i2b2 Clinical Datawarehouse
at the Pompidou University Hospital
in Paris (APHP - HEGP)

Paul Avillach, MD, PhD
Hôpital Européen Georges Pompidou (HEGP)
Faculté de Médecine René Descartes, Université Paris 5
INSERM UMR_S 872 eq22

CBMI – Harvard Medical School

ErasmusMC University
Location

- Paris 15th district

Within AP-HP

- HEGP is the most recent acute care hospital within the 38 AP-HP hospitals
- HEGP meets the needs of the 600,000 inhabitants of the Paris south-west
HEGP background

Opening: July 2000-

Hôpital Laennec (1634)  Hôpital Boucicaut  Hôpital Broussais
HEGP BDW

Production environment

CIO: Pr Patrice Degoulet
Pr Anita Burgun

HER/CDW integration

EHR: Operational Database (ODS)

EHR: Mirrored Database

Clinical Data Warehouse (CDW)

External Databases

ETL suite (Talend Open Studio)

Real time requests

Data Analysis
Data Mining

i2b2/tranSMART tools

Business Object

IBM Ilog Rules

R
<table>
<thead>
<tr>
<th>Type of data</th>
<th>H</th>
<th>OV</th>
<th>Start date</th>
<th># unique patients</th>
<th># values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic (age, sex, Hospital vital status)</td>
<td>X</td>
<td>X</td>
<td>1971</td>
<td>606 524</td>
<td></td>
</tr>
<tr>
<td>Vital signs (temperature, blood pressure, weight, ...)</td>
<td>X</td>
<td>X</td>
<td>2000</td>
<td>141 164</td>
<td>14 213 951</td>
</tr>
<tr>
<td>Diagnostic codes (DRG ICD10)</td>
<td>X</td>
<td></td>
<td>1995</td>
<td>305 369</td>
<td>2 626 792</td>
</tr>
<tr>
<td>Procedures (French CCAM codes)</td>
<td>X</td>
<td></td>
<td>2004</td>
<td>241 482</td>
<td>3 200 482</td>
</tr>
<tr>
<td>Clinical data (DxCare questionnaires)</td>
<td>X</td>
<td>X</td>
<td>1971</td>
<td>391 218</td>
<td>46 506 217</td>
</tr>
<tr>
<td>Free text reports*: Hospitalization, Surgery, consultations, ...</td>
<td>X</td>
<td>X</td>
<td>2004</td>
<td>289 614</td>
<td>1 961 985</td>
</tr>
<tr>
<td>Free text reports**: Imaging and pathology</td>
<td>X</td>
<td>X</td>
<td>2000</td>
<td></td>
<td>1 000 000</td>
</tr>
<tr>
<td>Pathology codes (ADICAP)</td>
<td>X</td>
<td>X</td>
<td>2000</td>
<td>73 173</td>
<td></td>
</tr>
<tr>
<td>Biology results (without antibiograms)</td>
<td>X</td>
<td>X</td>
<td>2000</td>
<td>338 068</td>
<td>88 607 301</td>
</tr>
<tr>
<td>Antibiograms</td>
<td>X</td>
<td>X</td>
<td>2000</td>
<td>39 040</td>
<td>4 058 842</td>
</tr>
<tr>
<td>Drug prescription (without Chemotherapy)</td>
<td>X</td>
<td>X</td>
<td>1988</td>
<td>88 567</td>
<td>2 612 742</td>
</tr>
<tr>
<td>Validation of Drug prescription by pharmacists</td>
<td>X</td>
<td></td>
<td>2002</td>
<td>67 151</td>
<td>1 691 137</td>
</tr>
</tbody>
</table>
HEGP CDW

Query Name:

Group 1
- (I21-I21) Infarctus aigu du myocarde

Group 2
- Dates
- Occurs > 0x
- Exclude

Group 3
- (I26-I26) Embolie pulmonaire

Query Status
- Finished Query: "IDM + EP"
- Patient Count - 73 patients
• 188 MD + Pharm trained

• IRB

Creation of an HEGP research ethical committee linked to the regional IRB
HEGP CDW

i2b2 CDW use (Jan. 2011-Dec. 2013)

• Level 1 studies: aggregated data (e.g. potential trial recruitment)
  ➢ Free access for all HEGP health professionals
  ➢ 2 209 requests

• Level 2 and 3 studies: access to patient level data
  ➢ Structured written project
  ➢ Validation by the HEGP IRB
  ➢ Transmission to the regional IRB
  ➢ Level 2: anonymized patient data
  ➢ Level 3: de-anonymized patient data

➢ IRB approval for 35 projects
Genome Wide Association Study
(1 Phenotype compared to ALL SNPs)

cases
(ex: systemic sclerosis)

cases DNA

controls

controls DNA

-log(P-value)

chromosomes

- HLA region (ch 6)

compare ALL SNPs to find differences between cases and controls
**Genome Wide Association Study**
(1 Phenotype compared to ALL SNPs)

- **cases** (ex: systemic sclerosis)
- **controls**

- cases DNA
- controls DNA

- HLA region (ch 6)

- compare ALL SNPs to find differences between cases and controls

**Phenome Wide Association Study**
(1 SNP compared to ALL Phenotypes)

- allele G patients group
- allele A patients group

- allele G patients phenotype
- allele A patients phenotype

- I21 (myocardial infarction)

- compare ALL DIAGNOSIS to find differences between cases and controls
PheWAS study on i2b2

Phenome-Wide Association Studies on a Quantitative Trait: Application to TPMT Enzyme Activity and Thiopurine Therapy in Pharmacogenomics

Antoine Neuraz¹,², Laurent Chouchana³, Georgia Malamut⁴, Christine Le Beller⁵, Denis Roche⁶, Philippe Beaune³,⁶, Patrice Degoulet¹,², Anita Burgun¹,², Marie-Anne Loriot³,⁶, Paul Avillac¹,²*

1 Biomedical Informatics and Public Health Department, University Hospital HEGP, AP-HP, Paris, France, 2 INSERM UMR_S 872 Team 22: Information Sciences to support Personalized Medicine, Université Paris Descartes, Sorbonne Paris Cité, Faculté de Médecine, Paris, France, 3 INSERM UMR-S 775, Université Paris Descartes, Sorbonne Paris Cité, Paris, France, 4 Gastroenterology Department, University Hospital HEGP, AP-HP, Paris, France, 5 Pharmacovigilance Center, University Hospital HEGP, AP-HP, Paris, France, 6 Biochemistry, Pharmacogenetics and Molecular Oncology Unit, University Hospital HEGP, AP-HP, Paris, France

Datatypes:
- ICD10
- Drug prescription
- Lab results
- Clinical Notes

Time points+++
Methodes: Selection of trait: enzymatic Activity TPMT

Thiopurine (DRUG) → 6-TIMP (active metabolite) → elimination

increased toxicity

Phenotype

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Low activity</th>
<th>Intermediate activity</th>
<th>Normal Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thiopurine dose</td>
<td>10 % dose</td>
<td>30 – 70 % dose</td>
<td>100 % dose</td>
</tr>
</tbody>
</table>

FDA & EMA recommendations?
ICD10

Very High TPMT activity VS others

Diabetes mellitus
Nutritional anemias

Odds Ratio *
• 1
• 2
• 3
• 4
• 5

256 ICD-10 based aggregated
Acknowledgments

Informatics & Public Health Dept:

- Eric Zapletal,
- Antoine Neuraz,
- Anita Burgun,
- Patrice Degoulet