAATCAGGGTCCAGCTGGCTCCCTAAACAGCAGCTGCTTCACTGTTGAAAGGCCTCTGAAGGCCAGGCCGGTGGCTCAT
GCTTGTAACTCCAGCACTGTGGAGGCCAGGTGGCGATCACCTGAGGTAGGGAGTCAGGACAGCCTGACCAACGTGGAGAAACCCATCTCTCTAAAAATAACAAAATTAGC
CAGGCCTGGTAGCACATACCTGTAACTCCAGCAGCTGGGAGGCTGAGGCAAGAGAATTGCTTGAACACCAGCAGGGAGTTGCTGAGCCAAGATCGAGCCATTGCACTCCAGC
AGGCACACAAGAGGCAAATGGCAGAACACTCCATCTCGAGAAAAAAAAGAATACTTCTGAAAGTATTATTACAAATAAGACTTGACCCATAAGGTAGAACGCA
ATGGGCCACGGAATCACTATTCCACAGTATACACCGAGTGCCCTGAAGTGCTGGCACTGCTCAGGATTGGGGCATATTGGTAAAAGAGAACGCAAGCCTGCTGCTCAGAT
GCAGGGAATGGGAAAAACAGGGAGACAGTTCTGTGAGATGTTGGAGTCGCTCAGTAGTATATTACTGAAATAGACCACTAACCTGGATGTCCTTTGGAAATG
TGCTGCGTCCAGGGCTGGGTTGGGGCCCAATGAACACTTGGCTCTGACATAGCTGTTGCCACACTCAGTGAACATGATCTTGCCTCACCGGCATCTCACCCAACT
CTCCCCGCCACAAACATACATCCCAGCCTGGGACCCCTAAAGGTGCTTCACTATTAGGTTGGCTGGCTACTGAAGTAAGTCTTGCACACTCAGAGGGATAGGAATTG
AATGAAGACATGAGATTCTCTGCGGGAGGCCTCTAGGAAATCTGTTGACTCACAGTTACTAATGTTGCTGCAGCCCCCACCTGGCCTGGCAGCCATACTTAGG
GCTTTGTAACTCTCATGTGAGGAACCTCAAATTAGACCTGGGTTGGAGGCAGTCCAGCTGGCTTTGGGGAGGTTTGTGCGAGGCATTCCAAAGTGTGCTGGCAGGATT
GTGTCACAGACACAGAGTAAACTTGTGAGGCTCCAAGTGACGCCATAGTTATTATAAGGTGACTGCACCTGCAGCACCAGCACTGCCTGGCTCACGTGCTCTGGTC
TCAGTATGGCGCTGTCCTGGGTTCTACAGTCTGAGCCTCTACCTGCTGGAAGCCCAGATCCCATTGTTGCAACCTAGTACCGGTGCCATACCAACGCCACCCCTGGACC
GGGTGAGTGCTGGCTAGCCCTGCTGAGCACATGGGAGCTGCCCTCTGGGCTCCCTTACCTGCTGGCTGTTGAGGACCCCCACTCCAGCTGCTCTTTCTC
TTCTGGGCCCCAGGGTGAATTCTCACAGCCCAGGGACTCTGGAGGCACCCCTGCCTCAAACACAGAAGCCTCACTGCAGAGTCCTCACGGAGGACGGTTGTGCTGGG
CTGGAGGGCTGCTGGGGCAATGACTGATCTCAGGGTGAGCTCTGCATGCGACTGCCACAGGGGCTCATCTCCCCTGAAACATCAGGGAGAGATGCTGAGTC
TCCTCCCAGCTGACAGTCAAAGATTGACATCAAGCCCCATCACCAGCTCCCCCTCTCCCAGATCACTGGCAAGTGGTTTATATGCACTGCCCTTCAAAACGAGGAGTAC
ATAAGTCGGTTAGGAGATCCAAGAACCTTCTTACTTACCCCAACAAGACAGAGGACACGATCTTCTCAGAGAGTACAGACCCGGTGAGAGCCCCATTCAAATGCAAC
CCGATCTCAGCTGCTGGCCAGAAGACCTGAGCAAGTCCCTCCTTCTGGCCTTGGCCTCCATGGTGGAAACCGGGAGGGTTGCTTAATCTCCACAGAACTCTGCC
CGGGACTGTGATGGGCAATTCTCCTCGATAACATTACTGTTTCTTCCGCTTCTGTTGACTTACAGGCAAGTGGTTTATATGCACTGCCCTTCAAAACGAGGAGTAC
AGCAGGAGAATGGGACCGTCCAGATACGGTGAGGGCAGCCCTCAGGCAGGGTTACCGTGGAAACAGGGCAGGCCAGCATAAGGTGGGGCTGGATGTAGAGCCCTGGAGG
CTTGGGACAGAGAAATAACCACTAACATTGTAGCTTACACGTGCTCAGAAAAAATCCCTAAGAAGACACTGAGAGAAATTAGATGAGGAAACATAAGAACAGAGACCTCAA
ATAGTTCCCAAGGTACACAGTTAAATTAGAACTAGAATTGAACTCCAGGCTGGCTTCACTGCTCTCTCACGCCCTTTAAGATCTTGCCTAAACCAATGGTAGA
AGCCTGTATGGAGAGGTTGACCTCAACTATGCTCCCATCACCGAGGGTGGCAGATGGCAGGGATCTGATGGAGCTGAACATGACATCTTACATCCCAGGCCCTCT
CTGGGCTCTTCTCTGTAAAACGGGAGAAAGGCCACTGACAGCCACAGTCTGAGGCTCTGAGATCTCATGTACAGAAAGTGTGCTGGCTGGAGCTGGCACGCAG
CAGGGGCTGGCACCGTGGCCAAAGGAGACCCGGCCTCACTGATGGCTTGTGGCCCCGGACACATTCTCTCAGAGGGCTTTAAGACCTTGTGTTGGCTTACCTGGACGATGAGAAG
CTTAGGGACACCCCTAGGACTCCTCACCTGTAAAGACAGGCACATTGTGCCATCCATGTTCTACCCAGAGGCTTTAAGACCTTGTGTTGGCTTACCTGGACGATGAGAAG
AACTGGGGCTGTCTTCTATGGTAGGCATGCTTAGCAGCCCCAAACTCATGCCCTCTCAGGCCTACCCCCCATTACCCACCCCTGGCTGGCCCTAGAACCCAGCCCTCCC
TGGCCTCCGGGCCCCACCATGTCAGTCAGTCTCCTGCTCCCCCTGAGCTGACAAGCCAGAGACGACCAAGGAGCAACTGGGAGAGTTCTACGAAGCTCTGACTGCTT
TGCATTCCCAGGTCACTGTCATGACACGACTGGAAAAAGGTAACGCAAGGGATTGGACATTGCCACCTTGTCCATGGCCAACTTGGCAGCCCAGAGGCCAGAGCAGGA
AAGCTGCCAGGCAAGGCTGCACAGCTAGGAGATCTCTGCTTTAGGCACCTGCTCACTGAGGACAGCTGAGCTACAGAGGCCAGGGGGTTGGATGAGGCCAGGAGG
GAGAAGTCCCTGTGAAACCAGGGAGGACCTGAAAGCTAACAGGAGGGACAGCGTGGACCCAGGGGTTGGGGATTGGCAATTGGAGGGACGTAATGCGGGAGTTACCACTACA
GACGCTCCAAACCCCAGGCTTCAACCTCCACTCCCCGCTCATTTTAATACCCGTGCACTGGGAAATTGATACTGTGTTTCAATGTCACCCACACTGCAGCACGGG
ACAGTCACCATCCGATTTGTACAAATGAAAATTACTGTATAATGAGCTCTAACACTTTCTTAAACCTGTGTTGAAGACTGTGTTGTGCTGCCCTGCTCTAATA
CCTGTGAAATCACAGCACCGATGAGCTGGTCCAATTAAATATACATGAGTACTTCCATGACTATTCAAAGAAAACAATTCTTCCATTGCAACCTGAGATGACCAC
AGGGATGTGAACTACCTCTGCCCATCCCCAGCCAGGATCCTGGACAGGGCTTATGAACGCAACCACTGAGTCAGCTACCTGATCCACAGCCTGGCACCTCACTGCTGG
CTAGGGAGCCTGAAATGGGCTCAAGGCCACCCCTGCTCAGTTACATCATCTGCACTAGTAGGTTGAGGAGGAAATTGAGGAGCTGCAGCATAAGGGCCCTGAGGTACTATG
TGCTCAGTAAATGCCAGTGGTCTTAAGGTCTGAGCTCCATTGTAGGGCAAGTAAGCTGAGGTTGAGAGAAGAAAATGACTGCTTCAAGATCACCCAGCTGGAAAGTGA
CCAGGGTGGAGCCCTGGTGGAGCTGGAGCCACTGGAGAACGAGCAGCACGAGAACAGGAGGGAGGGAACTCAGCAGGACACAGCCTGGATCAGGACAGAGACTTGGGGCCA
CTGCCTCTAGGATAAGTGTGAGGCCAGGAGGAGCTGACCTCAGCTTTCCCTCACTTGCATCAATAAGCTTGCATCGGCTTGGAAACAGGAGGAGTACAATAAGTCGGTT
CAGGAGGCCAGCATAAGGTGGGGCTGGATGTAGAGCCCTGGAGGTTGGCACAGAGGCCACCCCTGGACCCGGTGAGTGCCTGGCTAGGCC
TGTCCCTGAGCACATGGGAGCTGCCCTCTCTGGCTCCCTTACCTGCTGGCTGAGGCCACCCACTCCAGCCCCAACTCTCCCCGCCACAAACATACATCCATGCC
CAGGAGGGTTCACCGTGGAACAGGGCAGGCCAGCATAAGGTGGGGCTGGATGTAGAGCCCTGGAGGTTGGCACAGAGGCCACCCCTGGACCCGGTGAGTGCCTGGCTAGGCC

Den livslange multi-morbiditetsrejse i sygdomsrummet



Sygdomsprogression, gen-netværk, og sygdomskorrelationer



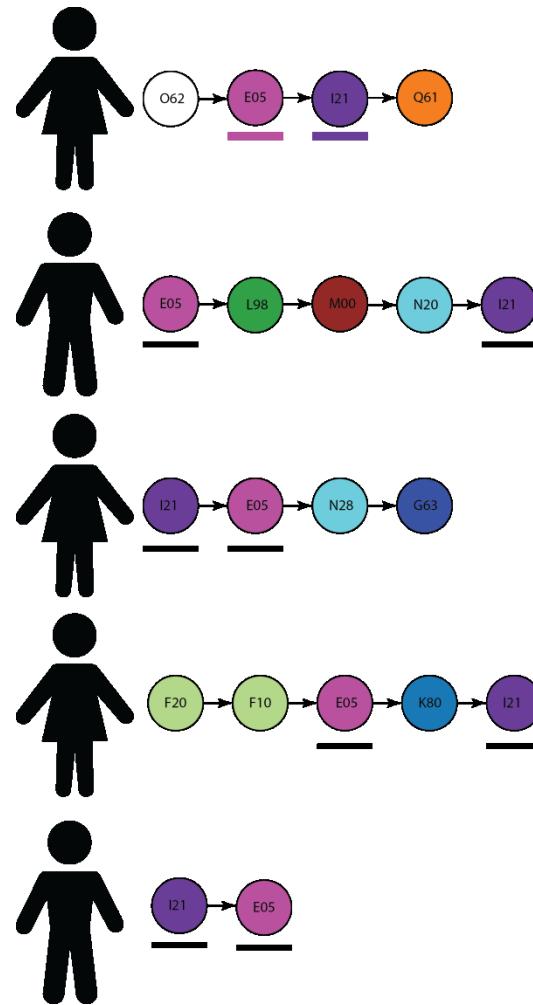
Væk fra primært at analysere sygdomme enkeltvis

Sygdomskorrelationer

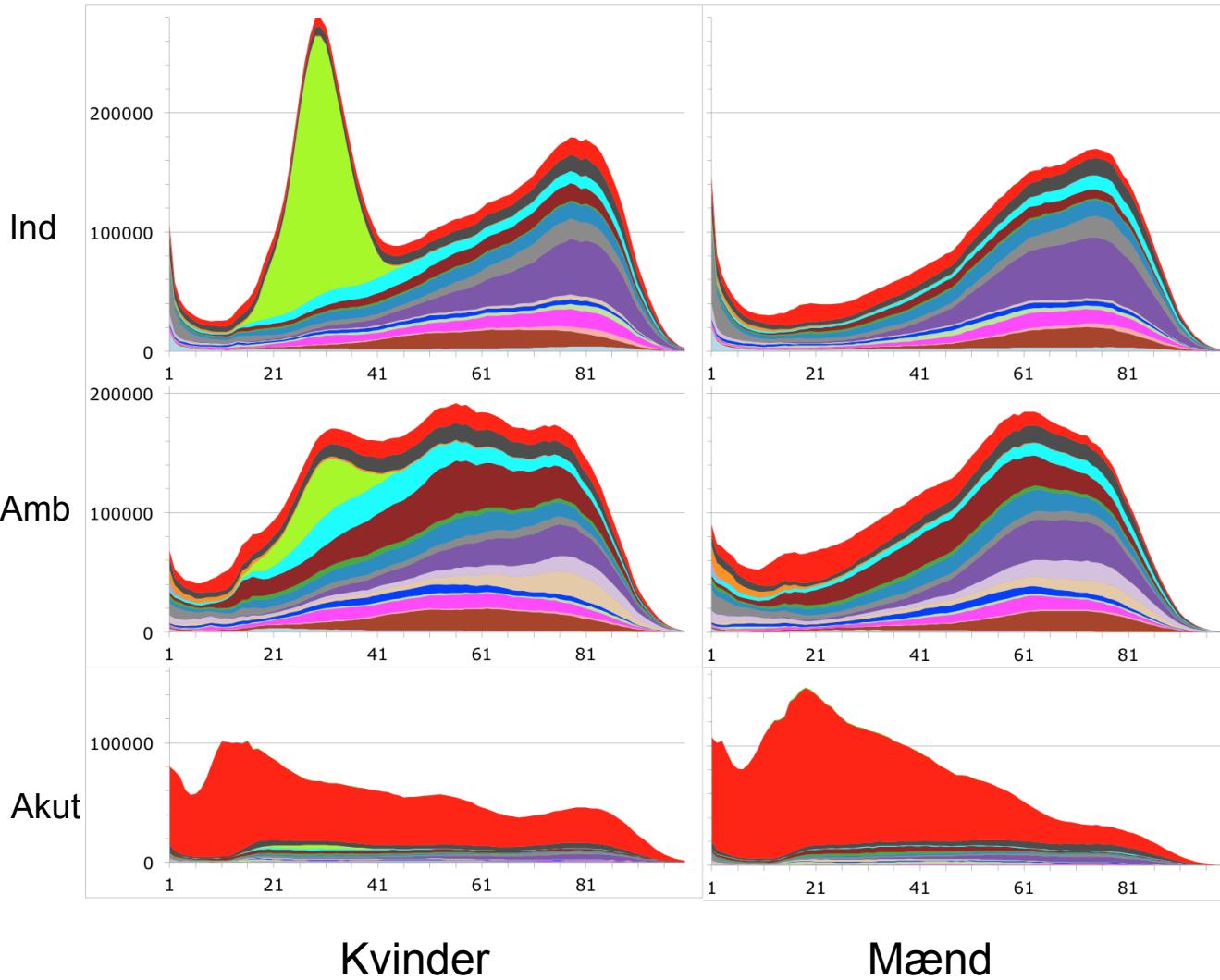
Sygdomstrajektorier

Sygdomstrajektorier for 7 millioner danskere

(ICD-10 era, 1994-2015)



Landspatientregistret (7M danskere) ICD-10 diagnose som funktion af alder



ICD 10 chapter coloring

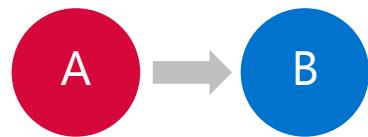
- 1: Certain infectious and parasitic diseases
- 2: Neoplasms
- 3: Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
- 4: Endocrine, nutritional and metabolic diseases
- 5: Mental and behavioural disorders
- 6: Diseases of the nervous system
- 7: Diseases of the eye and adnexa
- 8: Diseases of the ear and mastoid process
- 9: Diseases of the circulatory system
- 10: Diseases of the respiratory system
- 11: Diseases of the digestive system
- 12: Diseases of the skin and subcutaneous tissue
- 13: Diseases of the musculoskeletal system and connective tissue
- 14: Diseases of the genitourinary system
- 15: Pregnancy, childbirth and the puerperium
- 16: Certain conditions originating in the perinatal period
- 17: Congenital malformations, deformations and chromosomal abnormalities
- 18: Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
- 19: Injury, poisoning and certain other consequences of external causes
- 20: External causes of morbidity and mortality

6-7 million trajektorier

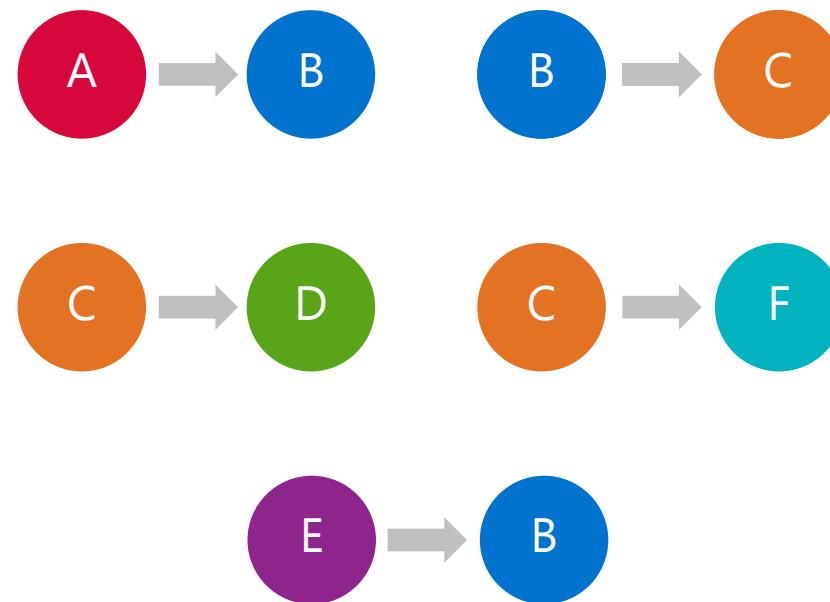


Hvilke er hyppigst ?

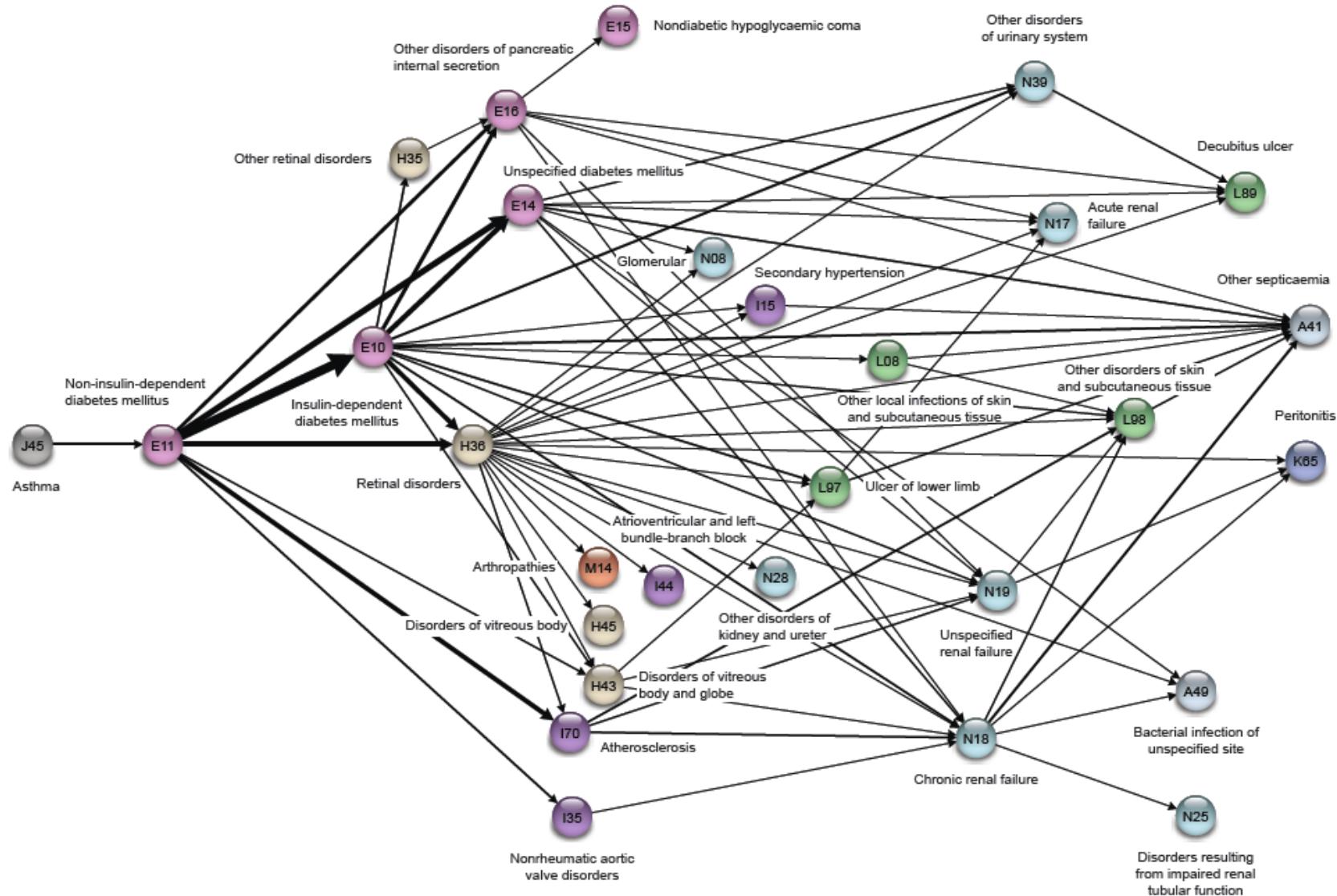
Hvilke diagnoser får man efter hinanden ?



Sygdomstrajektorier fra diagnose-par med signifikant retning

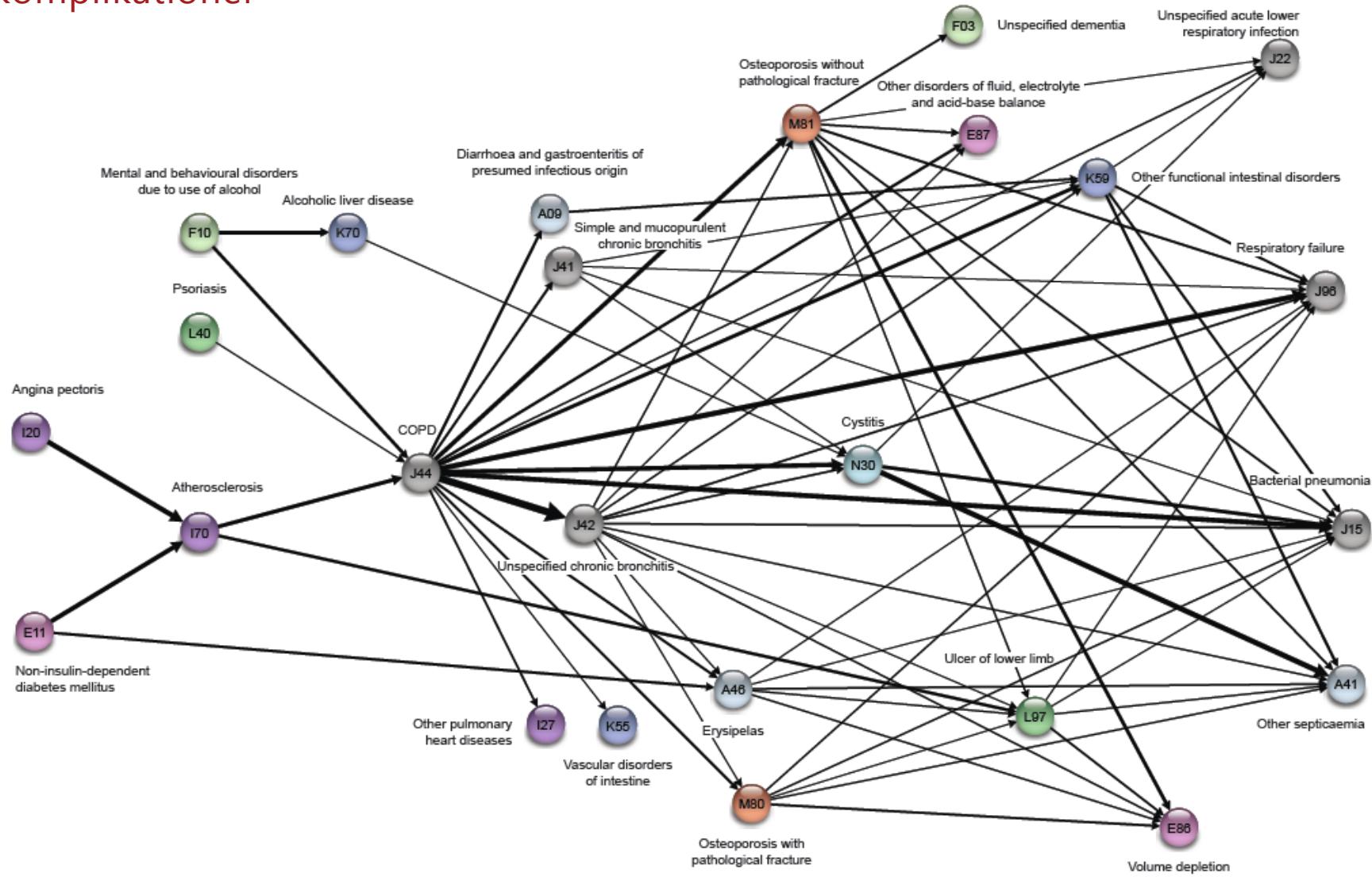


Diabetes trajektorie-netværk

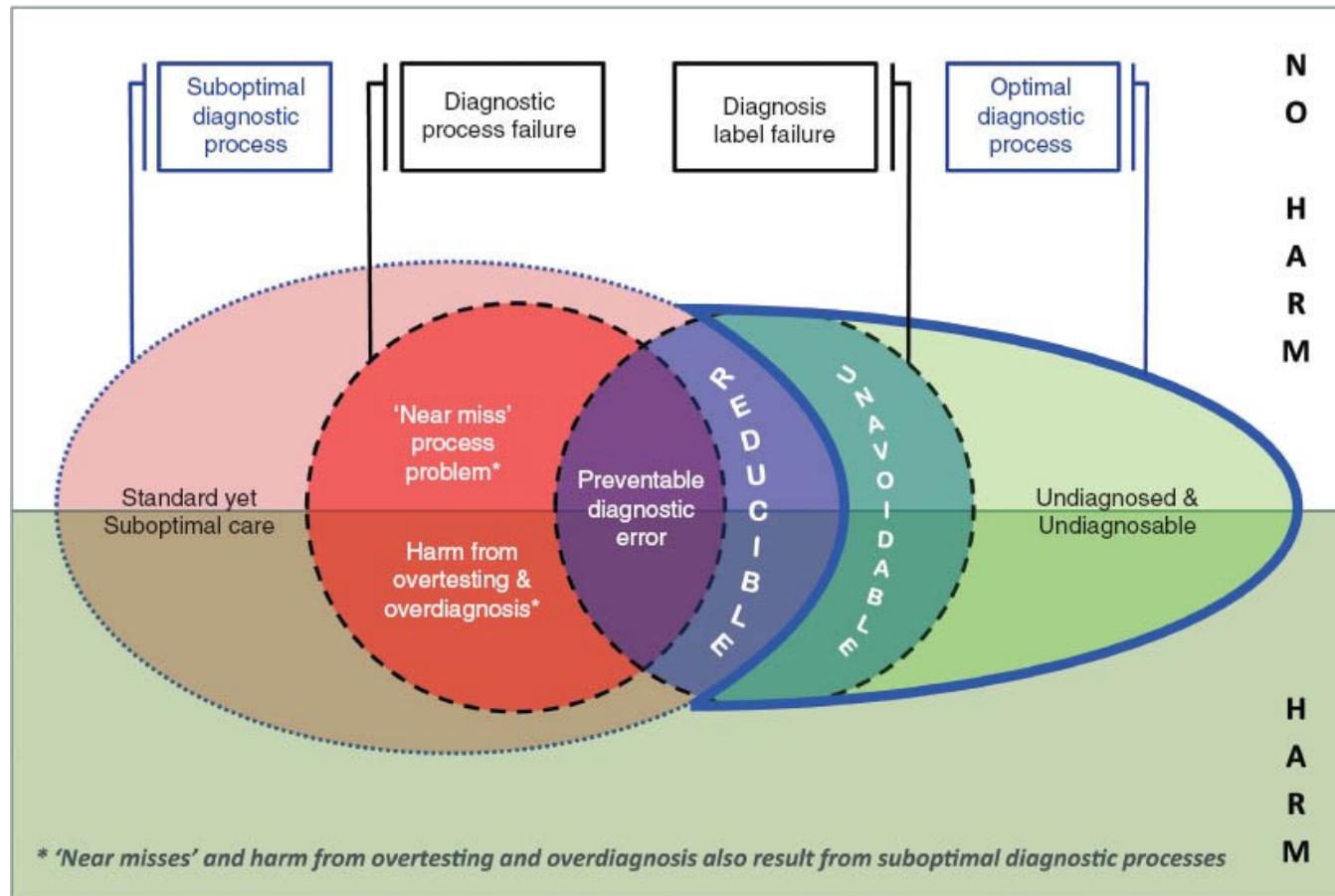


KOL trajektorier

med fem forudgående diagnose og nogle af de respiratoriske komplikationer

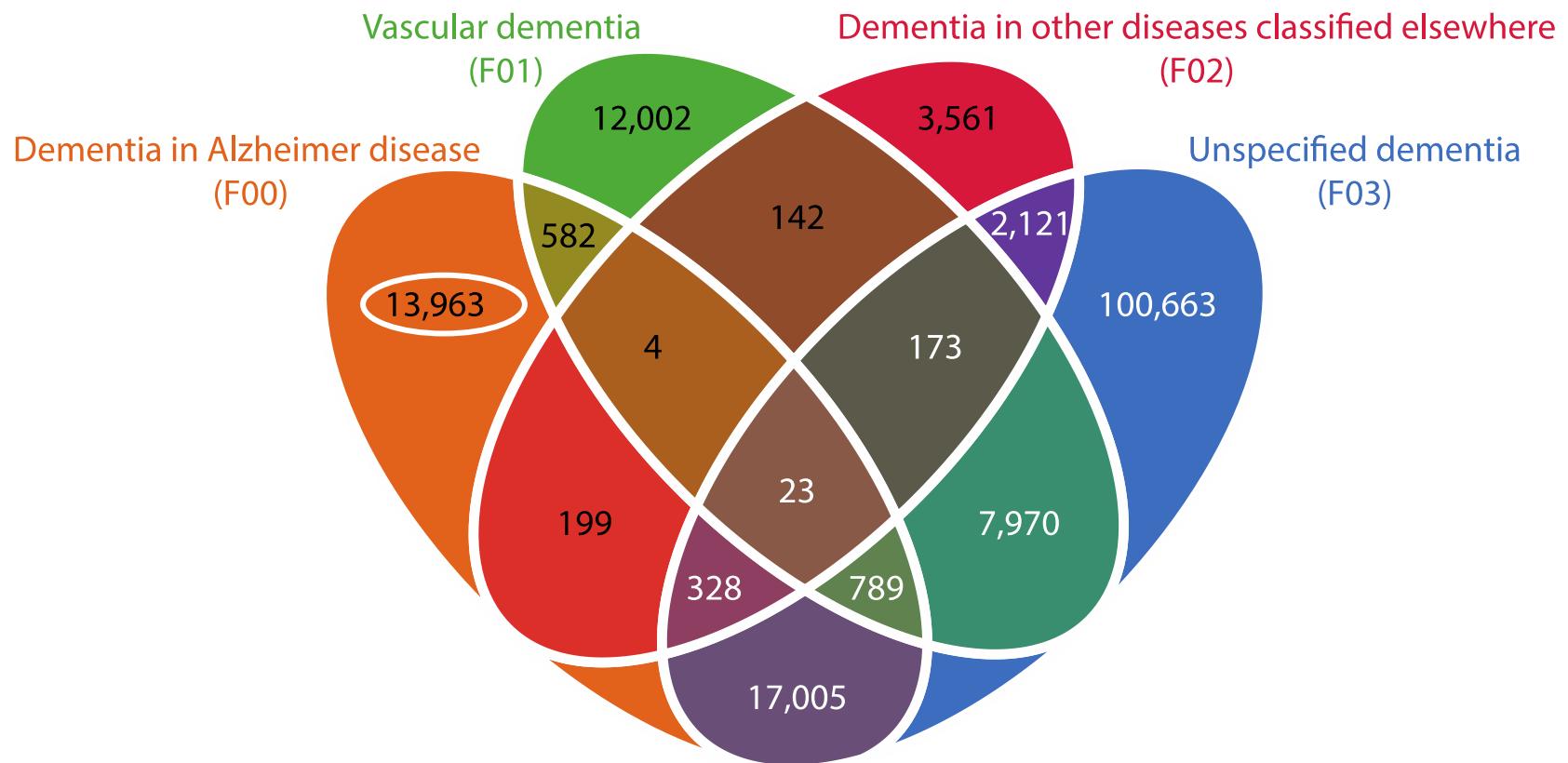


Precisionsmedicin skal reducere både overdiagnoser og underdiagnoser

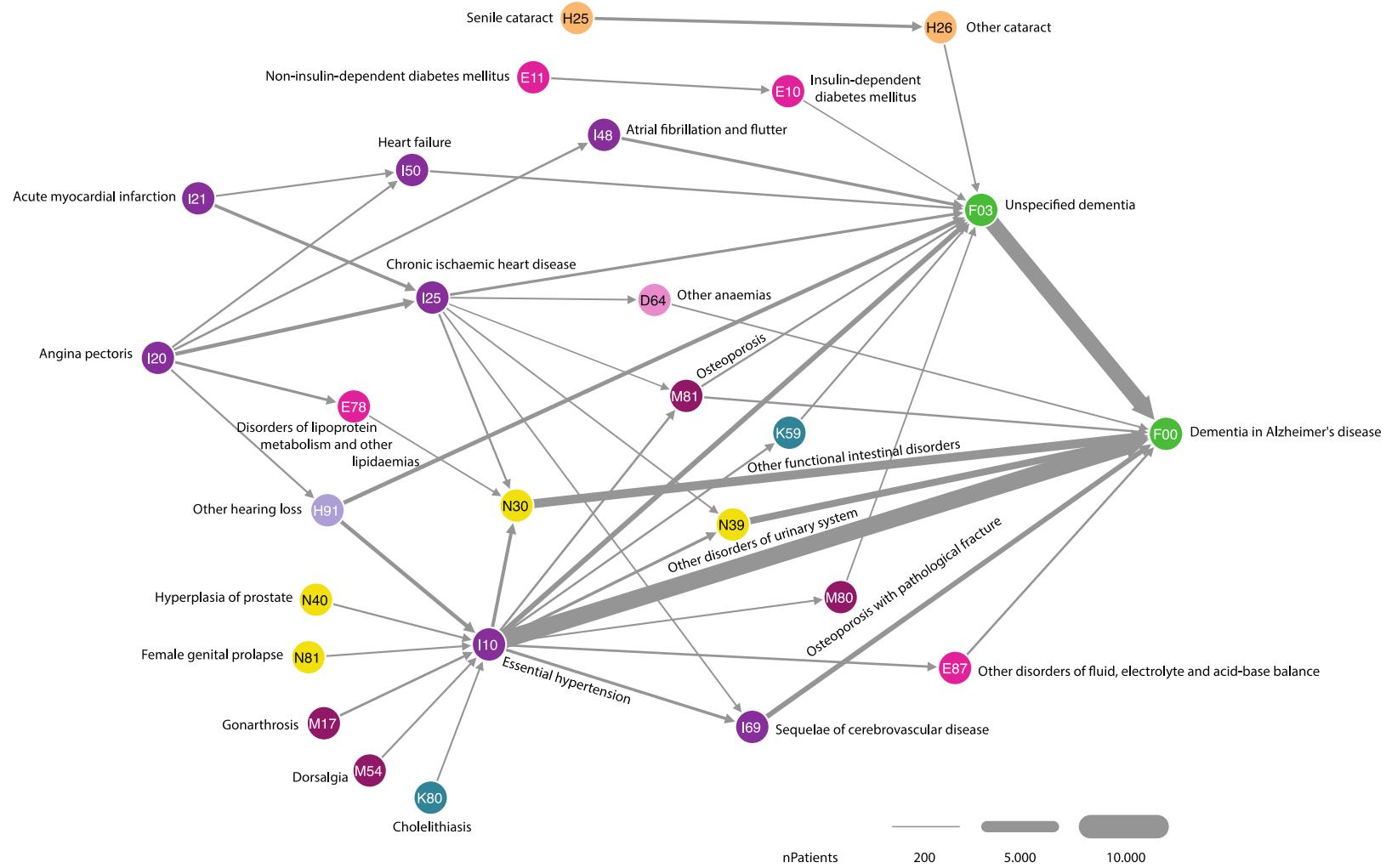


A unified conceptual model for diagnostic errors:
underdiagnosis,
overdiagnosis, and
misdiagnosis,
Diagnosis, David E.
Newman-Toker 2014|

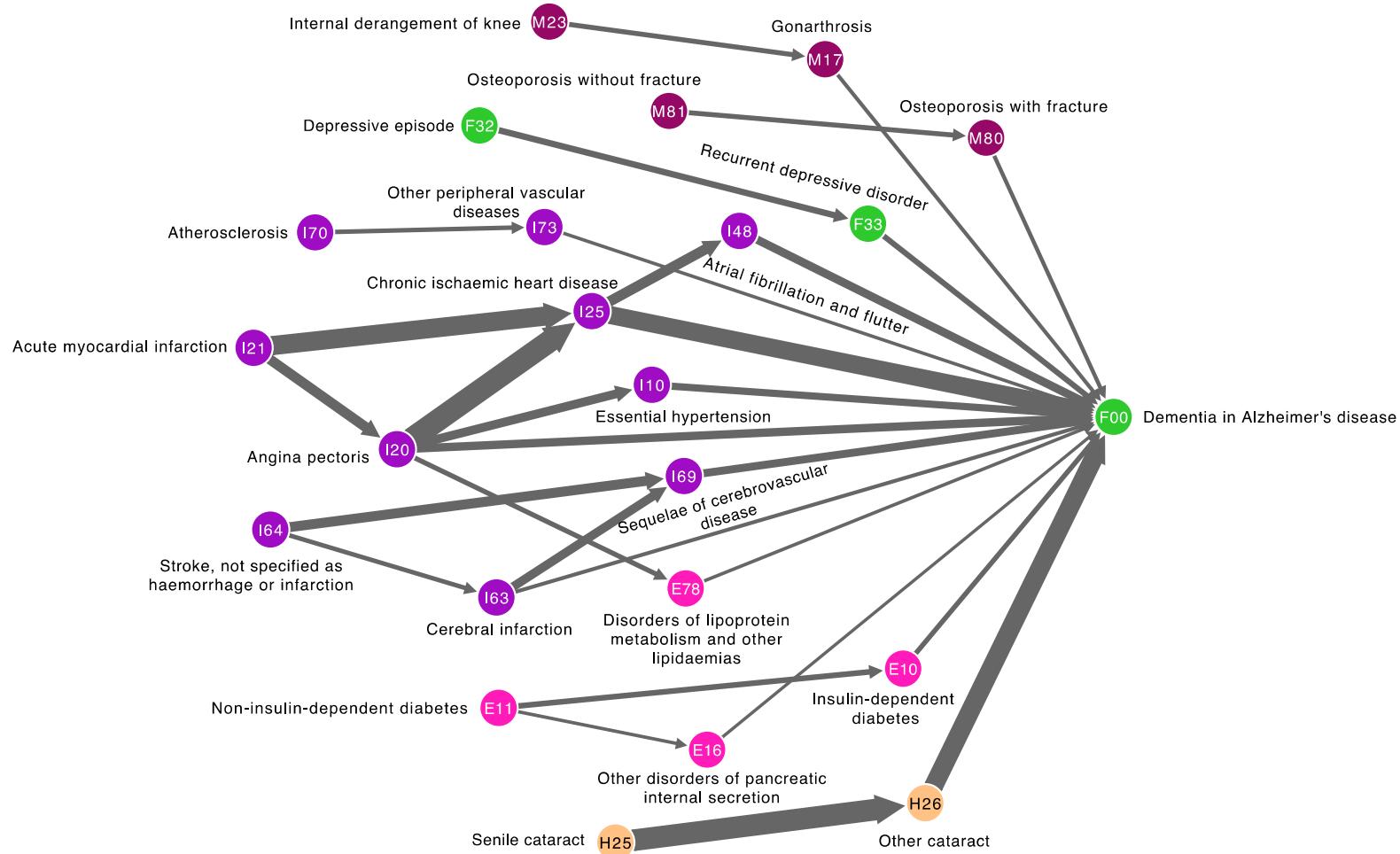
Demens: registerdata er ofte ikke tilstrækkeligt specifikke



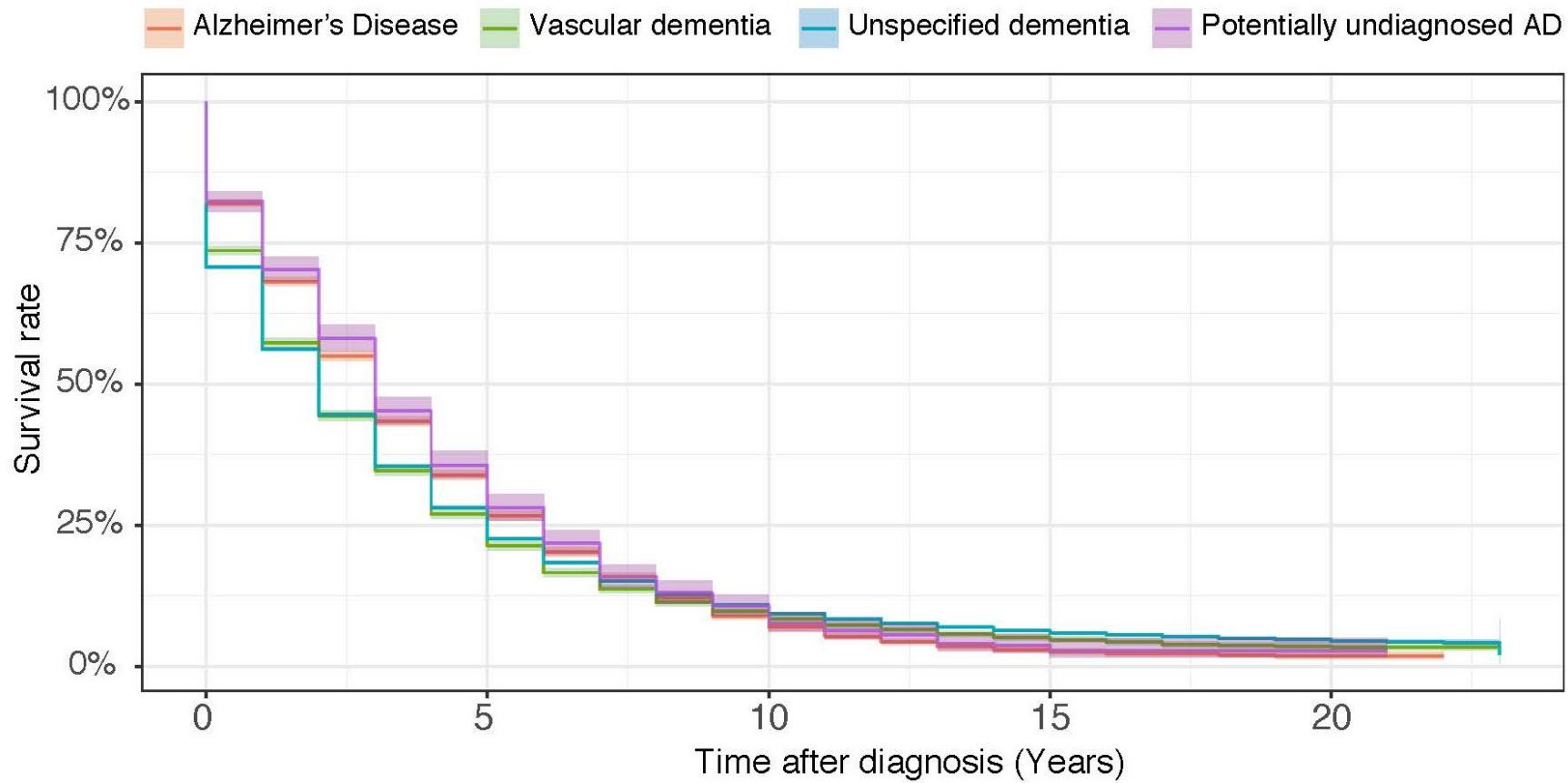
Alle patienter med F00, Demens i Alzheimer's sygdom



F00 patienter (ikke F01 Vaskulær/F02 Andre)



Overlevelse af demens patienter



DISEASE TRAJECTORY SEARCH:

ALL DIAGNOSES (UNION)

SEARCH:

FILTERS ▾

EDGE ANNOTATION:

PATIENTS RELATIVE RISK OFF

NODE ANNOTATION:

ICD CODE TEXT DESC. NONE

 INSTANT SEARCH PERFORMANCE ISSUES?

SEARCH

Information

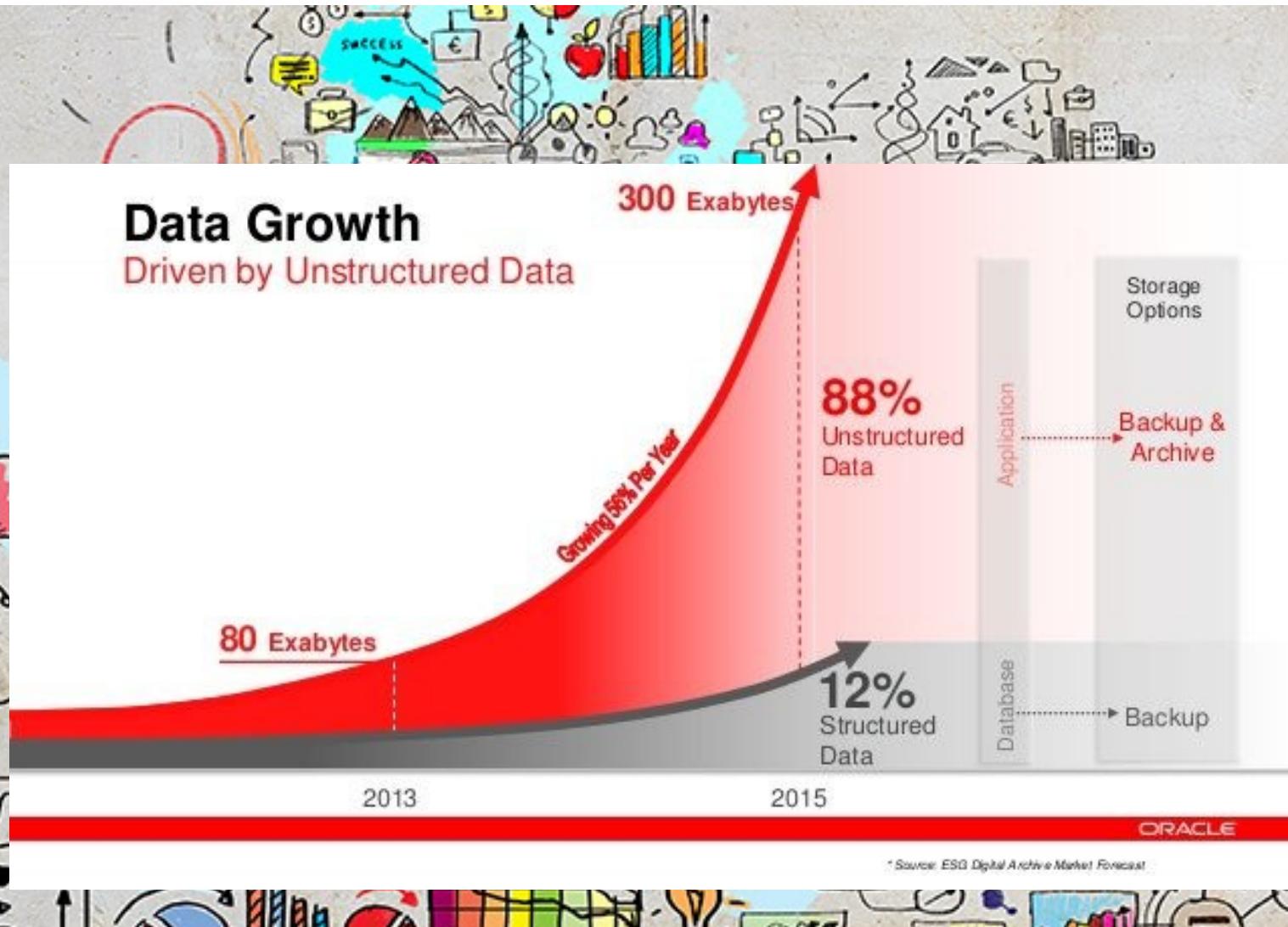
Data from: Danish National Patient Register (Landspatientregisteret)

Population: ~6,900,000 people

Fra registerdata til big data



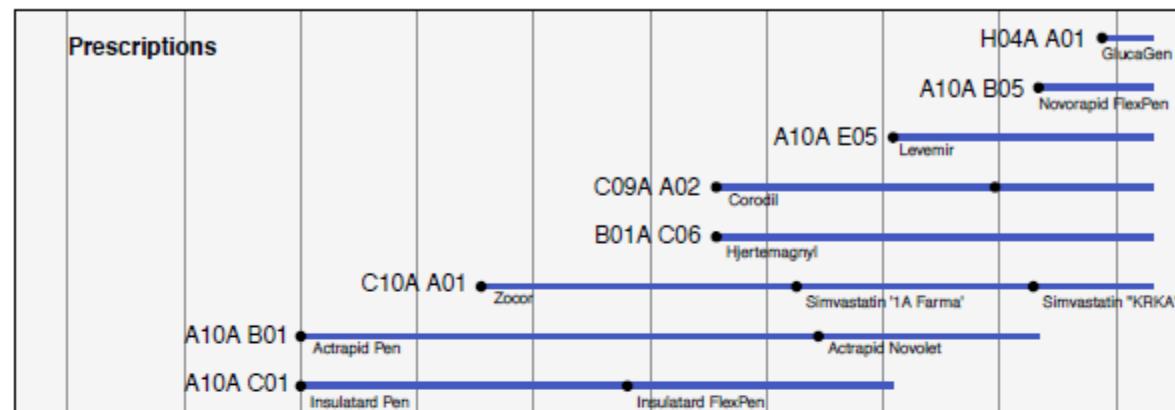
Fra registerdata til big data



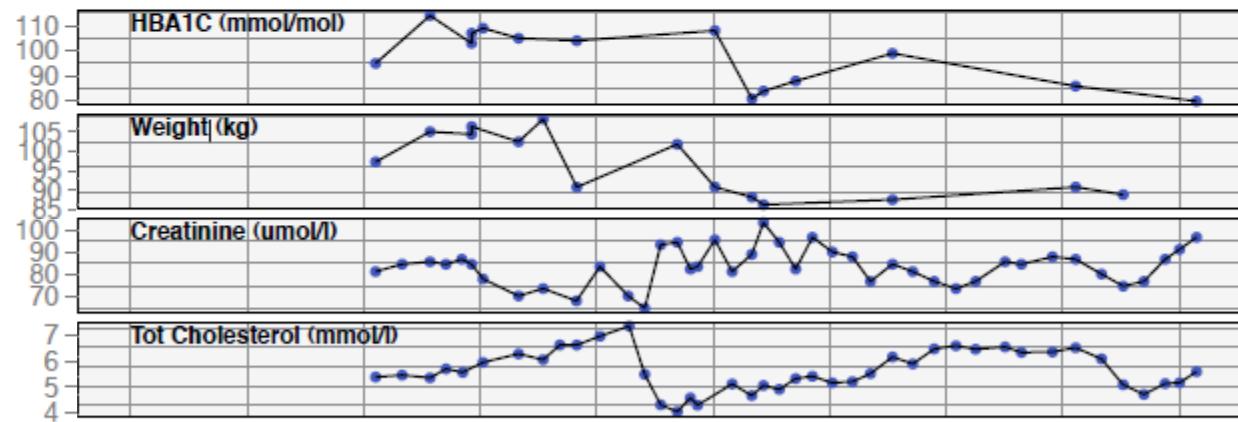
Diagnoses



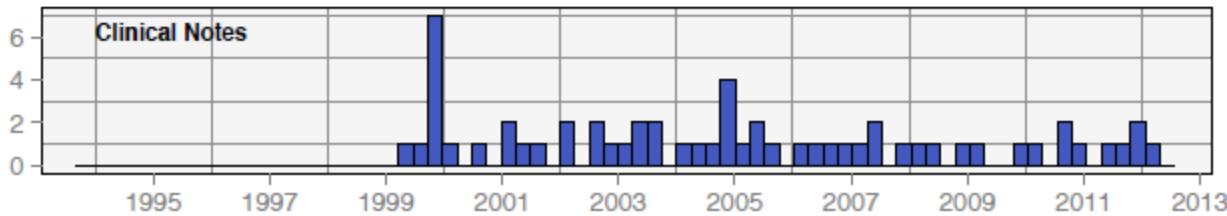
Drugs



Lab tests



Text



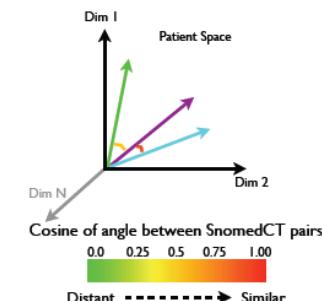
Præcise fænotyper fra text-mining af elektroniske patientjournaler

F20

F200

Deidentificeret
Journal tekst

det drejer sig om en 36-årig sygemeldt mand der overflyttes fra frederiksberg hospital, afdeling m.h.p. længerevarende rehabiliteringsophold. , er allergisk overfor kat og parfume, men tåler penicillin. er i besiddelse af en vis indsigt og virker svært forpint. ang. det at vi tilråder, at hun har brug for at være mere i afd. , siger hun til det, at det for hende er som at vælge mellem pest eller kolera. Har stadig mange spørgsmål omkring **skizofreni** og er meget bekymret for hvordan hendes fremtid ser ud. er meget plaget af tanketræghed og er bange for at det er et led i sygdommen. der siges til hende at det godt kan være bivirkning af risperdal men at der ikke laves om på medicinen, før vi har lært hende bedre at kende. Har **aldrig** haft **hallucinationer** på nogen af sanserne har været til lægesamtale idag. der snakkes en del om diagnose og at pernille har svært ved at forholde sig til at have diagnosen **skizofreni**, det virker som om pernille er blevet lidt mere afslappet, selvom hun stadig har gang i mange ting. pt. møder til samtale i dag, hvor vi gennemgår mit udkast til erklæringen til pensionskassen. endvidere udspørges der til pt.s diverse symptomer på **paranoid skizofreni**. i denne beskriver hun at "hendes største problem nok er den manglende sociale evne, som er en følge af sygdommen (**paranoid skizofreni**) og henviser til contras beskrivelse" Pt. Nævner sin **mor**, som han mener har en nervøs lidelse, muligvis **social fobi** pt. har her til aften angivet tiltagende **bivirkninger** i form af trækninger i nakken, indre uro og stivhed af fingre. pt. har fået svar på sit ekg, som viser sinus rytme med enkelte **ventrikulære ekstrasystoler** uforandret fra tidl. med baggrund i oplysninger om tidligere maniske episoder præget af irritabilitet, hyperaktivitet og øget seksuel interesse revurderes diagnosen til **bipolar affektiv sindslidelse**. følges i distrikt vest med psykologsamtaler. har i dag tydeligvis brug for en faglig forklaring på hendes symptomer. det drejer sig om **paranoia**, uvirkelighedsfølelses, influenssympt. og koncentrationsbesvær. det største problem er dog samværet med andre. det er specielt om natten det påvirker hendes **astma**., klg. desuden over uro i benene. ,xxx nævner på et tidspunkt, hun er bange for, tidlige tiders **spiseforstyrrelser** er ved at dukke op igen. xxx har haft **søvnbesvær** og har af vagtlægen i aftes fået tabl. imovane 7,5 mg med god effekt. kl 19, pinex, tabletter 500 mg indtaget dosis: 1 gram for hovednigne. pt. er henviset til at



Negation

Familie



BigTempHealth

Komplet journal-dump fra to Danske regioner (> 3.1M patienter)



Region Hovedstaden

2.14M patienter, 15.3M kontakter
2008-2017

1.35 milliarder journal sætninger

37.4M Diagnoser, 6.30M Operationer,
80.2M Undersøgelser, behandlinger,
procedurer

125M lægemiddelordinationer

Laboratoriesvar fra 5.3M patienter

Region Sjælland

899K patienter, 6.42M kontakter
2008-2017

514B milliarder journal sætninger

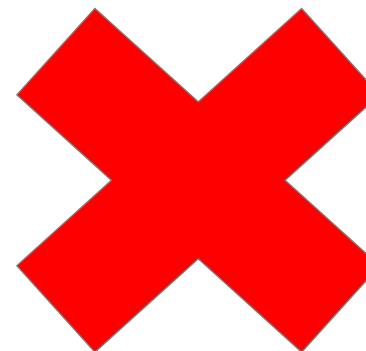
17.2M Diagnoser, 2.64M Operationer,
28.3M Undersøgelser, behandlinger,
procedurer

65M lægemiddelordinationer

Laboratoriesvar fra 2.2M patienter

Hvilke sygdomskorrelationer og symptomer er behandlingsrelaterede?

Polyfarmaci



Text mining af lægemidler, mulige bivirkninger, diagnoser

Removed ADR - no corresponding structured data

Behandlet med Zyprexa 5 mg fra 3. til 24.6.99 og 10 mg fra 24. til 29.6.99 med nogen effekt på tankeforstyrrelse; men seponeret pga appetitøgning. Herefter Risperdal 2 mg stigende til 4 mg i perioden 29.6. til 12.7.99, men seponeret på grund af uro i knappen og "ostekoldkæremimmelse". Herefter Orap 2 mg fra 2.8. stigende til 3 mg fra 30.8.99 med god effekt på tankeelko og tankemylden. Behandlet med Zoloft 50 mg fra maj 98 til maj 99 med noget virkning på depressive symptomer; men seponeret på grund af natlig svædtendens. Siden 14.7.99 Efexor 75 mg med nogen effekt på antallet og sværhedsgraden af kortvarige depressive episoder

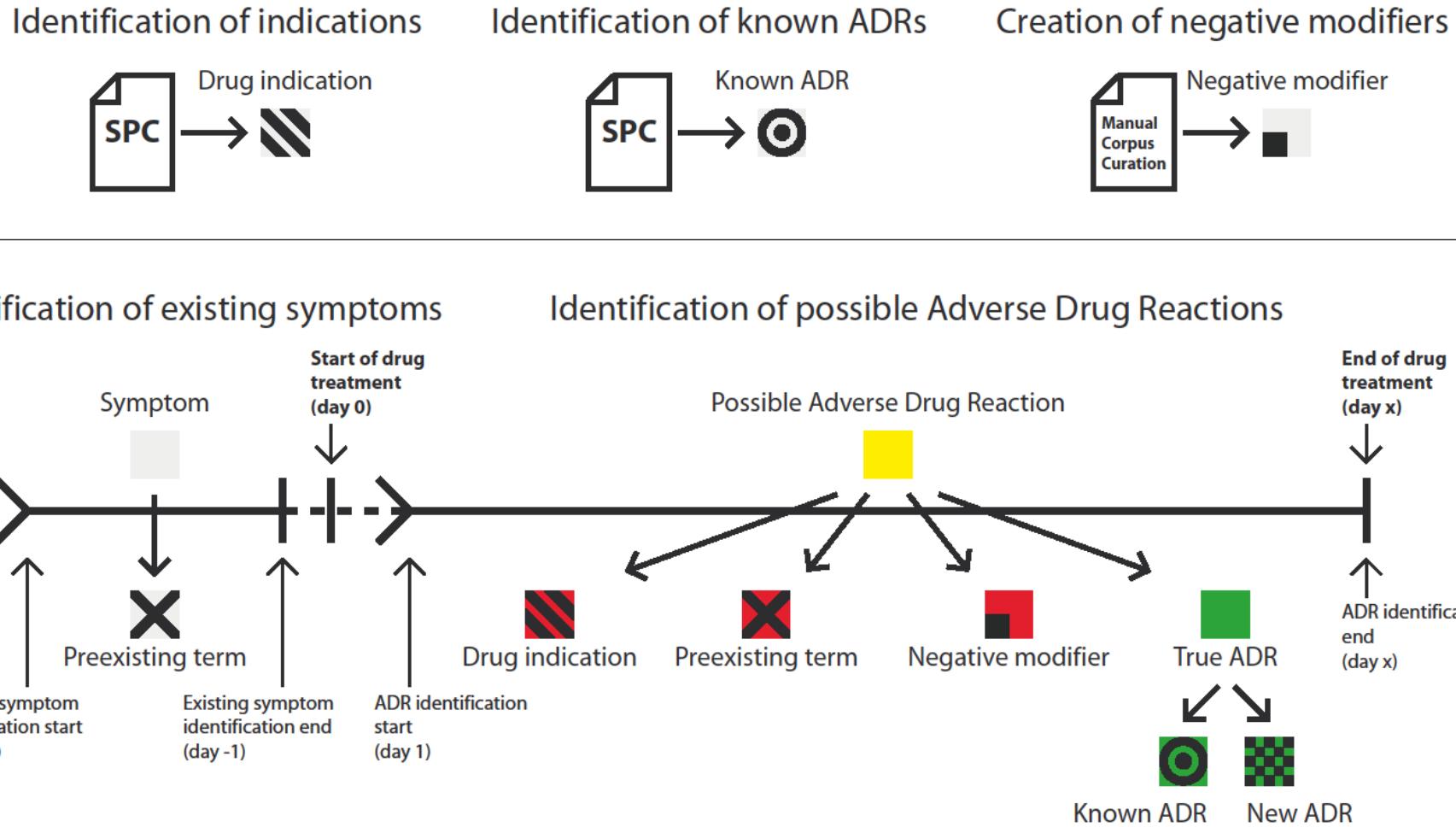
se venligst under allergier. Desuden forsøgt beh med Zyprexa, sep grundet vægtøgning, træthed og manglende effekt. Risperdal ord med nogen effekt, tillagt dogmatik (1999). Efterfølgende aurorix beh seponeret i 1999. Startede istedet remeron. Aktuel medicindosis, jvf udskrivningsnotat fra U12 samt EP/medicinliste.

tbl. leponex 100+0+0+200 mg, tbl. rivotril 0,5 +0+1+0 mg, tbl. arintapin 0+0+0+30 mg, tbl. domipramin 0+0+0+25 mg, tbl. imoclon 7,5 mg nocte, tbl. rivotril 2 mg p.m. max x 1 dgl, tbl. marevan a 2,5 mg efter skema, tbl. magnesia 1 g p.m., laxoberal dr. 7,5 mg /ml 15 ml p.m. max. link 150 mg /ml 15 ml p.m. max x 3 dgl. Figensaft 20+0+20+0 ml/Pt. er aktuelt, CAVE: tricykliske antidepressiva. Dette kan dog ikke bekræftes og pt. har tidl. fået imipramin, som han har fået godt, hvorfor der er ansægt om opfævelser af denne cave på højere niveau. Har tidl. fået zyprexa som blev sep. grundet vægtøgning, træthed og manglende effekt

Removed ADR - negation and subject identification

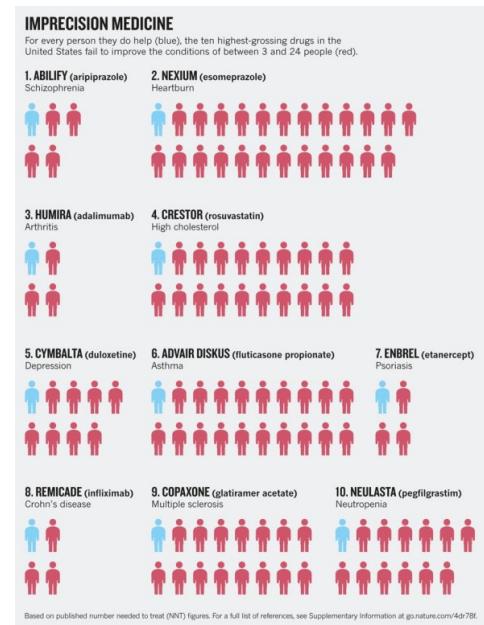
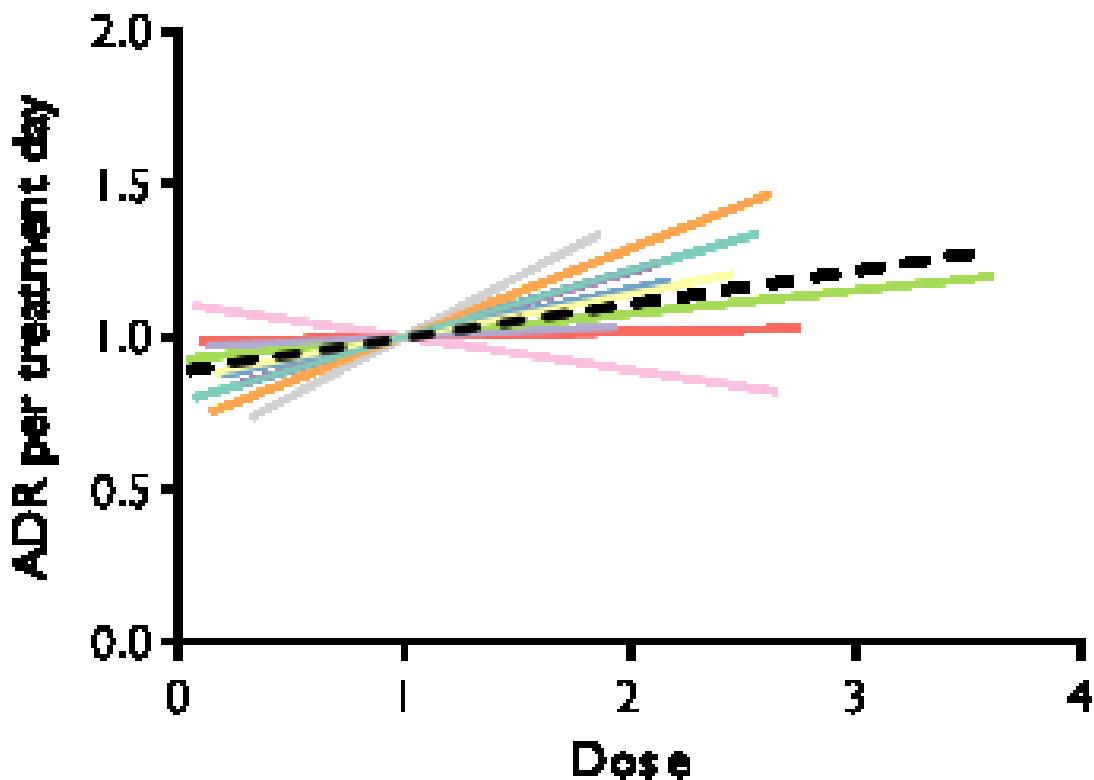
-Jeg mener fortsat, at han har brug for medicin, da han i går fx var meget vred og følte sig utryg og angst og har haft svært ved at sove. Dette synes pt. at accepterer. Jeg tilbyder herefter Zyprexa i stedet for Risperdal, pt. avisér dette, da han ved med patient har fået denne medicin og har fået øget appetit, dette vil han ikke. Har ikke tidligere fået antipsykotisk medicin. Accepterer hereafter Cisordinal, startende på en lille dosis. Accepterer også angstdampende medicin i dættiden. Angiver, at når han bliver vred kan han godt styre det. Synes det hjælp i går noget at få Nozinan. Virker fortsat garderet. Siger intet uopfordret. Sparsomt sygdomsindsigts. Er ikke medsladen. Er i dag heller latent aggressiv-

Text mining's workflow (7,500 lægemidler og 21,000 mulige bivirkninger)



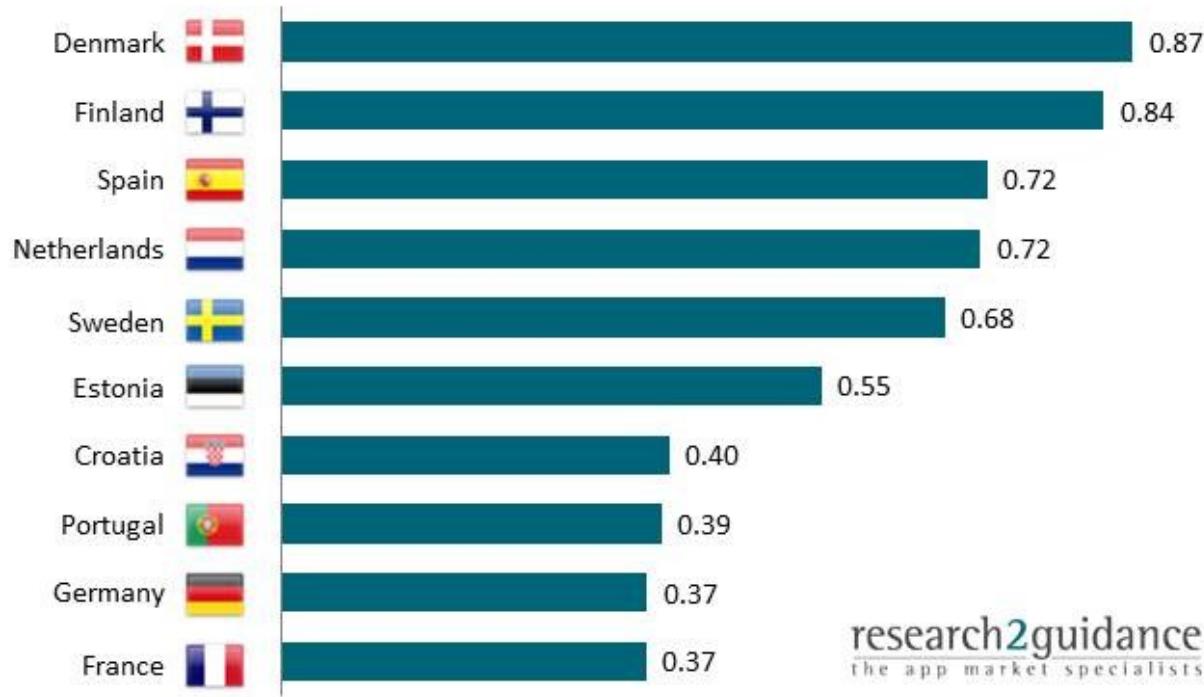
Sammenhængen mellem bivirkninger, dosis og lægemiddelkombinationer fra big data

Nature 520, 609–611
(30 April 2015)



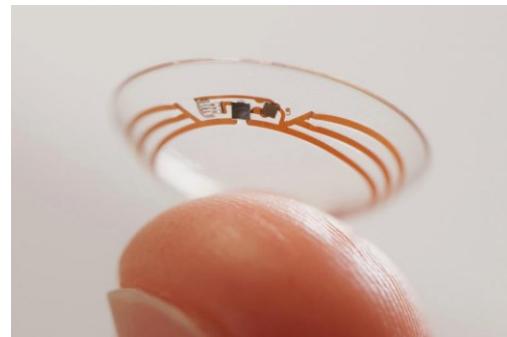
DENMARK IS THE LEADING COUNTRY IN EHEALTH ADOPTION

Top 10 EU countries by eHealth adoptions of patients and doctors

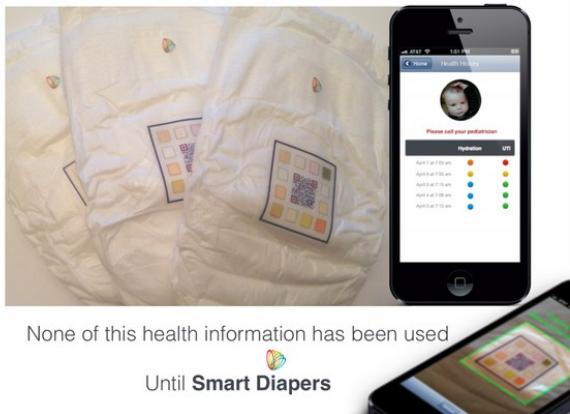


research2guidance
the app market specialists

eHealth adoption – doctors transferring prescription electronically, doctors electronically exchanging medical patient data with other healthcare professionals , patients making appointment via website, patients seeking online information about health



360 million diapers are changed every day



None of this health information has been used

Until Smart Diapers



Δ big data

- Sundshedsdata:
 - Redefinere fænotyper som trajektorier
 - Omklassificere under og overdiagnosticerede patienter
 - Bedre grundlag for at forudsige sygdomme?
 - Personlig medicin og genomdata
 - "Live data" versus registerdata
 - Patientgenererede data
- Inkludere hvad der ikke står i patientjournalen på nye måder:
 - Fødevareprofiler
 - Indkomst, ...
 - Uddannelse, ...
 - Wearable data

The screenshot shows the homepage of the 'Genome Medicine' journal. At the top, there's a search bar with the placeholder 'Search this journal for'. Below the search bar are navigation links: Home, Articles, Authors, Reviewers, About this journal, My Genome Medicine, Subscriptions, and Advanced Search. The main content area features a highlighted article titled 'Musings: The \$1,000 genome, the \$100,000 analysis?' by Elaine R Mardis. The article includes sections for Top, Musings, Competing interests, Acknowledgment, and Reference. To the right of the article, there's a sidebar with 'Viewing options' (Abstract, Full text, PDF [220KB]), 'Associated material' (PubMed record, Reader's comments), and 'Related literature' (Articles citing this article on Google Scholar, on PubMed Central, Other articles by authors on Google Scholar, and an ORCID ID). The footer of the page includes the journal's name, volume information (Volume 2, Issue 11), and copyright details (© 2010 BioMed Central Ltd).



COMPUTEROME

THE DANISH NATIONAL SUPERCOMPUTER FOR LIFE SCIENCE

Computing Power: **16,000+ cores**

- 525 x Thin nodes
- 28 core Intel CPUs with 128 GB DDR4
- GPU Node
 - Nvidia K40
- 29 x FAT nodes
- 32 core Intel CPUs with 1TB DDR3 RAM

Storage: 8PB+

- 8 PB extremely high performance storage
- +50GB/s aggregated bandwidth



Tak til

Journal og registerdata analyse

Anders Böck Jensen, CPR/KU
Isa Kirk, CPR/KU
Christian Simon, CPR/KU
Karina Basanik, CPR/KU
Mette Beck, CPR/KU
Peter Bjødstrup Jensen, CPR/KU
Robert Eriksson, CPR/KU
Isabella Friis Jørgensen , CPR
Pope Mosely, U. New Mexico
Tudor Oprea, U. New Mexico
Lars Juhl Jensen, CPR/KU
Francisco Simões Roque, CBS/DTU
Eva Roitmann, CBS/DTU

Danske Regioner

Karen Søeby, Hvidovre Hospital
Søren Bredkjær, Region Zealand
Mahad Huniche, Region Zealand
Stig Ejdrup, Region Zealand, Roskilde Hospital

Mental Centre Sct. Hans

Thomas Werge
Henriette Schmock
Thomas Hansen, Sct. Hans Hospital

Rigshospitalet

Anders Perner, Intensive care
Jan Bonde, Intensive care

Innovation Fund Denmark

RESEARCH, TECHNOLOGY & GROWTH



SEVENTH FRAMEWORK
PROGRAMME

Danish Agency for Science
Technology and Innovation
Ministry of Science
Technology and Innovation

VILLUM FONDEN

innovative
medicines
initiative



novo nordisk fonden