The Novo Nordisk Foundation
Center for Protein Research

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Disease Trajectory Analysis of Health Records and Registries

Søren Brunak
Center for Biological Sequence Analysis
Technical University of Denmark
brunak@cbs.dtu.dk
&
Novo Nordisk Foundation Center for Protein Research
University of Copenhagen
www.cpr.ku.dk

Disease

– journey versus destination

Cancer
Diabetes
Obesity
Mental disorders

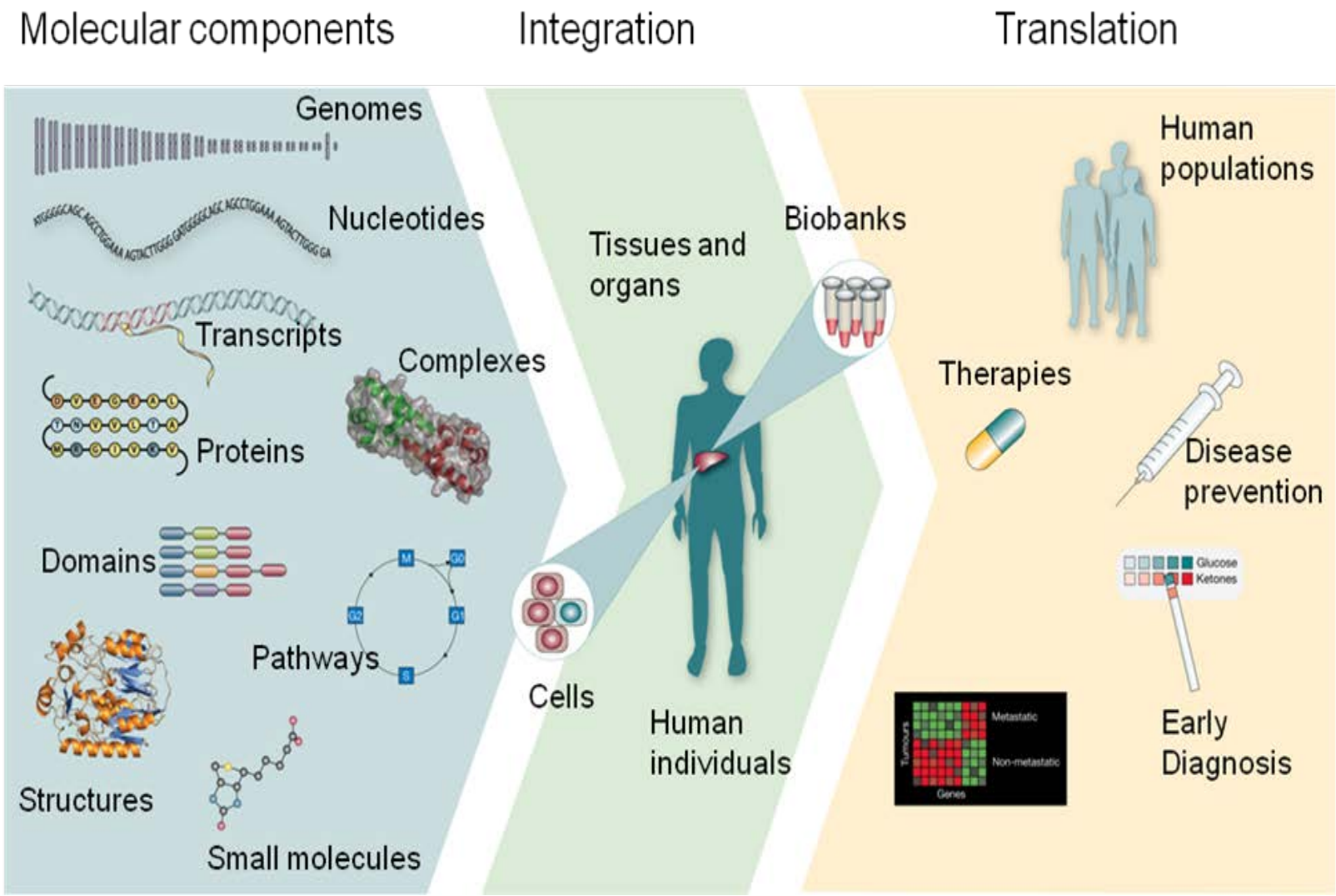
Beyond single disease analysis
(a la GWAS)

Disease-disease correlations

Disease-trajectories

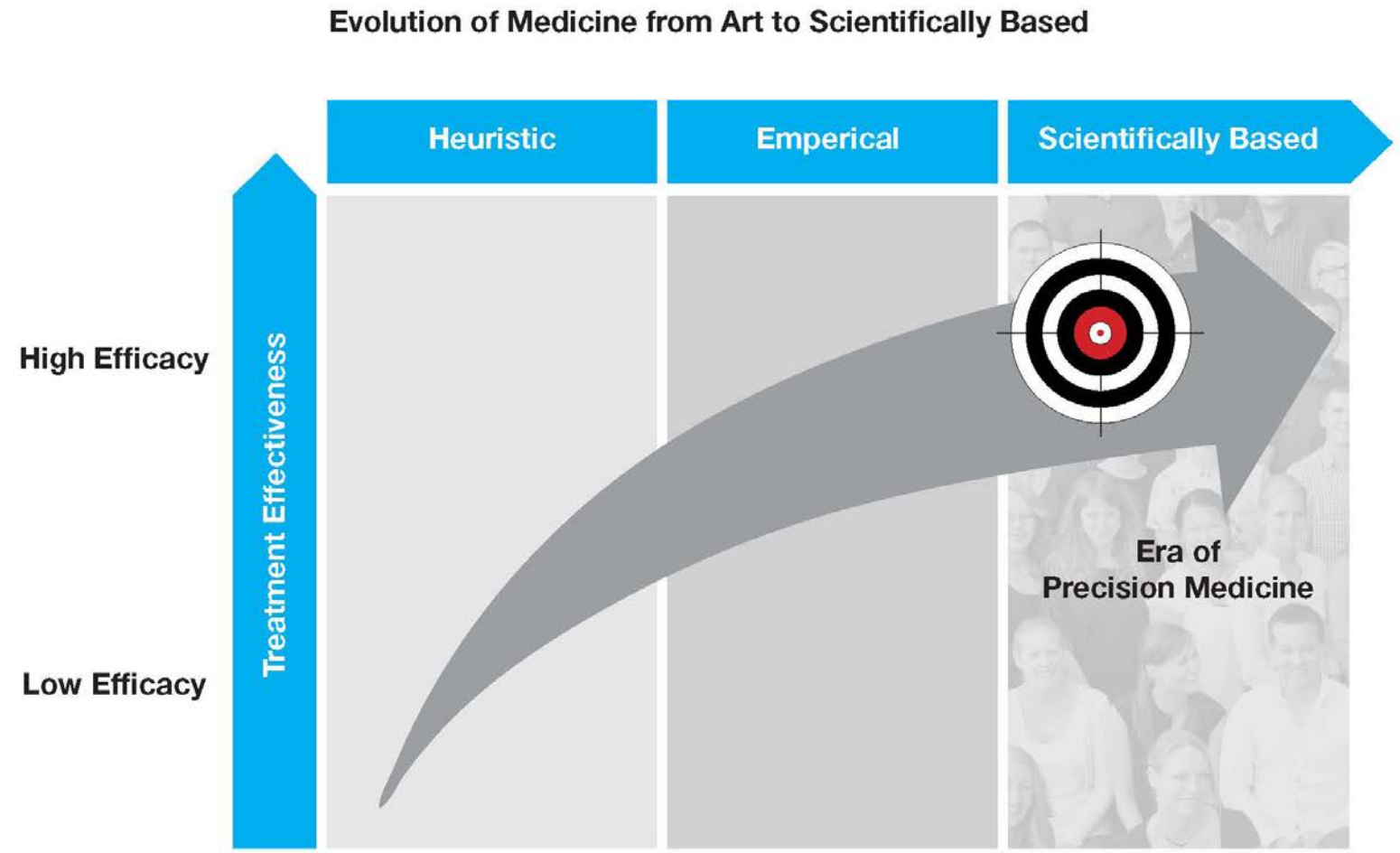
What is potentially solely genetic and what is possibly treatment related?

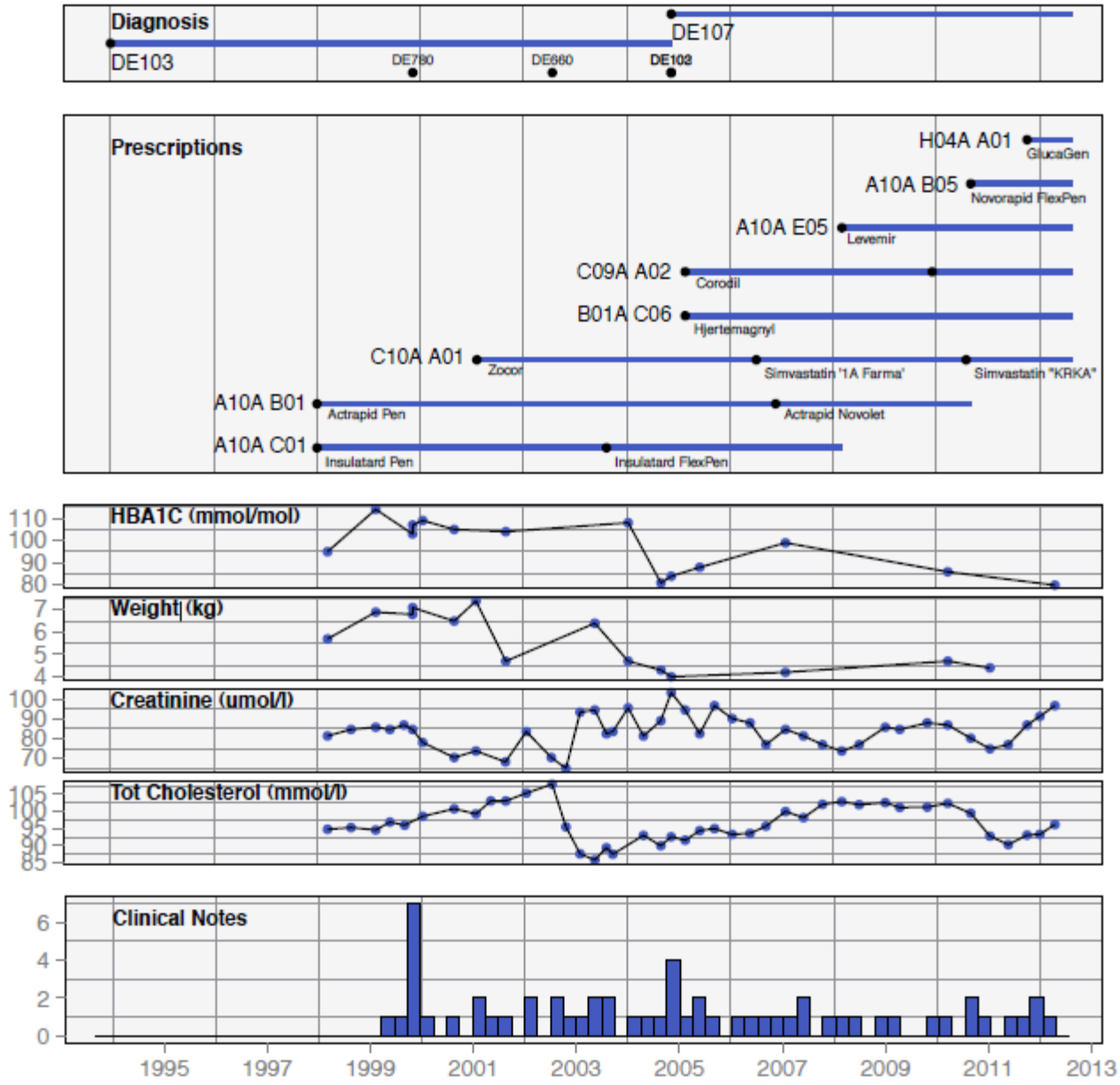
From molecules to phenotypes …
… from phenotypes to molecules

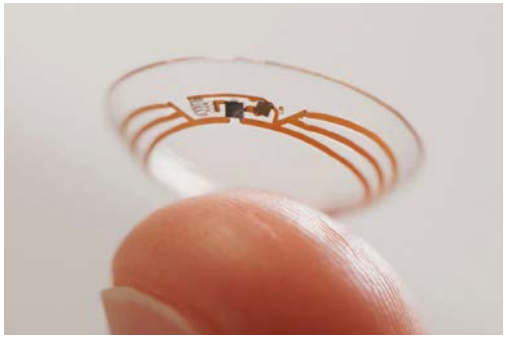


biomedbridges.eu

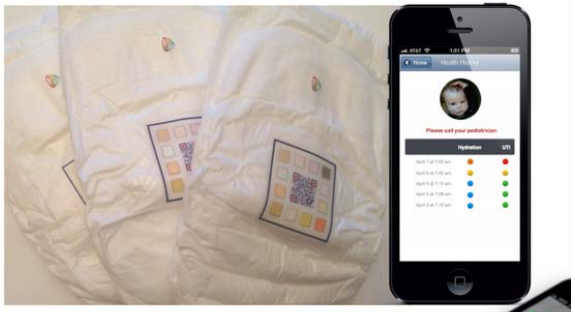
Precision medicine




Diagnosis
DE103
DE780
DE660
DE102
DE107
Prescriptions
H04A A01
GlucaGen
A10A B05
Novorapid FlexPen
A10A E05
Levemir
C09A A02
Corodil
B01A C06
Hjertemagnyl
C10A A01
Zocor
Simvastatin '1A Farma'
Simvastatin "KRKA"
A10A B01
Actrapid Pen
Actrapid Novolet
A10A C01
Insulatard Pen
Insulatard FlexPen
HBA1C (mmol/mol)
110
100
90
80
Weight (kg)
7
6
5
4
Creatinine (umol/l)
100
90
80
70
Tot Cholesterol (mmol/l)
105
100
95
90
85
Clinical Notes
6
4
2
0
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013



360 million diapers are changed every day



None of this health information has been used

Until **Smart Diapers**

Complexity of temporal health data – we all move towards becoming ICU patients

General practitioner

Outpatients

Inpatients

Intensive care

Personal identification number enables temporal analysis

From Wikipedia, the free encyclopedia

The Danish Personal Identification number is a national identification number, which is part of the personal information stored in the Civil Registration System.

Was established in 1968.

It is a ten-digit number with the format DDMMYY-SSSS, where DDMMYY is the date of birth and SSSS is a sequence number. The first digit of the sequence number encodes the century of birth (so that centenarians are distinguished from infants), and the last digit of the sequence number is odd for males and even for females.

Any person registered as of 2 April 1968 or later in a Danish civil register, receives a personal identification number.

The civil register list only persons who:

- Are born in Denmark of a mother already registered in the civil register, *or*
- Have their birth or baptism registered in a 'Dansk Elektronisk Kirkebog (DNK)' (Danish electronic church-book), *or*
- Reside legally in Denmark for 3 months or more (non-Nordic citizens must also have a residence permit)

Danish citizens, including newborn babies, who are entitled to Danish citizenship, but are living abroad, do not receive a personal ID number, unless they move to Denmark.

What is a precise phenotype?

Home » Harvard Health Blog » Overweight and healthy: the concept of metabolically healthy obesity

Overweight and healthy: the concept of metabolically healthy obesity

POSTED SEPTEMBER 24, 2013, 4:31 PM
Patrick J. Skerrett, Executive Editor, *Harvard Health*

Carrying too many pounds is a solid signal of current or future health problems. But not for everyone. Some people who are overweight or obese mange to escape the usual hazards, at least temporarily. This weight subgroup has even earned its own moniker—metabolically healthy obesity.

Health professionals define overweight as a body-mass index (BMI) between 25.0 and 29.9, and obesity as a BMI of 30 or higher. (BMI is a measure of weight that takes height into consideration. You can **calculate your BMI here**.)

Most people who are overweight or obese show potentially unhealthy changes in metabolism. These include high blood pressure or high cholesterol, which damage

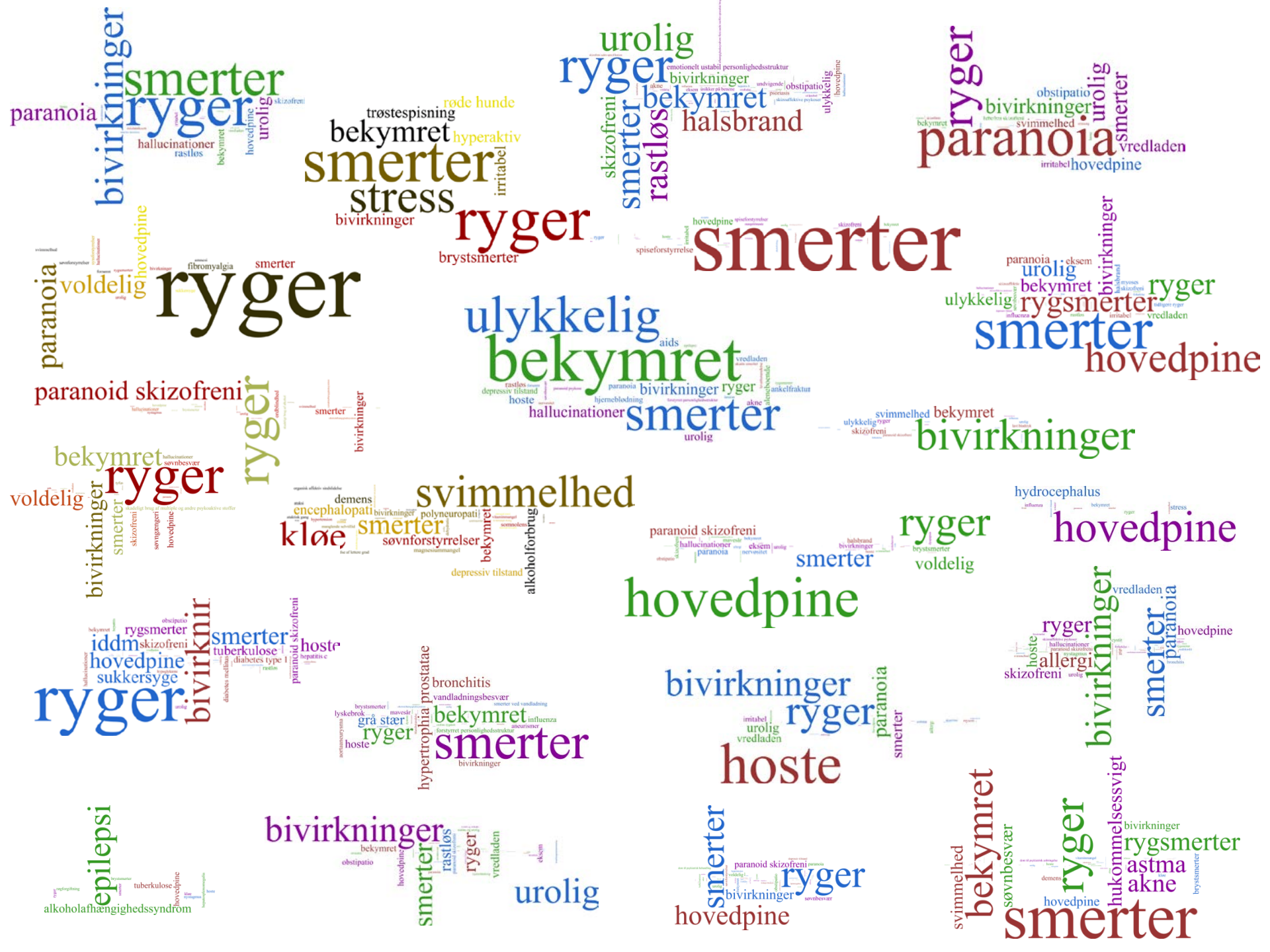
Wordfrequencies in Danish patient records

Testis cancer

Mental disorders

Male infertility

Wordfrequencies in Danish patient records from 2 x ~6,000 T1D and T2D patients (Steno Diabetes Center)

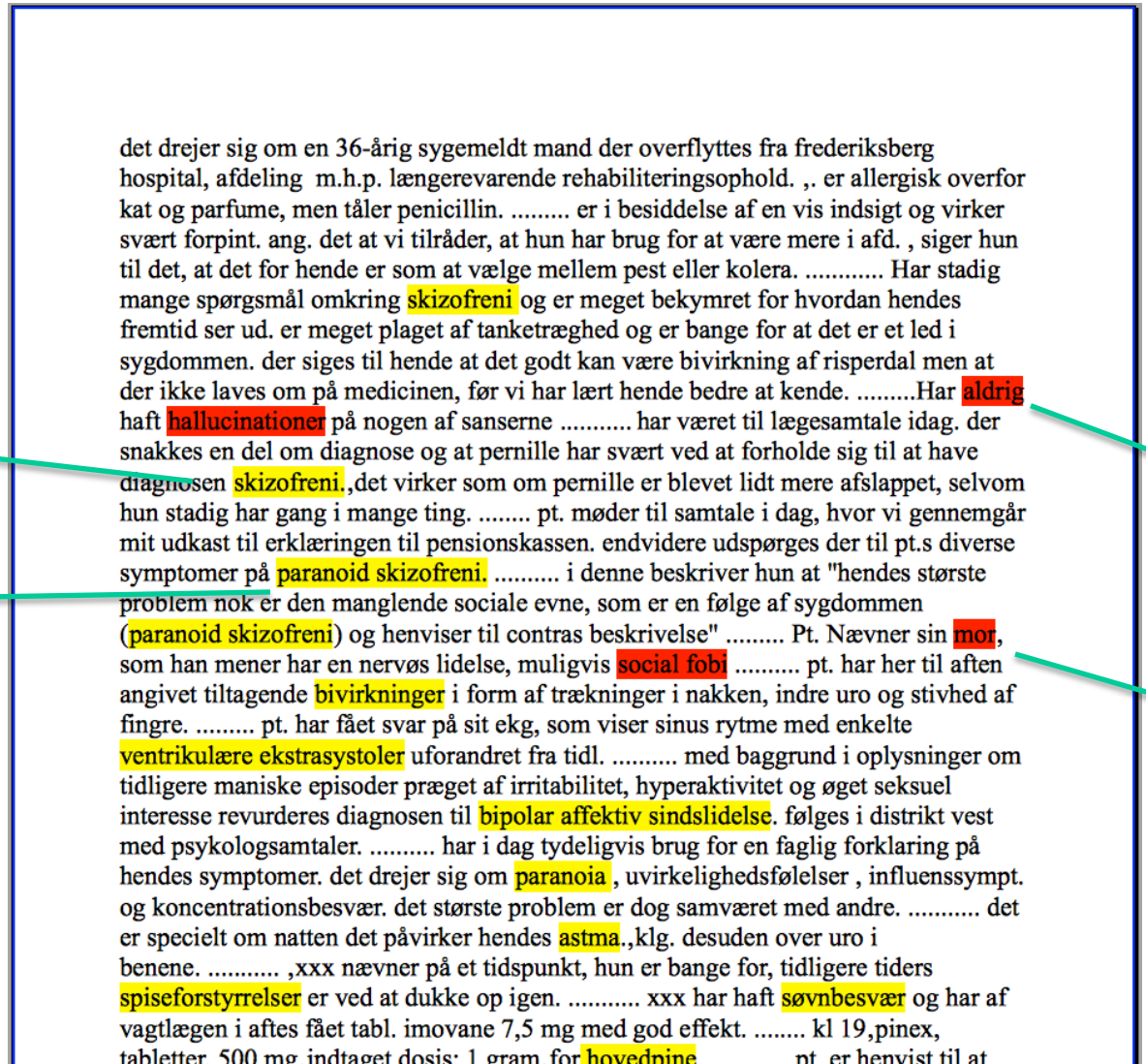

bivirkninger
smerter
ryger
paranoia
hallucinationer
rastløs
bekymret
hovedpine
urolig
skizofreni
trøstespisning
røde hunde
bekymret
hyperaktiv
smerter
irritabel
stress
bivirkninger
ryger
brystsmerter
urolig
ryger
bekymret
skizofreni
smerter
rastløs
halsbrand
obstipatio
bivirkninger
svimmelhed
paranoia
vredladen
irritabel
hovedpine
hovedpine
spiseforstyrrelse
smerter
paranoia
voldelig
hovedpine
fibromyalgia
smerter
ryger
paranoia
eksem
urolig
bekymret
bivirkninger
ulykkelig
rygsmerter
ryger
smerter
hovedpine
ulykkelig
aids
bekymret
vredladen
rastløs
depressiv tilstand
paranoia
bivirkninger
ryger
hoste
hjerneblødning
ankelfraktur
hallucinationer
smerter
urolig
paranoid skizofreni
smerter
bivirkninger
ryger
bekymret
hallucinationer
søvnbesvær
ryger
svimmelhed
bekymret
ulykkelig
skizofreni
bivirkninger
bekymret
ryger
voldelig
bivirkninger
smerter
skizofreni
søvngængeri
hovedpine
demens
encephalopati
bivirkninger
polyneuropati
svimmelhed
kløe
smerter
søvnforstyrrelser
bekymret
somnolens
depressiv tilstand
alkoholforbrug
paranoid skizofreni
hallucinationer
paranoia
eksem
nervøsitet
ryger
brystsmerter
smerter
voldelig
hydrocephalus
hovedpine
hovedpine
vredladen
bivirkninger
smerter
paranoia
hovedpine
ryger
hoste
allergi
skizofreni
rygsmerter
iddm
skizofreni
hovedpine
sukkersyge
tuberkulose
diabetes type 1
smerter
hoste
paranoid skizofreni
ryger
bivirknir
bronchitis
vandladningsbesvær
prostatae
hypertrophia
bekymret
influenza
grå stær
ryger
hoste
smerter
bivirkninger
bivirkninger
ryger
paranoia
smerter
irritabel
urolig
vredladen
hoste
epilepsi
alkoholafhængighedssyndrom
tuberkulose
hovedpine
bivirkninger
bekymret
obstipatio
hovedpine
smerter
rastløs
ryger
vredladen
eksem
urolig
smerter
paranoid skizofreni
ryger
bivirkninger
søvnbesvær
hovedpine
svimmelhed
bekymret
søvnbesvær
ryger
bivirkninger
rygsmerter
astma
akne
hukommelsessvigt
hovedpine
brystsmerter
smerter

Controlled vocabulary: ICD-10

ICD10 – International Classification of Disease

Chapter	Blocks	Title
I	A00-B99	Certain infectious and parasitic diseases
II	C00-D48	Neoplasms
III	D50-D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
IV	E00-E90	Endocrine, nutritional and metabolic diseases
V	F00-F99	Mental and behavioural disorders
VI	G00-G99	Diseases of the nervous system
VII	H00-H59	Diseases of the eye and adnexa
VIII	H60-H95	Diseases of the ear and mastoid process
IX	I00-I99	Diseases of the circulatory system
X	J00-J99	Diseases of the respiratory system
XI	K00-K93	Diseases of the digestive system
XII	L00-L99	Diseases of the skin and subcutaneous tissue
XIII	M00-M99	Diseases of the musculoskeletal system and connective tissue
XIV	N00-N99	Diseases of the genitourinary system
XV	O00-O99	Pregnancy, childbirth and the puerperium
XVI	P00-P96	Certain conditions originating in the perinatal period
XVII	Q00-Q99	Congenital malformations, deformations and chromosomal abnormalities
XVIII	R00-R99	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
XIX	S00-T98	Injury, poisoning and certain other consequences of external causes
XX	V01-Y98	External causes of morbidity and mortality
XXI	Z00-Z99	Factors influencing health status and contact with health services
XXII	U00-U99	Codes for special purposes

Mine ICD10 dictionary terms from the medical record



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 , uvirkelighedsfølelser , 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, tidligere 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 hovedpine. pt. er henvist til at

F20

F200

Negation

Family

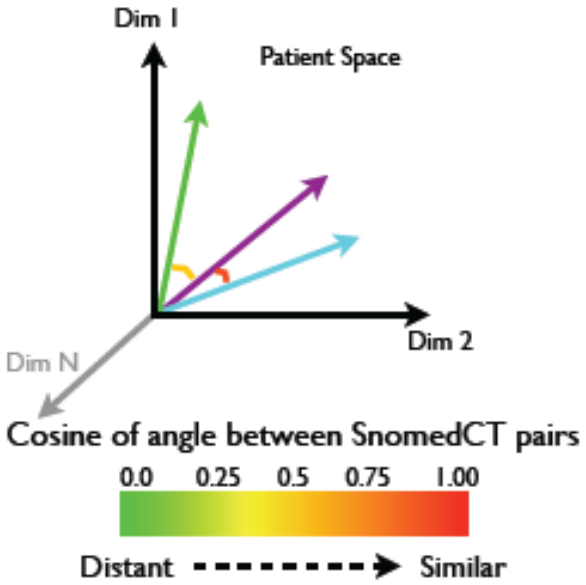
Compare patients by ICD10 terms mined and assigned in terminology space

Patient 1

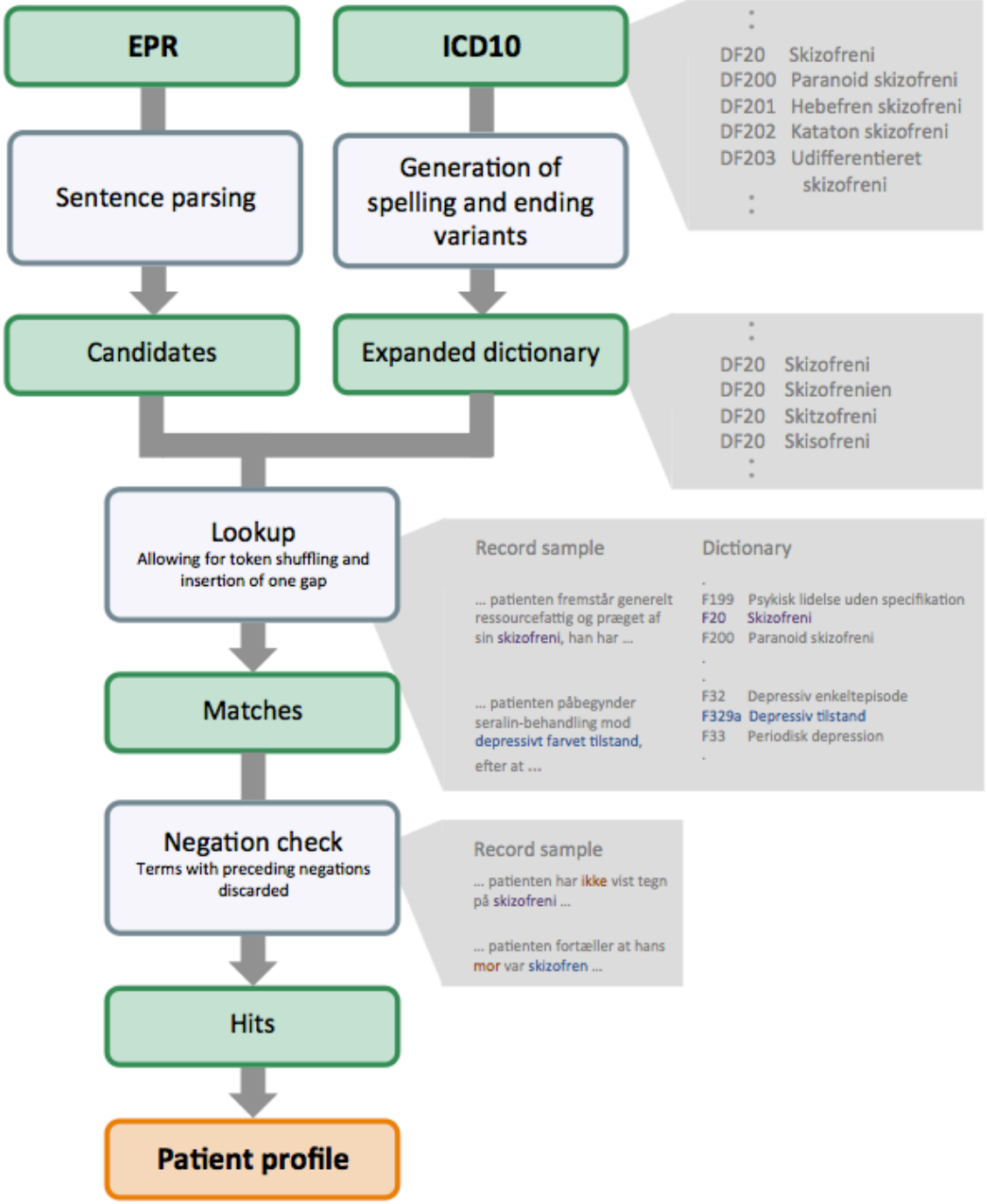
Patient 2

Patient 3

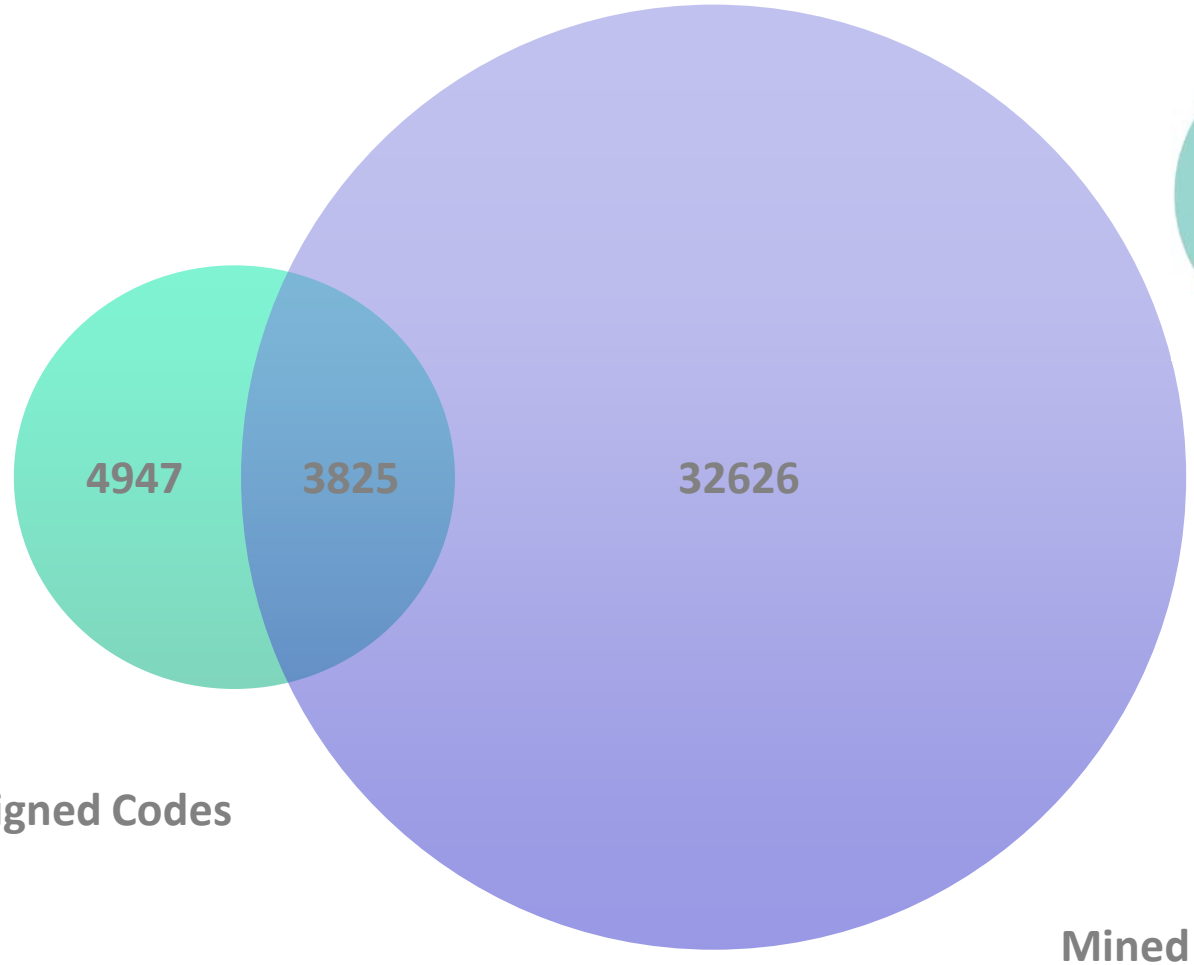
PID	248153007	88311004	386633000	...
pat1	1	0	0	
pat2	0	1	0	
pat3	0	0	1	
...				

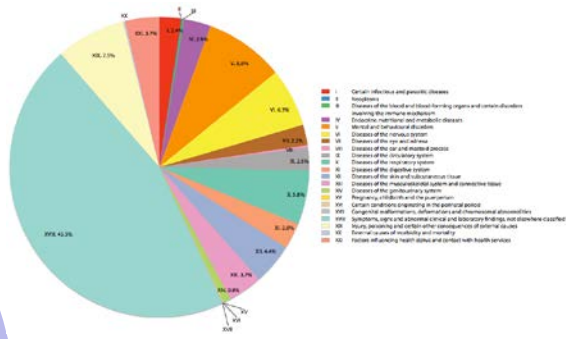


Roque et al. PLoS Comp. Biol. 2011,
Jensen et al., Nature Rev. Genet. 2012


EPR
ICD10
Sentence parsing
Generation of spelling and ending variants
DF20 Skizofreni
DF200 Paranoid skizofreni
DF201 Hebefren skizofreni
DF202 Kataton skizofreni
DF203 Udifferentieret skizofreni
Candidates
Expanded dictionary
DF20 Skizofreni
DF20 Skizofrenien
DF20 Skitzofreni
DF20 Skisofreni
Lookup
Allowing for token shuffling and insertion of one gap
Record sample
... patienten fremstår generelt ressourcefattig og præget af sin skizofreni, han har ...
... patienten påbegynder seralin-behandling mod depressivt farvet tilstand, efter at ...
Dictionary
F199 Psykisk lidelse uden specifikation
F20 Skizofreni
F200 Paranoid skizofreni
F32 Depressiv enkeltepisode
F329a Depressiv tilstand
F33 Periodisk depression
Matches
Negation check
Terms with preceding negations discarded
Record sample
... patienten har ikke vist tegn på skizofreni ...
... patienten fortæller at hans mor var skizofren ...
Hits
Patient profile

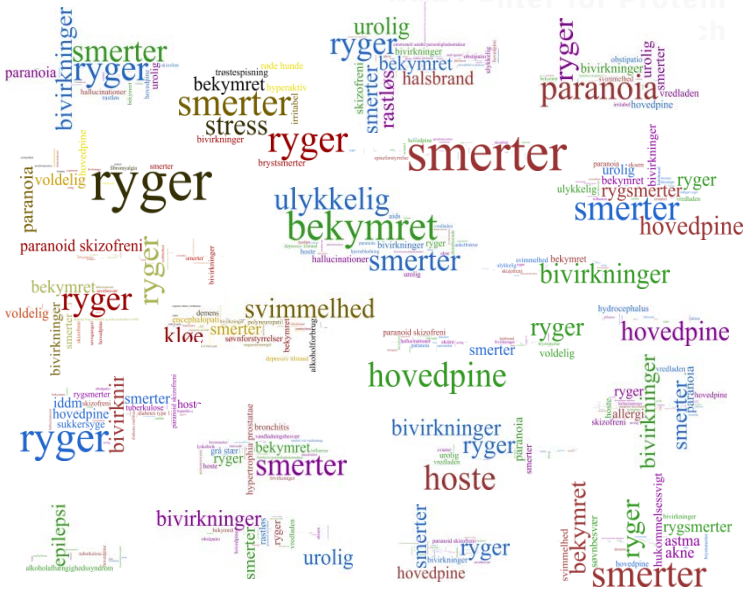
Mined and assigned ICD10 codes

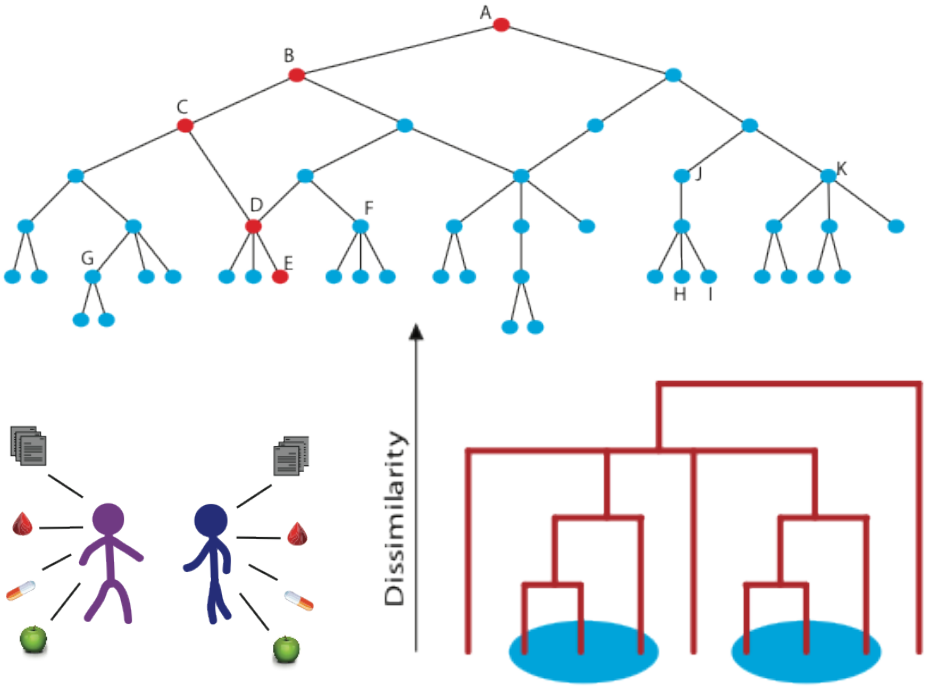




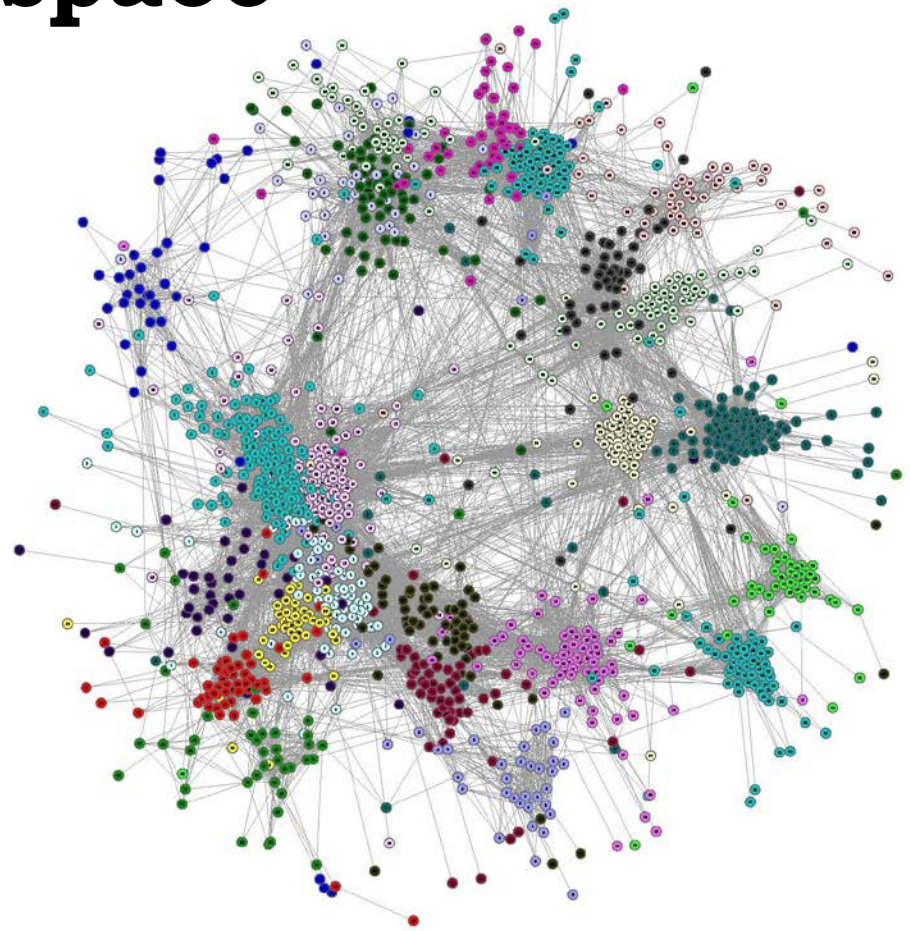
Patient similarity metrics

- **Patient phenotype similarities:**
  - Patient similarity measures beyond “words” (using knowledge from ontologies to quantify the similarity of clinical features, biochemical data etc.)
  - Semantic harmonization and phenotype harmonization, benchmarking and semantic interoperability e.g. across language barriers for meta-analysis



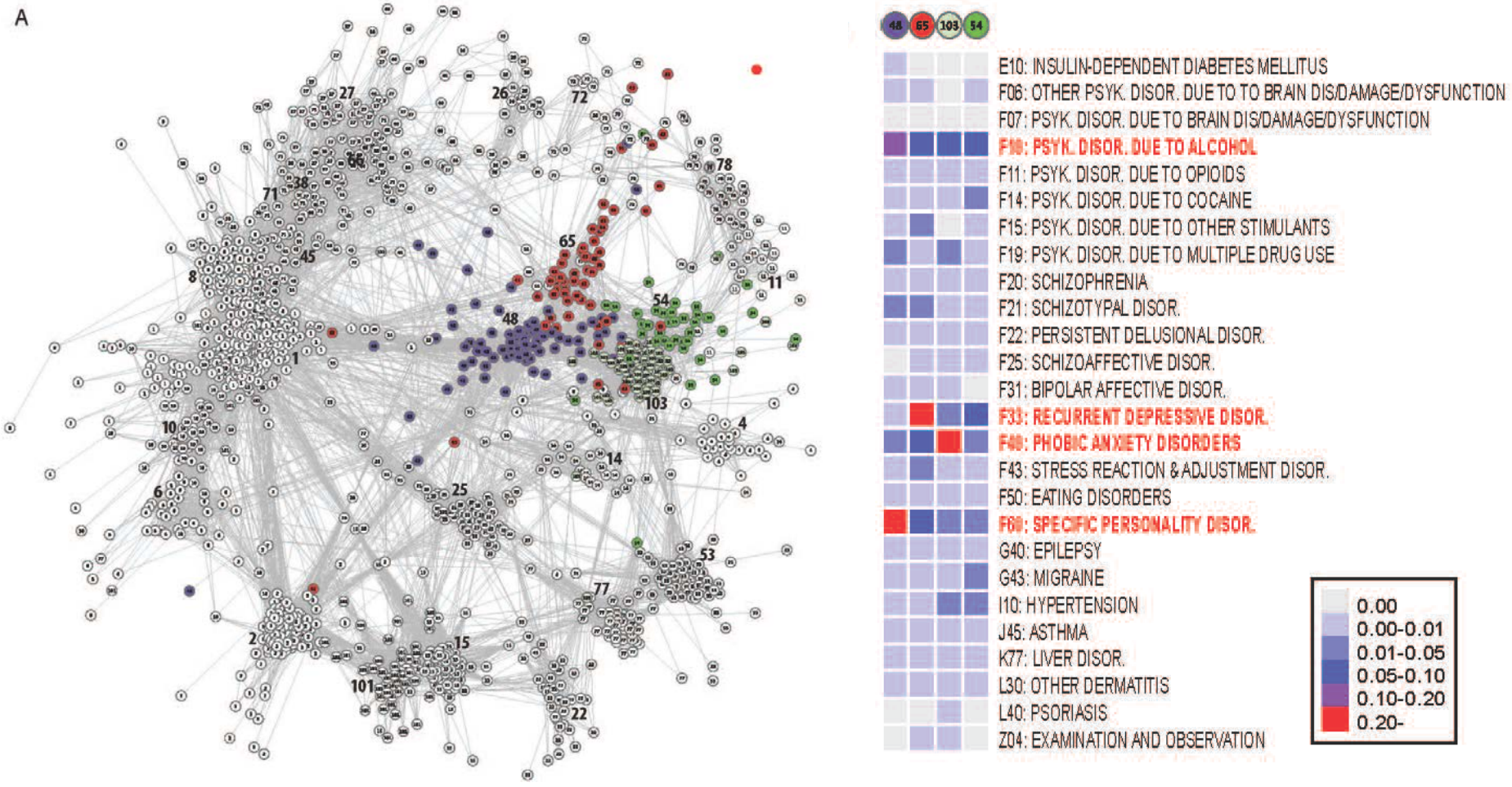


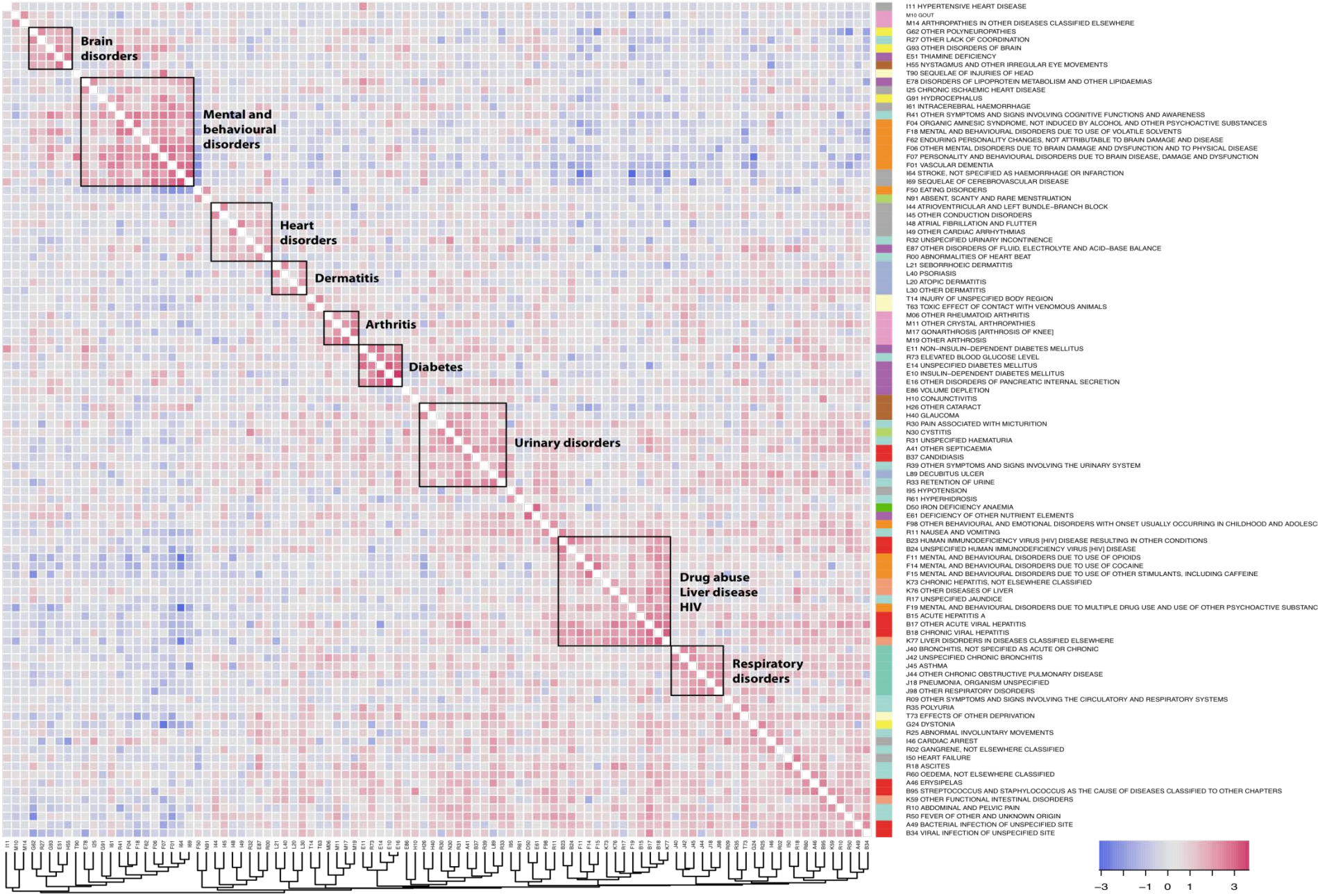
Clustering in medical terminology patient space



Roque et al.
PLoS Comp. Biol. 2011
Jensen et al.,
Nature Rev. Genetics 2012

Alcohol and depressive disorders, anxiety disorders, and other personality disorders





Roque et al. PLoS Comp. Biol. 2011,
Jensen et al., Nature Rev. Genet. 2012

Significant comorbidities among complex and Mendelian disorders (110M patients, registry data)

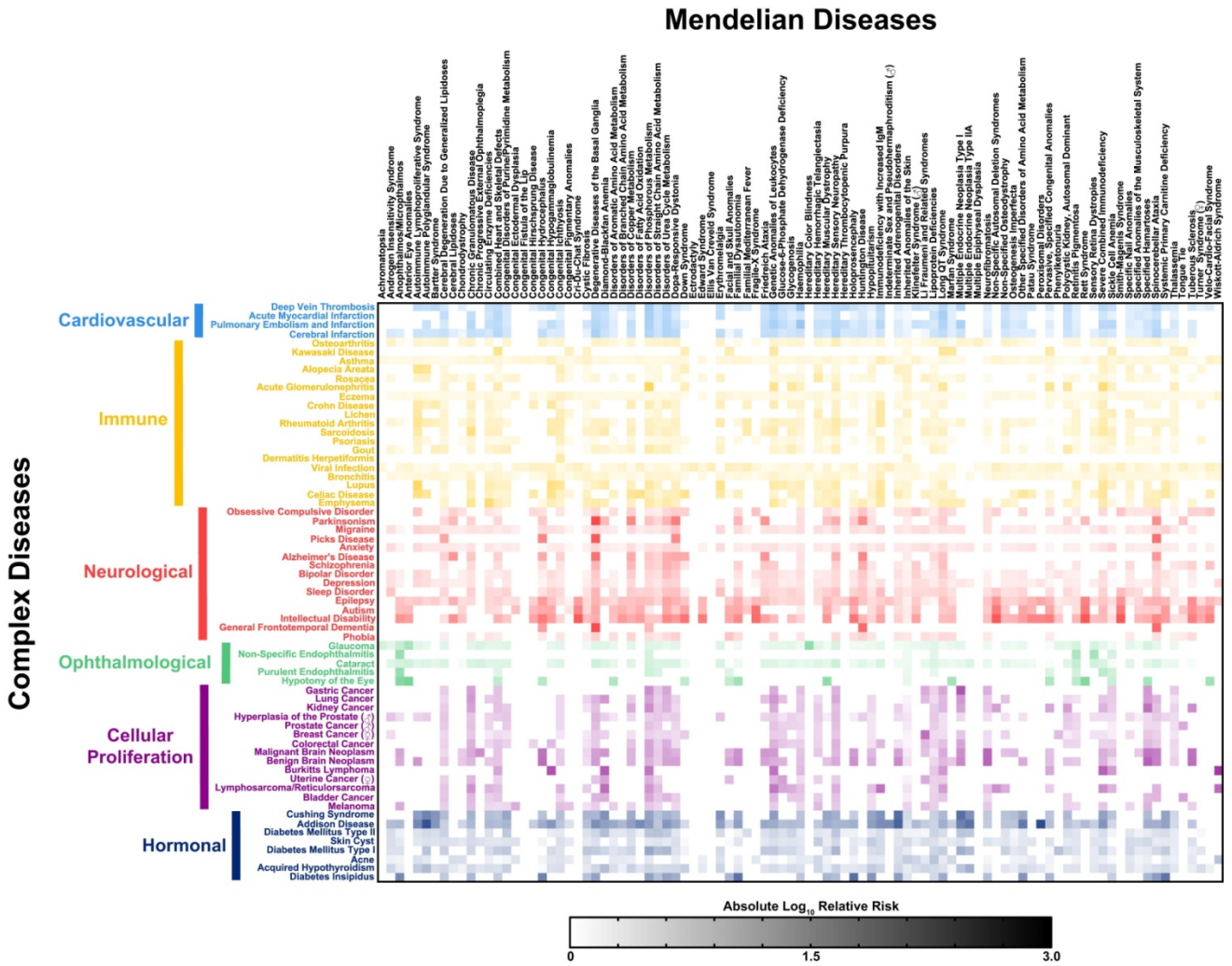


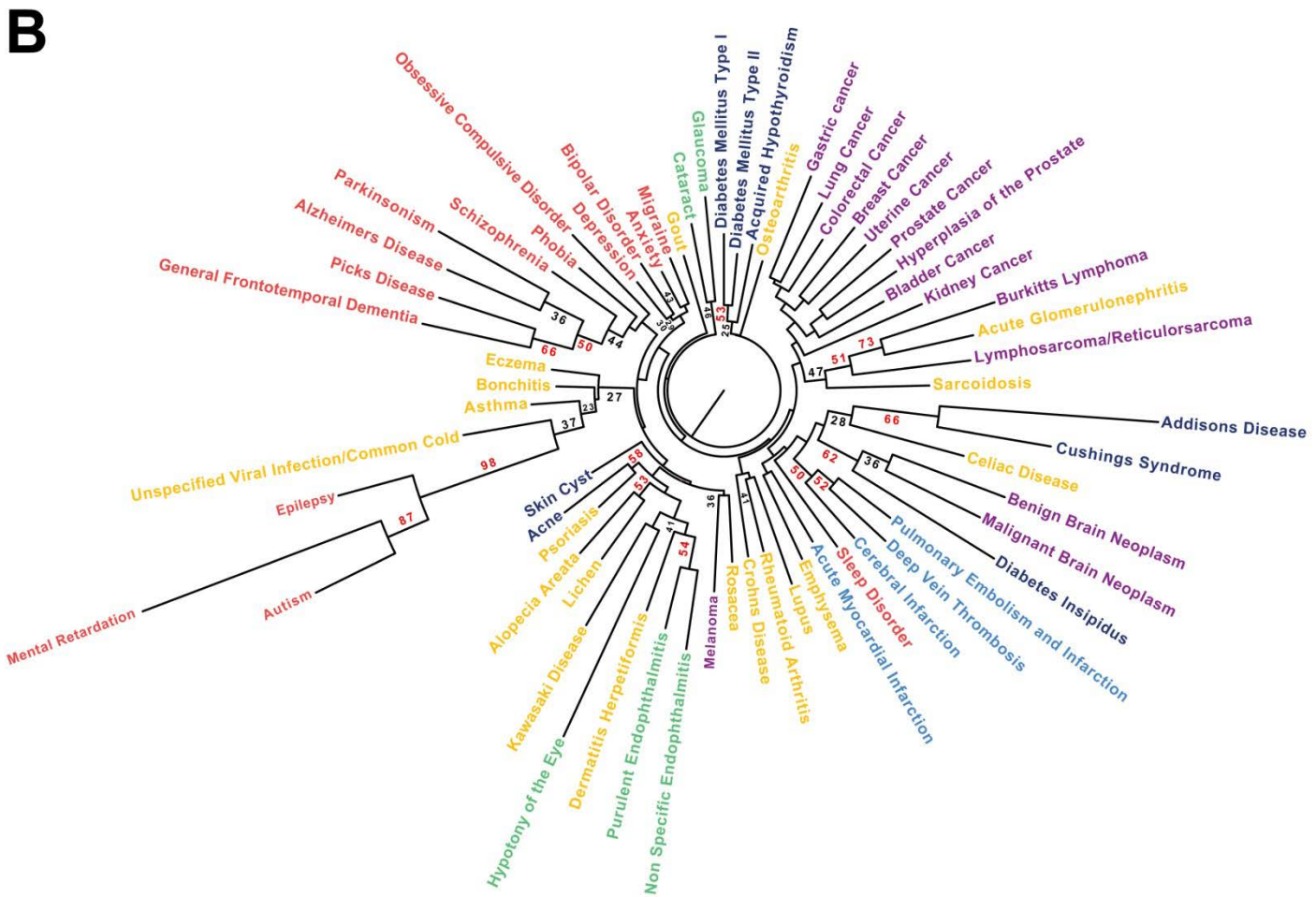
Table 1. The Clinical Record Data Sets Utilized in This Study

Data Set	Description	Encoding Type	Number of Unique Patients
CU	Columbia University, 1985–2003	ICD9	1,505,822
DK	Denmark; database covering most of the country's population	ICD10	6,214,312
NYPH	New York Presbyterian Hospital and Columbia University; 2004–present	ICD9	767,978
SU	Stanford University	ICD9	806,369
TX	University of Texas at Houston	ICD9	1,599,528
UC	University of Chicago	ICD9	146,989
USA	MarketScan insurance claims data set	ICD9	99,143,849
MED	Medicare database	ICD9	13,039,018
	Total:		123,223,865

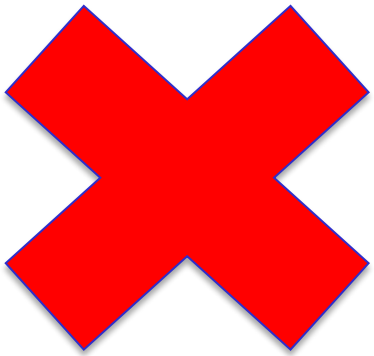
This table provides a brief description, the ICD encoding type, and the size of each data set. The MED data set was used for comparison and was not included in the full meta-analysis.

Blair et al., Cell 2013

Similarities of complex diseases computed from comorbidity profiles to Mendelian disorders

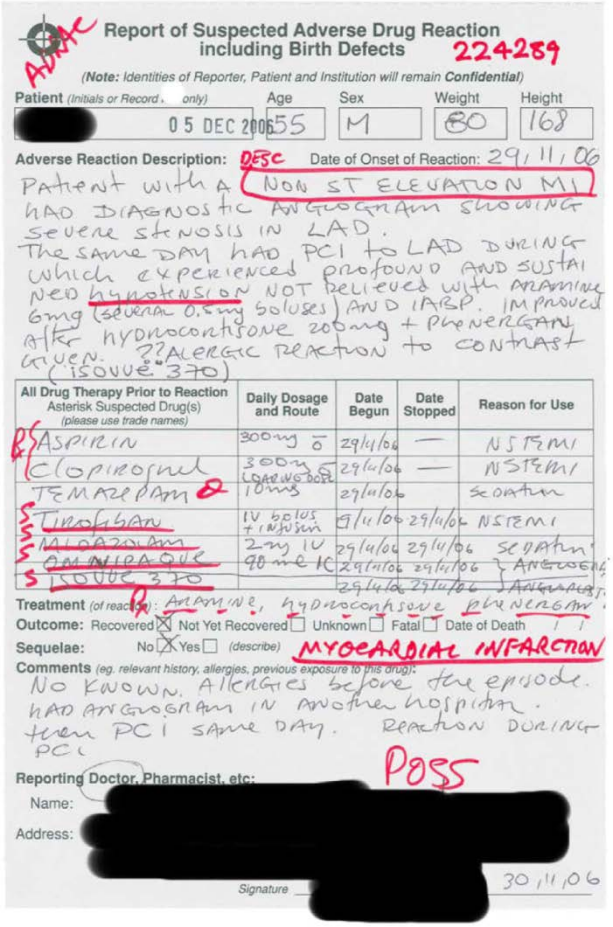


Which disease-disease and symptom correlations are treatment related?



Spontaneous reports

- **Heavily trusted**
- **Underreporting and biases**
- **Data quality issues**



Report of Suspected Adverse Drug Reaction including Birth Defects 224289

(Note: Identities of Reporter, Patient and Institution will remain Confidential)

Patient (Initials or Record only) [redacted] 05 DEC 2006 | Age 55 | Sex M | Weight 80 | Height 168

Adverse Reaction Description: DESC Patient with a NON ST ELEVATION MI had diagnostic angiogram showing severe stenosis in LAD. The same day had PCI to LAD during which experienced profound and sustained hypotension not believed with aramine 6mg (several 0.5mg boluses) and IABP. Improved after hydrocortisone 200mg + phenergan given. ??Allergic reaction to contrast (isovue 370)

Date of Onset of Reaction: 29/11/06

All Drug Therapy Prior to Reaction Asterisk Suspected Drug(s) (please use trade names)	Daily Dosage and Route	Date Begun	Date Stopped	Reason for Use
PS ASPIRIN	300mg O	29/11/06	—	NSTEMI
clopidogrel	300mg LOADING DOSE	29/11/06	—	NSTEMI
TEMAZEPAM O	10mg	29/11/06		SEDATION
S tirofiban	IV BOLUS + INFUSION	29/11/06	29/11/06	NSTEMI
S MIDAZOLAM	2mg IV	29/11/06	29/11/06	SEDATION
S OMNIPAQUE	90 ml IC	29/11/06	29/11/06	ANGIOGRAM
S ISOVUE 370		29/11/06	29/11/06	ANGIOGRAM

Treatment (of reaction): aramine, hydrocortisone phenergan

Outcome: Recovered ☒ Not Yet Recovered ☐ Unknown ☐ Fatal ☐ Date of Death / /

Sequelae: No ☒ Yes ☐ (describe) MYOCARDIAL INFARCTION

Comments (eg. relevant history, allergies, previous exposure to this drug): No known allergies before the episode. had angiogram in another hospital. then PCI same day. Reaction during PCI

Reporting Doctor, Pharmacist, etc: POSS

Name: [redacted]

Address: [redacted]

Signature [redacted] 30/11/06

Summaries of Product Characteristics (SPCs)

Every medicine pack includes a patient information leaflet (PIL).

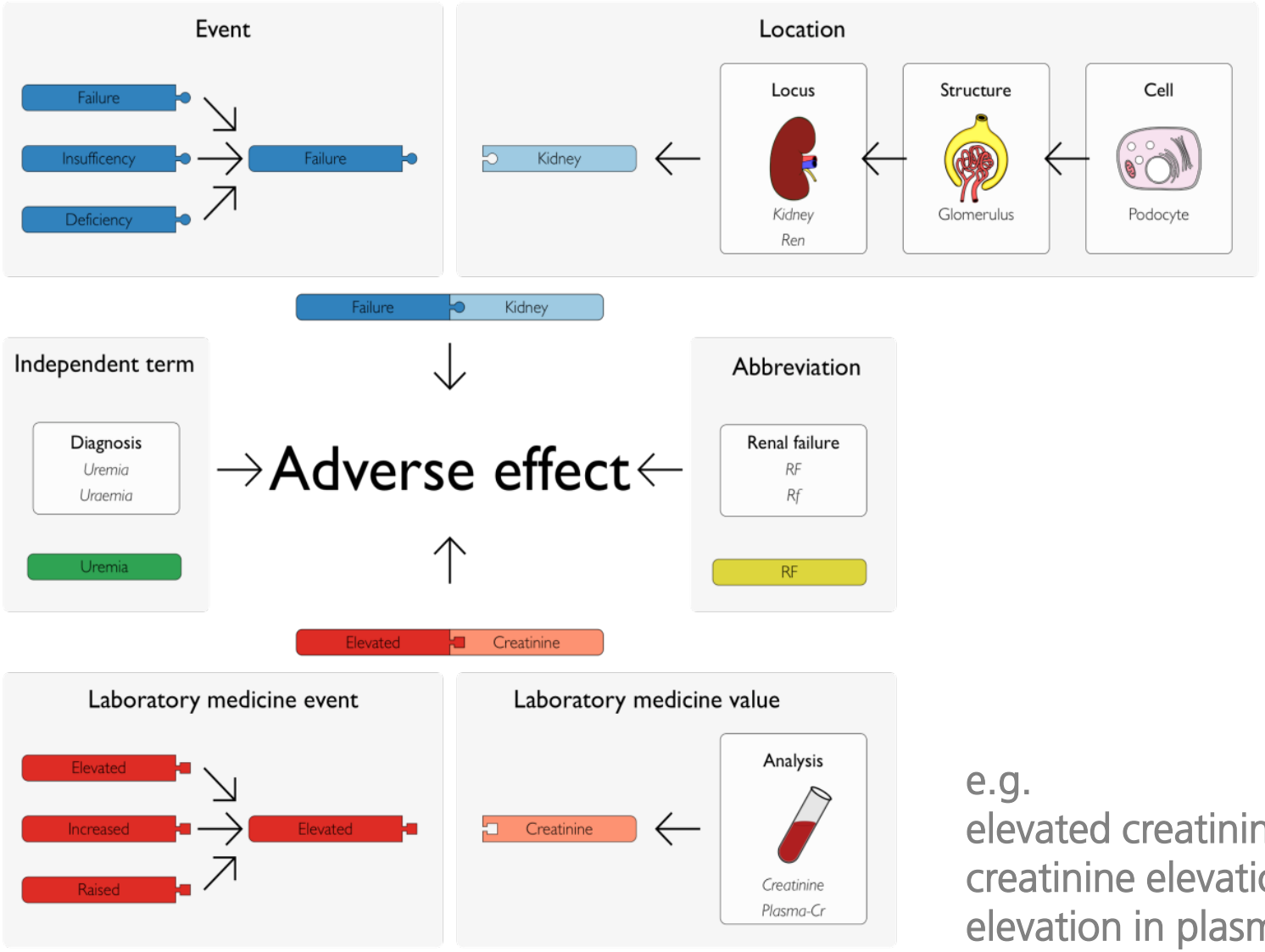
PILs are based on the Summaries of Product Characteristics (SPCs) which describe a medicinal product's properties and the conditions attached to its use.

Controlled ADR/ADE vocabulary from SPCs

ADR dictionary characteristics

- Descriptions extracted from **7,446 drug SPCs.** Including drugs approved by both European Medicines Agency and Danish Medicines Agency.
- **21,342 uniquely spelled ADRs** was used to construct the dictionary.
- Current version can match about 8,000 different ADRs.
- Final dictionary can match > **$4 .10^{12}$** ways to describe ADRs before applying fuzzy matching.

Adverse effect concept linkage



e.g.
elevated creatinine,
creatinine elevation
elevation in plasma-Cr
are recorded identically

Text mining of drug names, ADE/ADRs, diagnoses, …

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 tankeforstyrrelser, 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 kroppen og "ostek-lokkefornemmelse". Herefter Orap 2 mg fra 2.8. stigende til 3 mg fra 30.8.99 med god effekt på tankeekko og tankemylder. Behandlet med Zoloft 50 mg fra maj 98 til maj 99 med noget virkning på depressive symptomer, men seponeret på grund af natlig svedtendens. Siden 14.7.99 Efexor 75 mg med nogen effekt på antallet og sværhedsgraden af kortvarige depressive episoder

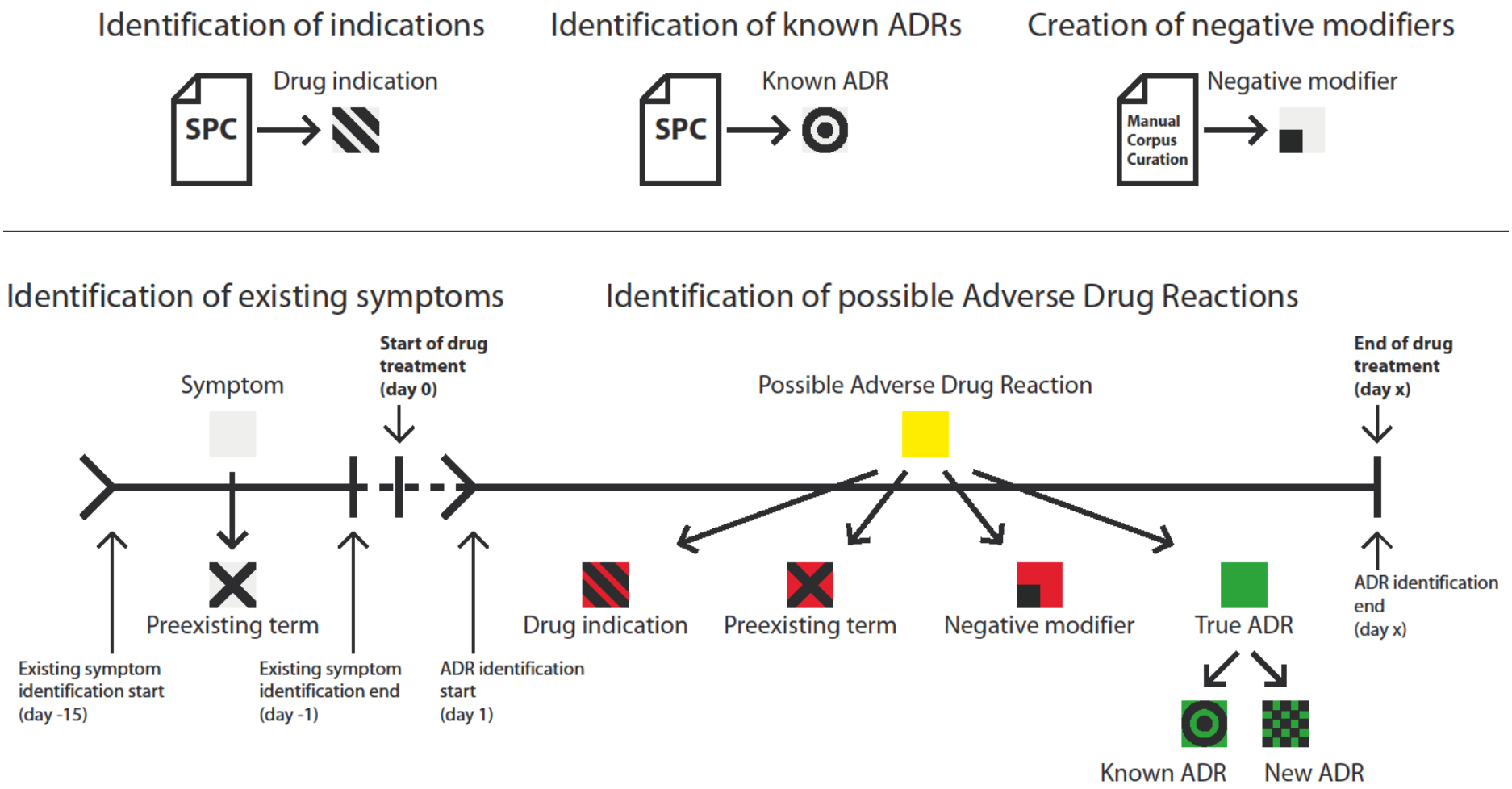
se venligst under allergier. Desuden forsøgt beh med Zyprexa. sep grundet vægtøgning, træthed og manglende effekt. Risperdal ord med nogen effekt tillagt dogmatil(1999). Efterfølgende aurorix beh seponeret i 1999. Startede istedet remeron. Aktuel medicindois, jvf ud-skrivningsnotat fra U12 samt EPJmedicinliste.

tbl. leponex 100+0+0+200 mg,tbl. rivotril 0,5 +0+1+ 0 mg,tbl. arintapin 0+0+0+30 mg,tbl. klomipramin 0+0+0+25 mg,tbl. imoclone 7,5 mg nocte,tbl. rivotril 2 mg p.n. max x 1 dgl. tbl. marevan a 2,5 mg efter skema,tbl. magnesia 1 g p.n. laxoberaldr. 7,5 mg /ml 15 dr. p.n. mix. link 150 mg /ml 15 ml p.n. 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 tålt godt, hvorfor der er ansøgt om ophævelser 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

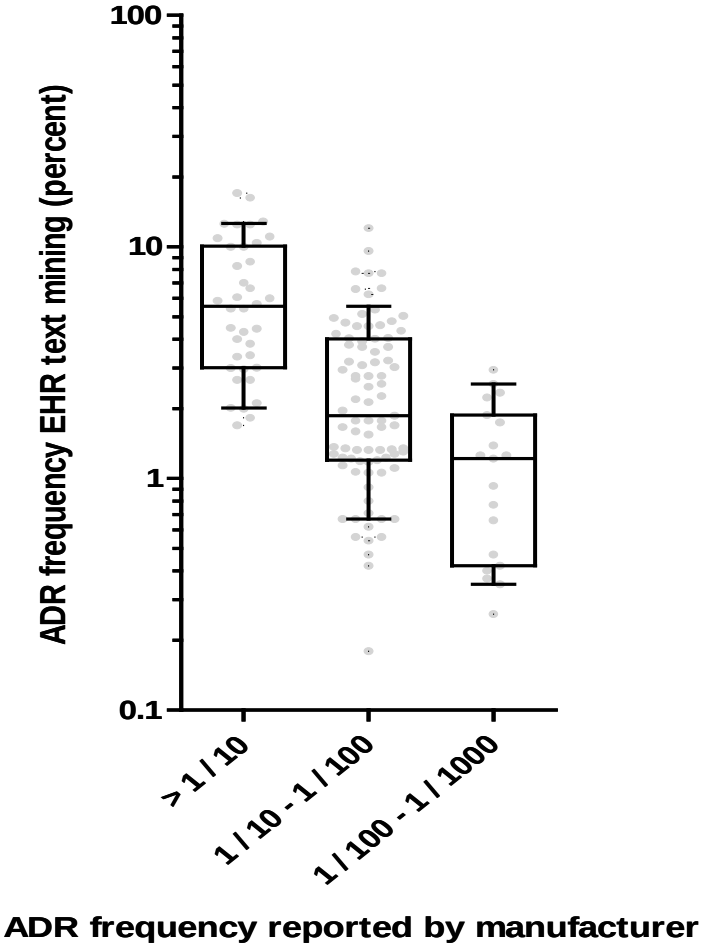
...Jeg mener fortsat, at han har brug for medicin, da han i går fx var meget vred og følte sig utryg og angst og har haft svært ved at sove. Dette synes pt. at accepterer. Jeg tilbyder herefter Zyprexa i stedet for Risperdal, pt. afviser dette, da han ved medpatient har fået denne medicin og har fået øget appetit, dette vil han ikke. Har ikke tidligere fået antipsykotisk medicin. Accepterer herefter Cisordinol, startende på en lille dosis. Accepterer også angstdæmpende medicin i dagtiden. Angiver, at når han bliver vred kan han godt styre det. Synes det hjalp i går noget at få Nozinan. Virker fortsat garderet. Siger intet uopfordret. Sparsomt sygdomsindsigt. Er ikke vredladen. Er i dag er heller latent aggressiv...

Text mining Adverse Drug Reactions
(using 7,500 drug names and 21,000 ADRs)



Eriksson et al. Drug Safety 2014

Frequency of Adverse Drug Reactions from text mining



Comparison between the 150 most statistically significant extracted ADR frequencies and the stated frequencies in the SPC by the manufacturer. Dots are showing the individual extracted frequencies.

Out of the 150 are 6 left out, according to the SPC these are having a frequency < 1/1000 and the corpus is not large enough to detect these with a satisfying frequency.

Recall of 75,1% and a precision of 95,0%. (manual curation of 200 records)

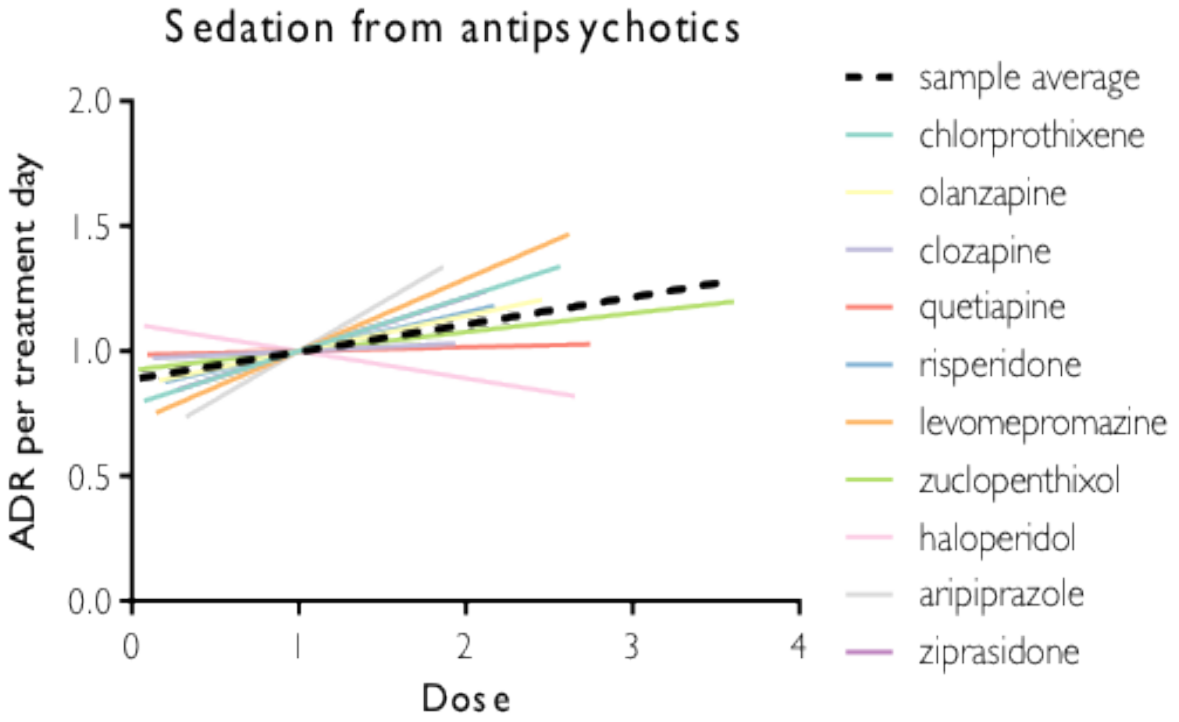
Eriksson et al. Drug Safety 2014

ADR-dose dependencies

Dosages from structured medication data

Sedation is the most occurring ADR in the corpus

Drugs selected are 10 antipsycotics (a class known to cause sedation)



ADRs and doses are normalized on multiples of the minimum dose prescribed of each drug.

Plot for 21 days steady dosage data is visualized, sample average slope 0.1105 (95% CI, 0.03085-0.1901), non-zero slope p-value was 0,0074, all individual drug slopes are positive except for haloperidol.

Eriksson et al. Drug Safety 2014

Possible ADRs?

Drug substance	ADE	p-value
Dipyridamole	Visual impairment	4.375e-04
Simvastatin	Personality changes	8.408e-08
Citalopram	Psychosis	8.807e-04
Bendroflumethiazide	Apoplexy	8.46e-03
Chlordiazepoxide	Nystagmus	4.03e-08

X1	Y1	p1
X2	Y2	p2
....		

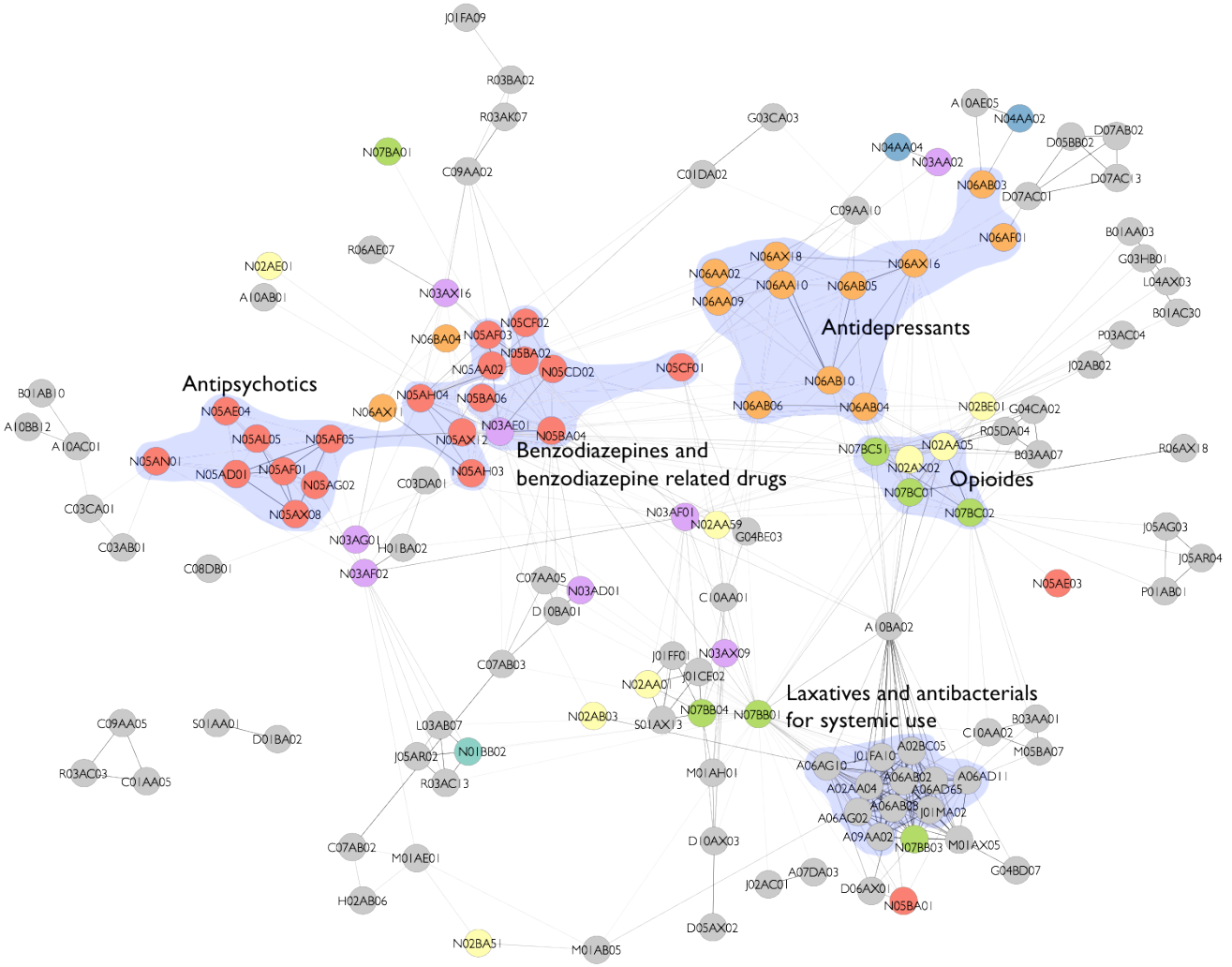
p-values are multiple testing corrected

Drug-ADR similarities

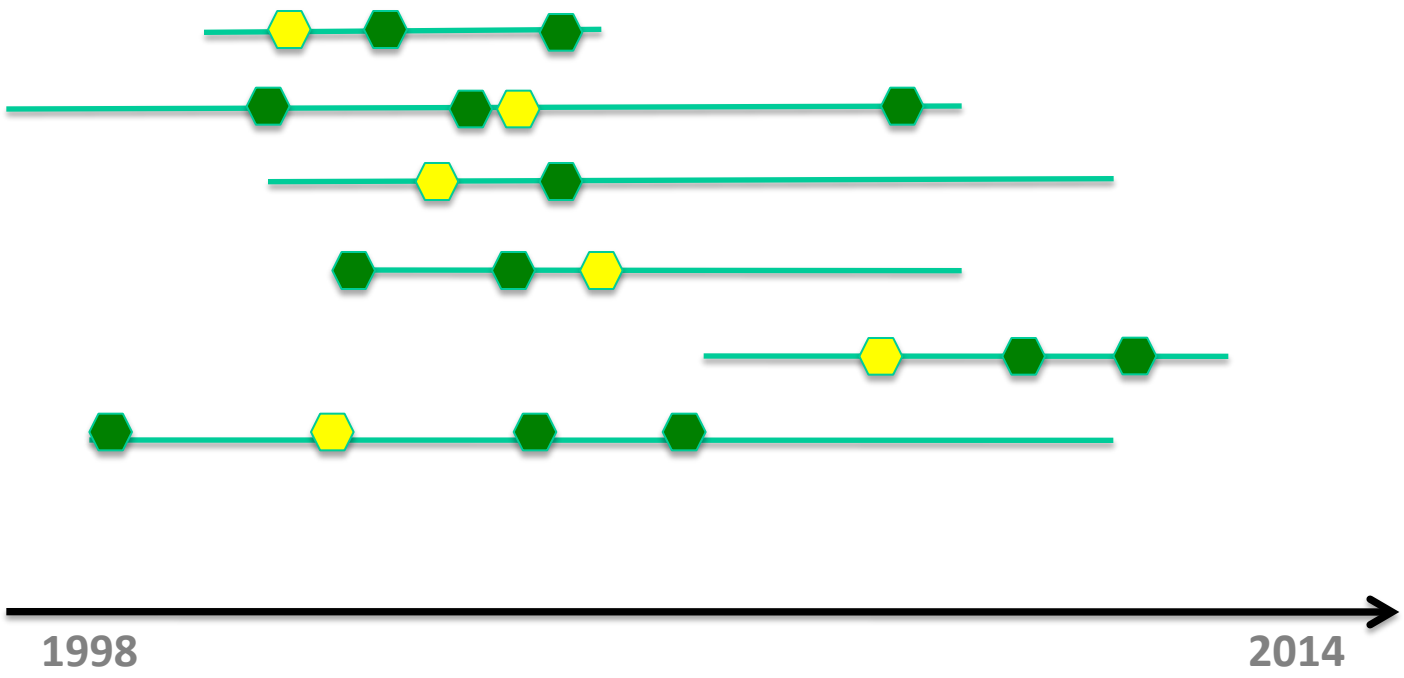
(ADRs text mined by temporal analysis of EHRs)

- N01 Anesthetics
- N02 Analgesics
- N03 Antiepileptics
- N04 Anti-parkinson drugs
- N05 Psycholeptics
- N06 Psychoanaleptics
- N07 Other nervous system drugs
- All other ATC codes

13 years of drug use at Mental centre Sct. Hans ATC code coloring. Edges show Drug ADR profile similarity, darker edge indicates stronger similarity. Network contains 500 strongest edges (Jaccard index). "N03 Antiepileptics" are scattered, not showing any clustering. This is expected as antiepileptics is a very diverse drug classein terms of ADRs. Laxatives and antibacterials for systemic use group since both cause diarrhea and stomach ache and other gastrointestinal problems



Comorbidity trajectories in individual patients



Typical development
E.g. type 2 diabetes > problems with foot > amputation

Yellow dot: Debut of a given ICD10 code, green debut of other diseases
Danish Discharge Registry, 6.2 million patients

Danish Patient Registry

Following up in registry data

15 years of ICD10 hospitalization history for 6.2 million Danes

- 68 million records (in/out of ward)
- 45 million admissions (in/out of hospital)
- 119 million diagnosis-record-associations
- 18 million surgical procedures
- 131 million treatments & examinations

Non-hypothesis driven comorbidity analysis of registry data

Patient admission to hospital

▽

Symptoms and clinical findings

▽

Critical diagnoses written in patient journal

▽

Diagnosis verified/declined

▽

Procedures (operations and treatments)

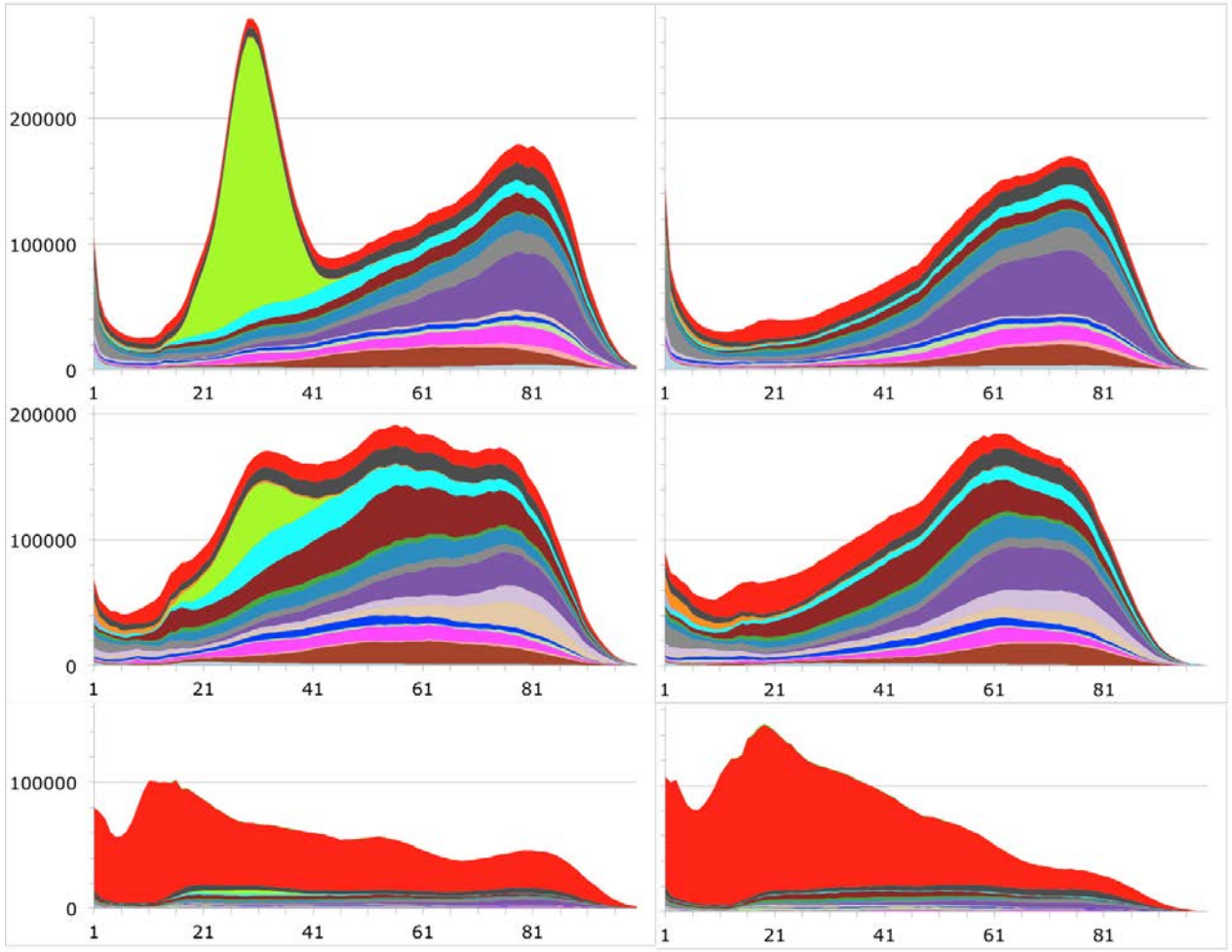
▽

Discharging physician makes a discharge summary including ICD-10 codes from patient journal

▽

Entered in the hospital's registry by a medical secratary ICD-10 codes sent to National Dishcarge registry

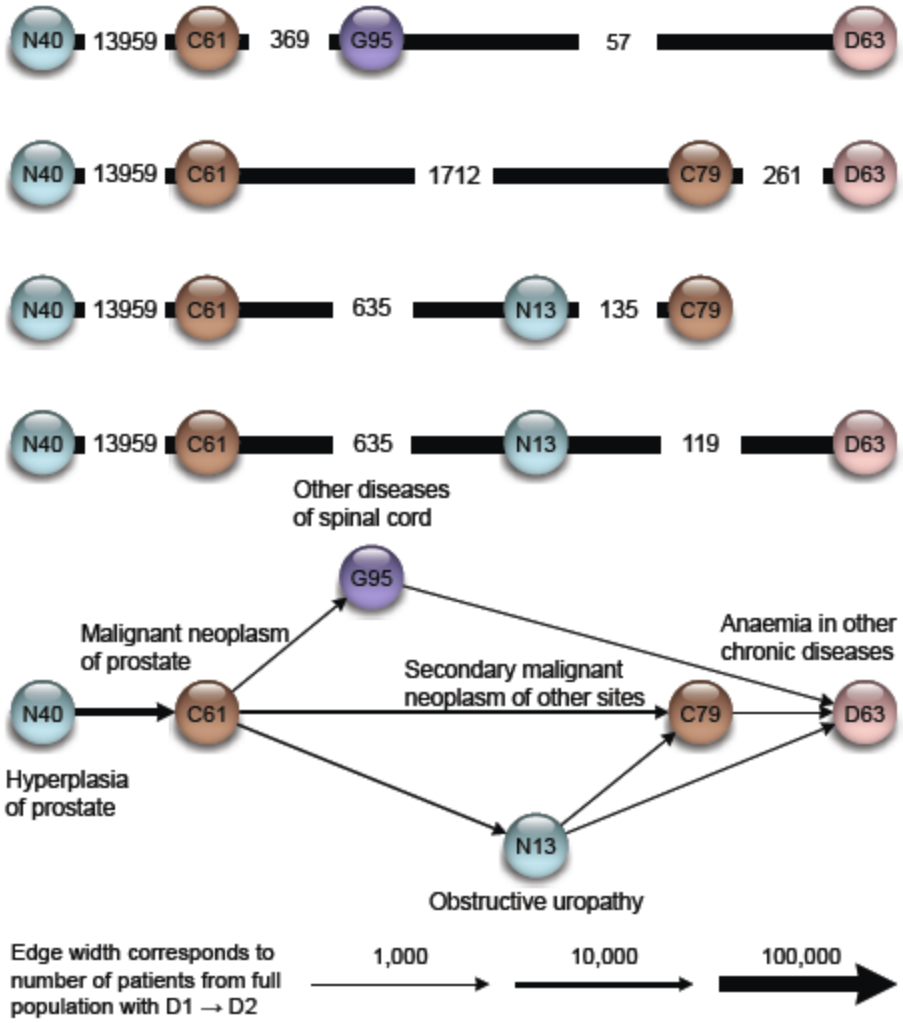
National Patient Registry (6.2M Danes) ICD10 diagnoses as a function of age



ICD 10 chapter coloring

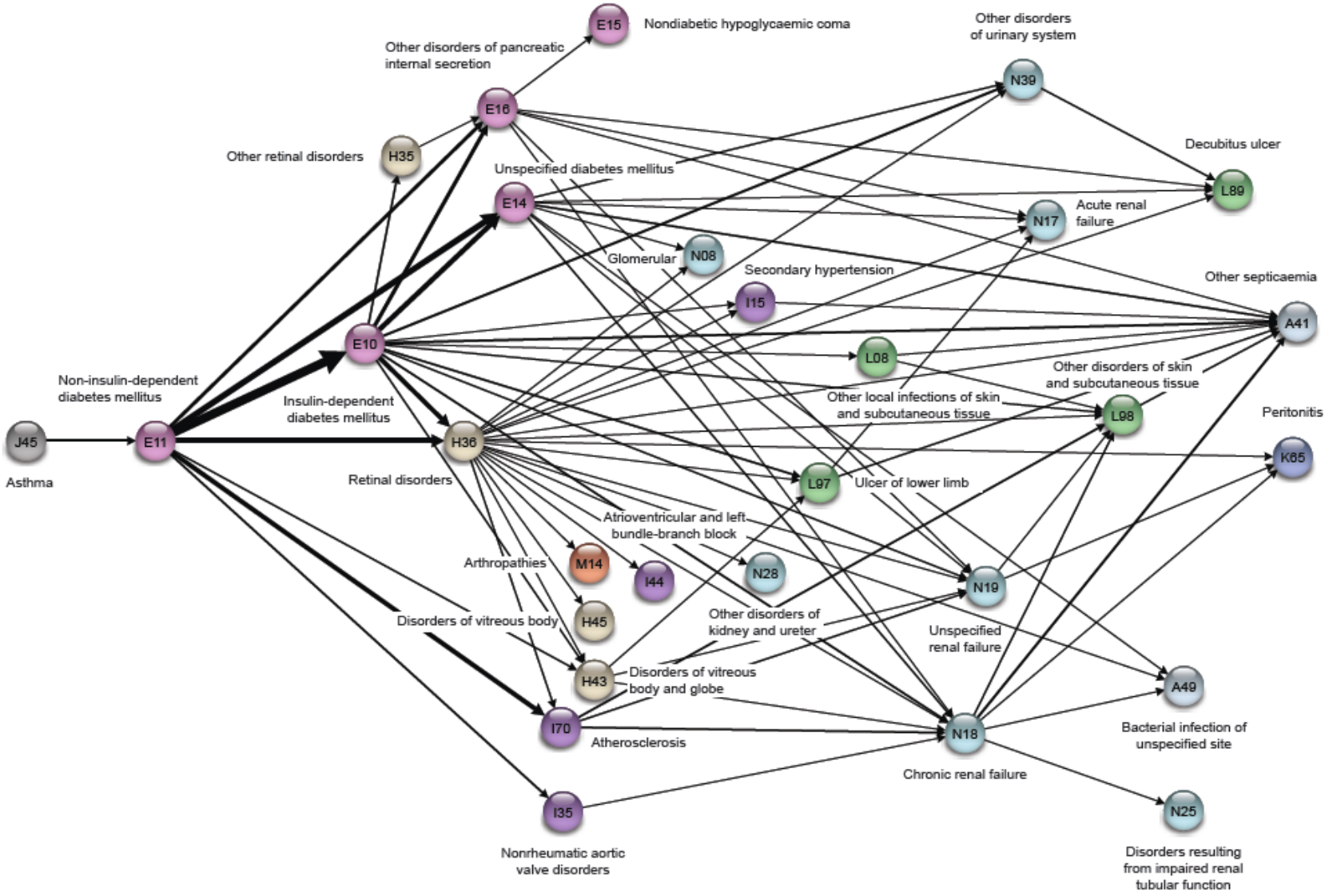
1: Certain infectious and parasitic diseases
2: Neoplasms
3: Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
4: Endocrine, nutritional and metabolic diseases
5: Mental and behavioural disorders
6: Diseases of the nervous system
7: Diseases of the eye 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 aboratory findings, not elsewhere classified
19: Injury, poisoning and certain other consequences of external causes
20: External causes of morbidity and mortality

Disease trajectories and trajectory-cluster for prostate cancer



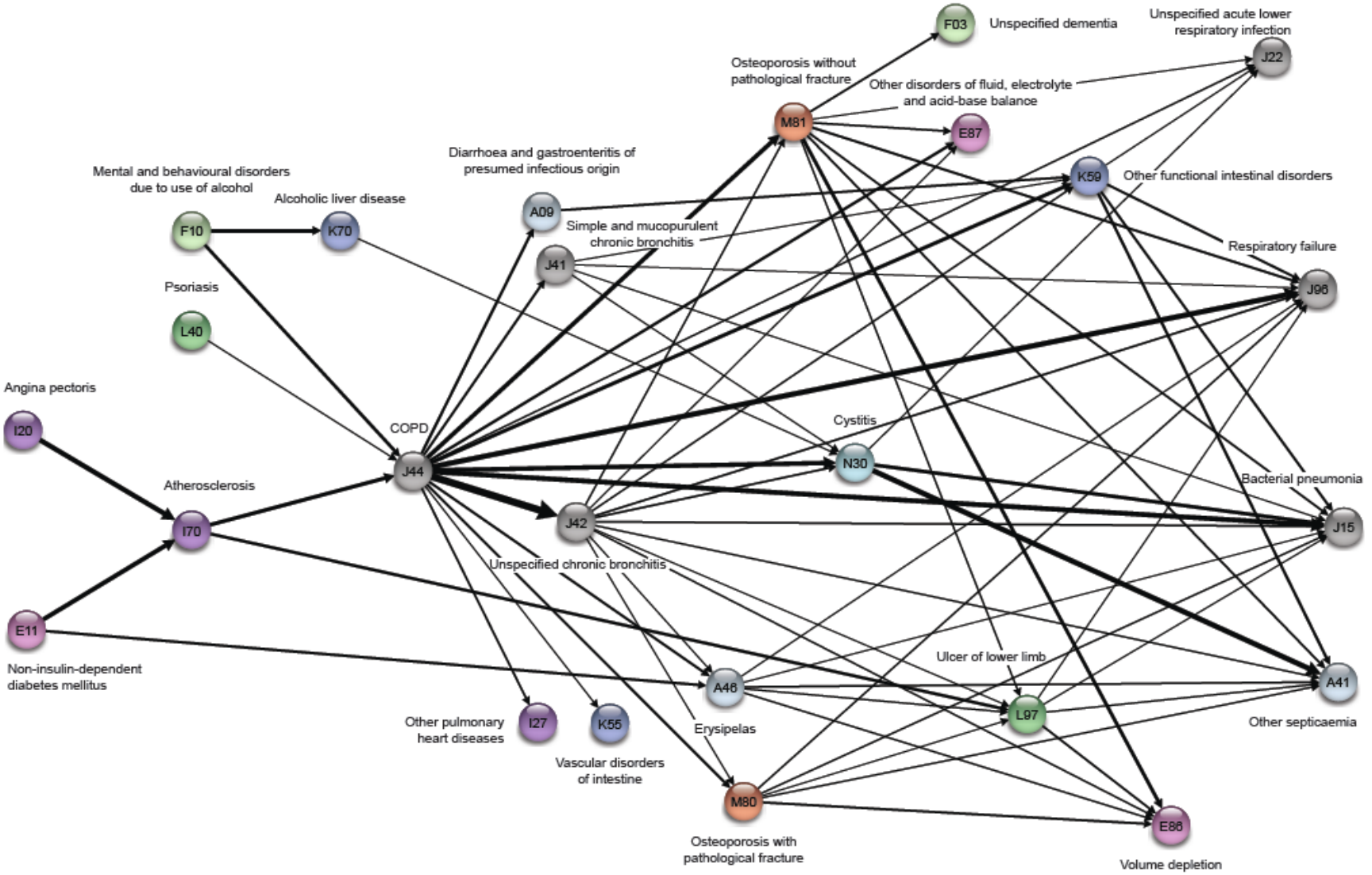
AB Jensen et al., Nature Comm., 2014

Diabetes trajectory network

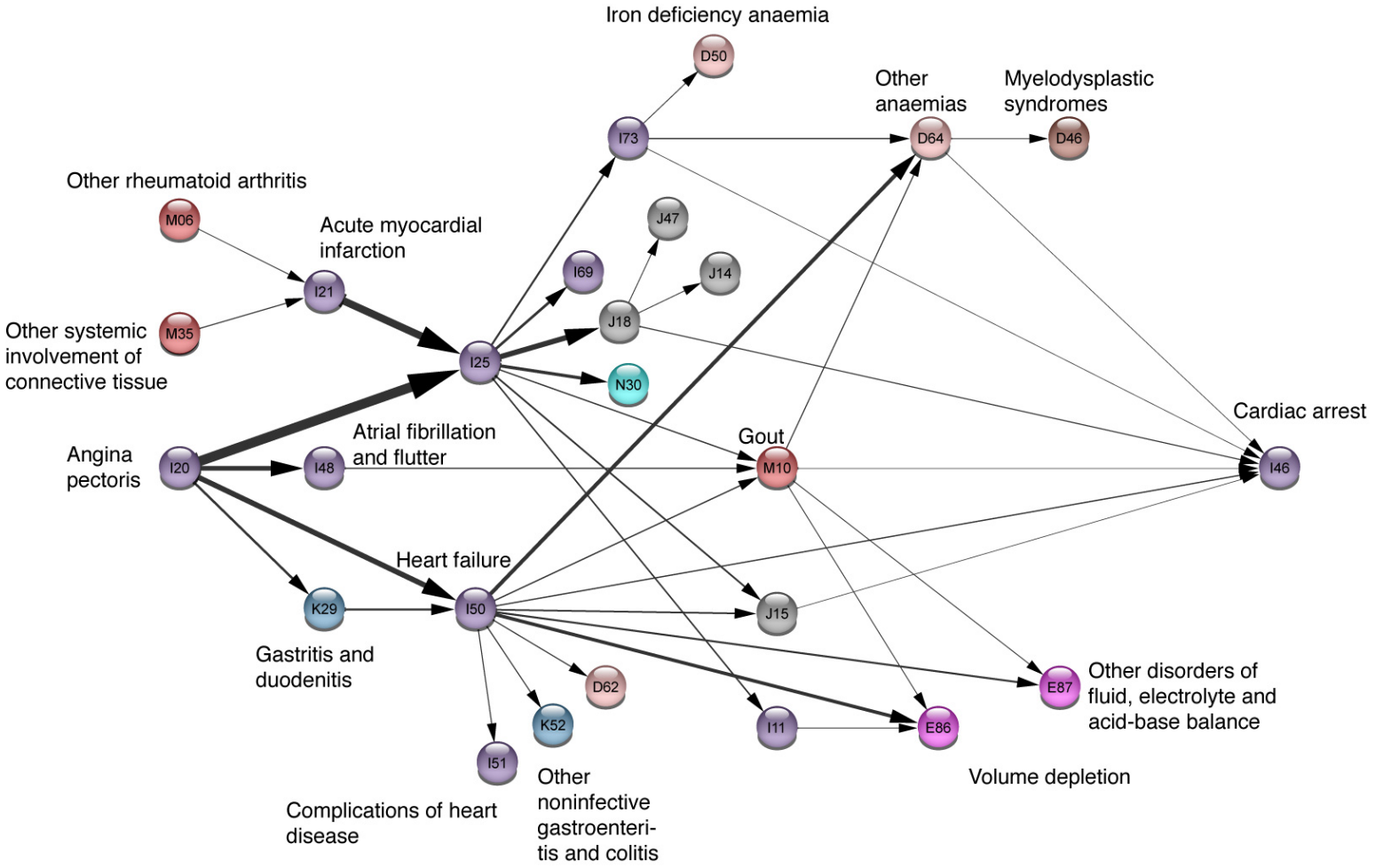


COPD trajectory cluster

with five preceding diagnoses leading to COPD and some of the possible outcomes



Cardiovascular trajectory network



AB Jensen et al.,
Nature Comm., 2014

Disease trajectory: towards an ontology

Reduced 6.2 million trajectories to

→ **1,171 trajectories of 4 steps**

Cover 140 diagnoses and >0.5 million unique patients

Clustering identified 5 major disease comorbidity clusters which incorporate ~600 of the trajectories

Musings

Highly accessed

The $1,000 genome, the $100,000 analysis?

Elaine R Mardis

Correspondence: Elaine R Mardis emardis@wustl.edu

▾ Author Affiliations

The Genome Center at Washington University School of Medicine, 4444 Forest Park Blvd, St Louis, MO 63108, USA

Genome Medicine 2010, **2**:84 doi:10.1186/gm205

The electronic version of this article is the complete one and can be found online at: http://genomemedicine.com/content/2/11/84

Published: 26 November 2010

Genome Medicine
Volume 2
Issue 11

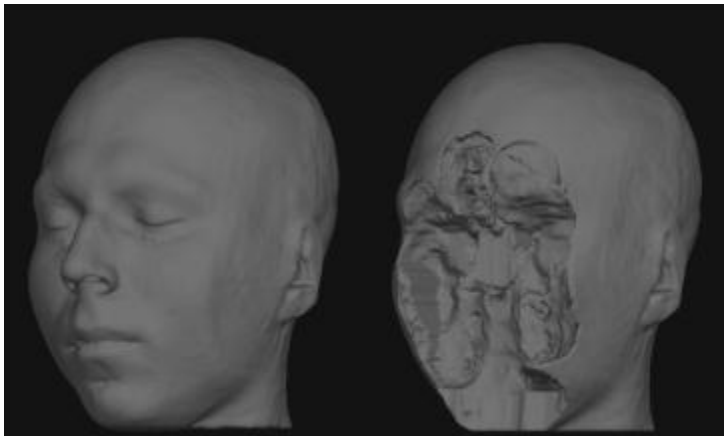
Viewing options
Abstract
Full text
PDF (220KB)

Associated material
PubMed record
Readers' comments

Related literature
Articles citing this article
on Google Scholar
on PubMed Central
Other articles by authors
▸ on Google Scholar
▸ on PubMed

De-identification

In the Nordic countries and in some other countries we can minimize the effort on rediscovering what we already know



Acknowledgements

EPR and registry analysis

Peter Bjødstrup Jensen, CPR/KU
Robert Eriksson, CBS/DTU/CPR/KU
Anders Bøck Jensen, CBS/DTU/CPR/KU
Teresa A Ajslev, U. Copenhagen
Pope Mosely, U. New Mexico
Tudor Oprea, U. New Mexico
Henriette Schmock, Sct. Hans Hospital
Lars Juhl Jensen, CPR/KU
Thomas Werge, Sct. Hans Hospital
Francisco Simões Roque, CBS/DTU
Eva Roitmann, CBS/DTU
Anders Juhl, Rigshospitalet, Copenhagen
Marlene Dalgaard, Rigshospitalet, Copenhagen
Massimo Andreatta, Copenhagen
Thomas Hansen, Sct. Hans Hospital
Karen Søeby, Hvidovre Hospital
Søren Bredkjær, Region Sealand
Thorkild IS Sørensen, U. Copenhagen

Steno Diabetes Center & Hagedorn

Peter Rossing
Henrik Ullits Andersen
Regine Bergholdt
Thomas Almdal
Flemming Pociot
Torben Hansen, now U. Copenhagen
Oluf Borbye Pedersen, now U. Copenhagen

SEVENTH FRAMEWORK PROGRAMME

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

VILLUM FONDEN

novo nordisk fonden

TRANSLATIONAL GENETICS

Mining electronic health records: towards better research applications and clinical care

Peter B. Jensen[1], Lars J. Jensen[1] and Søren Brunak[1,2]

Abstract | Clinical data describing the phenotypes and treatment of patients represents an underused data source that has much greater research potential than is currently realized. Mining of electronic health records (EHRs) has the potential for establishing new patient-stratification principles and for revealing unknown disease correlations. Integrating EHR data with genetic data will also give a finer understanding of genotype–phenotype relationships. However, a broad range of ethical, legal and technical reasons currently hinder the systematic deposition of these data in EHRs and their mining. Here, we consider the potential for furthering medical research and clinical care using EHR data and the challenges that must be overcome before this is a reality.

Clinical decision support (CDS). Software systems providing support for decision making to physicians through the application of health knowledge and logical rules to patient data.

Biobanks Central repositories of biological material that are mainly used for research. They facilitate the re-use of collected samples in different research projects.

Information technology has transformed the way health care is carried out and documented. Presently, the practice of health care generates, exchanges and stores huge amounts of patient-specific information. In addition to the traditional clinical narrative, databases in modern health centres automatically capture structured data relating to all aspects of care, including diagnosis, medication, laboratory test results and radiological imaging data.

This transformation holds great promise for the individual patient as richer information, coupled with clinical decision support (CDS) systems, becomes readily available at the bedside to support informed decision-making and to improve patient safety[1,2]. It is especially interesting when traditional health-care-sector data is linked with biobanks and genetic data[4].

Despite the great potential, researchers who wish to analyse large amounts of patient data are still faced with technical challenges of integrating scattered, heterogeneous data, in addition to ethical and legal obstacles that limit access to the data[5,6]. It is hoped that large-scale adoption of health information technology (HIT) infrastructure in the form of electronic health records (EHRs) and agreed standards for interoperability and schemes for privacy and consent, will improve this situation (TABLE 1). With incentives for improved public health and the expected health budget savings[7,8], these matters