Big Data for Public Health –
Public Data for Big Health

Stefan Thurner
with

Peter Klimek, Silke Aichberger, Anna Chmiel

supported by

EC FP7 projects LASAGNE and MULTIPLEX
date — patient ID — HCP ID — location — diagnosis — side diagnoses — prescription — price of generic drug/treatment — pharmacy ID — price of drug — date of purchase
100.000.000 lines per year
Data set

→ medical claims data

for every payed healthcare service there is one data line
Data set

- 8,000,000 patients
- 100,000,000 patient visits per year
- 2,000,000 hospitalisations per year
- 12,000 health care providers
- 6,102 diseases (ICD10 code)
- 1,171 drugs (ATC code)
- 255 hospitals
- 1,238 pharmacies
Network medicine
Co-morbidity networks

starting point: usually patients have more than one disease which diseases occur together? → co-morbidity networks

co-morbidity networks = health state of population (phenotype)

what can we learn from co-morbidity networks?
What is a co-morbidity network?

coopercurrence of diseases in population

diseases ‘linked’ if many patients have both at same time
Co-morbidity network of Austria
The co-morbidity network: children
The co-morbidity network: adults age 40-48
Empirical finding

the way individual diseases occur = diffusion on these networks

Prediction of health trajectories
Co-morbidity networks allow predictions
Co-morbidity networks allow predictions: DM

if have diabetes what is odds ratio to have any other disease?

Co-morbidity network of diabetes explains ...

- analysis equivalent to 40,000 individual epidemiological studies
- check which co-morbidity is causal
- confirm controversial relation of increased risk for Parkinson
- gender differences in progression of congestive heart failure
- females lower risk of hypertension during fertile age
- type 1 diabetes leads detection of depressions
- schizo-affective disorders lead type 2 diabetes, suggesting similar pathogenic or medication-related mechanisms

Efficacy of prevention
Co-morbidity networks and prevention

identify co-morbidities – check ‘causality’ – treat cause
New classification of diseases
What is Diabetes?

Observe:

diabetes co-occurs with other diseases in robust patterns

→ allows us to classify diabetes differently
New “types” of diabetes – defined by co-morbidity

define new phenomenological types of DM through co-morbidity
Which drug / therapy works?
• take a disease for which 2 therapies exist A and B
• compute all co-morbidities following therapy A
• compute all co-morbidities following therapy B
• compare: follow up costs, hospitalization time, co-morbidities
How genetic is your disease?
What explains a disease?

- genetic factors
- metabolic factors
- environmental / toxicogenetic factors
- epigenetic factors
How genetic is diabetes?

Genes associated with diabetes type 2

- HHEX/IDE/KIF11
- TCF7L2
- KCNJ11
- MTNR1B
- HNF1A
- FTO
- GCKR
- PPARG
- ADCY5
- CDKAL1
- SLC30A8
- CRY2
- FADS1

type 2 diabetes 25% hereditary

5-10% of variance explained by gene variants*

→ hard to tell!

* ME Travers MI McCarthy, Human Genetics 130 41-58 (2011)
Compare co-morbidity and genotype networks

if co-morbidity network is “similar” to genetic network → yes it is genetic
Ranking of likelihood of genetic cause in multi-factoral diseases

• compute “similarity” between phenotypic and genetic network
• hypothesis: the more similar – the “more” genetic influence
• rank multi-factoral diseases wrt similarity in pheno-geno networks

only take cases that are unlikely to be of statistical origin

\[ p < 0.00001 \]
<table>
<thead>
<tr>
<th>Condition</th>
<th>Code</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1 diabetes mellitus</td>
<td>E10</td>
<td>0,50</td>
</tr>
<tr>
<td>Transient cerebral ischemic attacks and related syndromes</td>
<td>G45</td>
<td>0,50</td>
</tr>
<tr>
<td>Benign neoplasm of colon, rectum, anus and anal canal</td>
<td>D12</td>
<td>0,33</td>
</tr>
<tr>
<td>Eating disorders</td>
<td>F50</td>
<td>0,33</td>
</tr>
<tr>
<td>Polycythemia vera</td>
<td>D45</td>
<td>0,25</td>
</tr>
<tr>
<td>Other diseases of intestine</td>
<td>K63</td>
<td>0,25</td>
</tr>
<tr>
<td>Other cerebrovascular diseases</td>
<td>I67</td>
<td>0,21</td>
</tr>
<tr>
<td>Other and unspecified diseases of blood and blood-forming organs</td>
<td>D75</td>
<td>0,21</td>
</tr>
<tr>
<td>Other congenital malformations of heart</td>
<td>Q24</td>
<td>0,20</td>
</tr>
<tr>
<td>Malignant neoplasm of heart, mediastinum and pleura</td>
<td>C38</td>
<td>0,16</td>
</tr>
<tr>
<td>Mesothelioma</td>
<td>C45</td>
<td>0,16</td>
</tr>
<tr>
<td>Specific personality disorders</td>
<td>F60</td>
<td>0,16</td>
</tr>
<tr>
<td>Overweight and obesity</td>
<td>E66</td>
<td>0,13</td>
</tr>
<tr>
<td>Other cardiac arrhythmias</td>
<td>I49</td>
<td>0,13</td>
</tr>
<tr>
<td>Cerebral infarction</td>
<td>I63</td>
<td>0,13</td>
</tr>
<tr>
<td>Type 2 diabetes mellitus</td>
<td>E11</td>
<td>0,11</td>
</tr>
<tr>
<td>Secondary parkinsonism</td>
<td>G21</td>
<td>0,11</td>
</tr>
<tr>
<td>Other and unspecified myopathies</td>
<td>G72</td>
<td>0,11</td>
</tr>
<tr>
<td>Congenital malformations of cardiac chambers and connections</td>
<td>Q20</td>
<td>0,10</td>
</tr>
<tr>
<td>Other congenital malformations of eye</td>
<td>Q15</td>
<td>0,09</td>
</tr>
<tr>
<td>Congenital malformations of aortic and mitral valves</td>
<td>Q23</td>
<td>0,09</td>
</tr>
<tr>
<td>Parkinsons disease</td>
<td>G20</td>
<td>0,08</td>
</tr>
<tr>
<td>Essential (primary) hypertension</td>
<td>I10</td>
<td>0,07</td>
</tr>
<tr>
<td>Anoph</td>
<td>Q11</td>
<td>0,06</td>
</tr>
</tbody>
</table>
Classify multi-factoral diseases

- do the same with metabolic, environmental, pathway networks

→ every disease gets assigned 3 numbers:

- genetic rank
- toxicogenetic rank
- pathway importance
Side effects – personalized
Side effect networks

you have disease $x \rightarrow$ get medication $y \rightarrow y$ causes disease $z$
Side effects for diabetes treatments: Insulin

Diabetes E11
- C04
- C10
- C18
- C22
- pancreas x 10
- C25
- C34
- peritoneum x 7
- C48
- C50
- C56
- brain x 4
- C71
- lymph x 2
- C77

- Insulin
- Medformine
- Statine
- Insulin + Statine
Side effects for treatment: Insulin + Statines

Diabetes E11
- C04
- C10
- C18
- C20
- C25
- C34
- C48
- C50
- C61
- C77
- C78

- Insulin
- Medformine
- Statine
- Insulin + Statine

colon x 0.7
prostate x 0.5
lymph x 0.6
second. x 0.5
What are the side effects of Metformin?

Diabetes E11
C04
C10
C18
C20
C25
C34
C50
C55
C61
C77
C78

uterus x 2.7
second. x 2.4

Insulin
Metformine
Statine
Insulin + Statine
Disclaimer

NO medical statement are made here!

- This reflects the *status quo* in the population only
- No understanding why
- No mechanism clarified
- No medical understanding
- Need experts for this
A telescope into the past
Unexpected causes for diabetes?

take all \( \sim 300,000 \) diabetes patients. Fraction of patients in population given birth date? \( \rightarrow \) famines in Austria

Message: mother suffers hunger in pregnancy \( \rightarrow \) baby develops diabetes in later life

S Thurner et al. PNAS 110, 4703-4707, (2013)
A window into the past II
A window into the past II
Vizualize healthcare system
The healthcare system is ...

- network of patient flows
- network of information flows
- network of cash flows
- it is a co-evolving multi-layer network!
Patient-flow network

- many patient flows are medically reasonable – many are not
- health care costs can be completely transparent – if wanted
- patient flows + comorbidity across age $\rightarrow$ future costs
Patient-flow network as we like it

- General practitioner
- Medical specialist
- Pharmacy
- Hospital
Patient-flow network
Patient-flow network of Austria 2006
Patient-flow network: patterns
With this information one can monitor ...

- regional differences in quality of care / prescriptions / success rates / diagnoses / costs / transparency /
- ‘cyclic flows’: which ones are medically necessary?
- success of prevention schemes (medical & economic)
- nation-wide behaviour of patients: drug use, vaccination rates
- visiting frequency as function of accessibility of HCP
- optimal health care coverage densities
Vision

• 1:1 agent based model of the Austrian health care system
• use data to fully calibrate the model
• estimate how individual patients take decisions
• estimate how HCP take decisions
• make policy experiments: insurers and politicians
Summary

• begin to predict health trajectories from co-morbidity networks
• see which medication works
• make gender differences visible
• compute personalized side effects
• new classification of diseases in terms of co-morbidity
• compute medical and economical value of prevention schemes
• quantify resilience, robustness, sustainability of health care system
Collaboration partners – experts

Alexandra Kautzky-Willer, MUW
Gottfried Endel, Hauptverband
Miriam Leitner, MUW
Irmgard Schiller-Frühwirth, Hauptverband
Herwig Ostermann, Gesundheit Österreich
Klaus Kratochwill, MUW
Statistics

• Phenotype NW: $\phi$ is correlation coefficient (binary), Kramers coefficient

• Relative Risk:

$$\frac{a/(a + c)}{b/(b + d)}$$

• Odds Ratio:

$$\frac{P(A)(1 - P(B))}{P(B)(1 - P(A))}$$

where $P(A) = a/(a + c)$ and $P(B) = b/(b + d)$