SNOMED CT Support
Interoperability in Chinese Cancer Clinical Lab Test Data

Zhi Wang, Zhouguang Hui, Qiang Guo, Yibo Gao, Jianyang Wang, Fang Wang, Mengchun Gong, Wenzhao Shi
Digital China Health
PART 1

Introduction of DCH
Honor of DCH

- **Partnership**
  - Fudan Pediatrics Hospital: Fudan Pediatrics—Digital China Health Research Center
  - Broad Institute
  - Philips

- **Project**
  - 2015: 863 Project《Data Analysis and Application for Malignant Tumor Big Data》
  - 2016: National Key Research and Development Plan《Clinical Cohort Research for Rare Diseases》
  - 2017: National Cancer Center and the platform
  - 2018: NKRDP《The Research on the Construction of Clinical Big Data Platform and Biological Sample Library》
  - 2018: NKRDP《New Service and Solution of Artificial Intelligence-based Clinical Decision Support》
National Health and Wellness Committee of the People's Republic

National Health Medical Big Data Security Management Committee

China Health Medical Big Data Industry Development Group Corporation

China Health Medical Big Data Technology Development Group Corporation

China Health Medical Big Data Co., Ltd.

DCH

Other Company

Health Care Big Data Industry Alliance
Vice President

Health Care Big Data Population Informationization Professional Committee

Health Medical Big Data Oncology Committee
Vice President

Chronic Disease Prevention Alliance

Geriatrics Big Data Professional Committee
Vice President
National Cancer Center

Data Center and Platform

① NCC

② Provincial Tumor Hospital

③ All Tumor Hospital

④ Tumor Dept. for all general hospital

1 30 174 1000+2000
The Value of Clinical Big Data

- Hospital
- Pharma
- Insurance
Real World Data from Hospital

- Collection
- Unstructured Data
- Distributive distribution
- Heterogeneity
Real World Data

- Information Model + Knowledge Model
- Data quality:
  - Data Correlation:
    - Data is collected not for research
    - Need semantic standardization
  - Data Reliable:
    - Accuracy
    - Integrity
    - Control of Bias
OHDSI CDM

OHDSI=Observational Health Data Science and Informatics
CDM=Common Data Model
OHDSI: Global group
Background

- OHDSI Common Data Model (CDM) will be used for the big data platform (Vinci) of DCH;
- The data from the tumor hospitals would be transferred and stored in OHDSI CDM;
- LOINC is the standard vocabulary of OHDSI CDM for the domain of lab test;
Terminology Mapping for LOINC

- LOINC is the standard vocabulary for lab test in OHDSI CDM.
- Prof. Zhang Lin translated the terms of LOINC to Chinese
- Six axes in LOINC: Component, Property, Timing, System, Scale and Method
- One to three axes information could be collected from information system of tumor hospitals
  - Component, System (sample)
  - single axial data comes from NLP
- It is difficult to exactly map the terms to LOINC
Objective

- To compare the mapping for lab test name from the tumor hospitals of China to LOINC and SNOMED CT
Methods

- 4451 laboratory terms were collected from 4 tumor hospitals.
- The cumulative frequency of all terms is 59285236.
- After sorting by frequency, we selected 643 high frequency items with 80% frequency to map to SNOMED.
- Relma 6.22 was used for the mapping of lab test name to LOINC, language set to Chinese.
- SNOMED International Browser was used for the mapping of lab test name to SNOMED CT.
- Two terminologists manually mapped the 643 terms to LOINC and SNOMED CT, among which there were 100 laboratory terms were specifically related to tumor test.
- Then a team of two senior terminologists checked the mapping result.
## Data Description

<table>
<thead>
<tr>
<th>Data Source</th>
<th>No. Of Lab Test Items</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor Hospital 1</td>
<td>1370</td>
<td>9127184</td>
</tr>
<tr>
<td>Tumor Hospital 2</td>
<td>1588</td>
<td>22986357</td>
</tr>
<tr>
<td>Tumor Hospital 3</td>
<td>843</td>
<td>7375863</td>
</tr>
<tr>
<td>Tumor Hospital 4</td>
<td>1339</td>
<td>12847618</td>
</tr>
</tbody>
</table>
Workflow for data processing

Data from Hospital → Data Preprocessing → Lab Test Items Selection → Standard Operation Protocol

LOINC Mapping → Result

SNOMED CT Mapping →
Result

- Among the lab test terms, 308 (47%) were complete matched to LOINC and 335 concepts (53%) unmatched.
- Most terms (92%) were mapped to SNOMED CT,
  - including 180 concepts (28%) mapped to pre-coordinated expression,
  - 411 terms (64%) mapped to post-coordinated expression,
  - 52 terms (8%) unmapped.
Discussion

- Due to the lack of information, Chinese Lab test items were hard to map to LOINC exactly;
- It is needed to find a way to map the terms to a standard vocabulary in order to perform;
- Post-Coordination might be a way to solve this issue.
- In order to solve this problem, we are trying to authoring local term with two or three attributes to represent the clinical meaning
- Carbohydrate antigen 125 (Procedure) and Cancer Antigen 125 (Substance)
  - Carbohydrate antigen 125 (Procedure) is inactive
  - Cancer Antigen 125 (Substance) is a “Substance”
### Result for SNOMED CT and LOINC Mapping

<table>
<thead>
<tr>
<th>Chinese Source Term</th>
<th>SPECIMEN</th>
<th>Source Unit</th>
<th>SCT ID</th>
<th>SCT Name</th>
<th>LOINCID</th>
<th>Component</th>
<th>Property</th>
<th>Time Aspect</th>
<th>System</th>
<th>Scale</th>
<th>Method</th>
<th>Class</th>
<th>SIRANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>尿β2-微球蛋白</td>
<td>尿液</td>
<td>mg/l</td>
<td>408227007</td>
<td>Urine beta 2 microglobulin level (procedure)</td>
<td>1953-9</td>
<td>MCnc</td>
<td>Pt</td>
<td>Urine</td>
<td>Qn</td>
<td>—</td>
<td></td>
<td>CHEM</td>
<td>0</td>
</tr>
<tr>
<td>乳酸脱氢酶</td>
<td>血清</td>
<td>iu/l</td>
<td>313854008</td>
<td>Serum lactate dehydrogenase measurement (procedure)</td>
<td>83077-8</td>
<td>MCnc</td>
<td>Pt</td>
<td>Urine</td>
<td>Qn</td>
<td>IA</td>
<td>CHEM</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>甲胎蛋白[定量]</td>
<td>血清</td>
<td>ng/ml</td>
<td>104404005</td>
<td>Alpha-1-fetoprotein measurement, serum (procedure)</td>
<td>14804-9</td>
<td>CCnc</td>
<td>Pt</td>
<td>Ser/Plas</td>
<td>Qn</td>
<td>Reaction: lactate to pyruvate</td>
<td>CHEM</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14805-6</td>
<td>CCnc</td>
<td>Pt</td>
<td>Ser/Plas</td>
<td>Qn</td>
<td>Reaction: pyruvate to lactate</td>
<td>CHEM</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>乙酰轻链k</td>
<td>血清</td>
<td>mg/l</td>
<td>444307003</td>
<td>Detection of ordinal level of free immunoglobulin light chain in serum or plasma specimen (procedure)</td>
<td>36916-5</td>
<td>MCnc</td>
<td>Pt</td>
<td>Ser</td>
<td>Qn</td>
<td>—</td>
<td>CHEM</td>
<td>386</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80515-0</td>
<td>MCnc</td>
<td>Pt</td>
<td>Ser</td>
<td>Qn</td>
<td>Nephelometry</td>
<td>CHEM</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
## Result for SNOMED CT and LOINC Mapping

<table>
<thead>
<tr>
<th>Chinese Source Term</th>
<th>Source Sample Term_SPECIMEN</th>
<th>Source Unit</th>
<th>SCT ID</th>
<th>SCT Name</th>
<th>LOINCID</th>
<th>Component</th>
<th>Property</th>
<th>Time_Aspect</th>
<th>System</th>
<th>Scale</th>
<th>Method</th>
<th>Class</th>
<th>SIRANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>癌胚抗原</td>
<td>血清Serum</td>
<td>ng/ml</td>
<td>60267001</td>
<td>Carcinoembryonic antigen measurement (procedure)</td>
<td>2039-6</td>
<td>MCnc</td>
<td>Pt</td>
<td>Ser/Plas</td>
<td>Qn</td>
<td>—</td>
<td>CHEM</td>
<td>312</td>
<td></td>
</tr>
<tr>
<td>糖抗原125</td>
<td>胸腹水Pleural fluid/Ascites</td>
<td>u/ml</td>
<td>80529009</td>
<td>CA 125 measurement (procedure)</td>
<td>83085-1</td>
<td>MCnc</td>
<td>Pt</td>
<td>Ser/Plas</td>
<td>Qn</td>
<td>IA</td>
<td>CHEM</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>糖抗原199</td>
<td>血清Serum</td>
<td>u/ml</td>
<td>40939009</td>
<td>Cancer antigen 19-9 measurement (procedure)</td>
<td>11210-2</td>
<td>ACnc</td>
<td>Pt</td>
<td>Body fld</td>
<td>Qn</td>
<td>—</td>
<td>CHEM</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15156-3</td>
<td>ACnc</td>
<td>Pt</td>
<td>Body fld</td>
<td>Qn</td>
<td>Dilution</td>
<td>CHEM</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24108-3</td>
<td>ACnc</td>
<td>Pt</td>
<td>Ser/Plas</td>
<td>Qn</td>
<td>—</td>
<td>CHEM</td>
<td>677</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>83084-4</td>
<td>ACnc</td>
<td>Pt</td>
<td>Ser/Plas</td>
<td>Qn</td>
<td>IA</td>
<td>CHEM</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Serum lactate dehydrogenase measurement (procedure)
尿β₂-微球蛋白-Urine beta-2 microglobulin

Urine beta 2 microglobulin level

- Urine protein test
- Fluid sample globulin measurement
- Beta-2-microglobulin measurement

- Has specimen
  - Urine specimen
- Component
  - Beta 2 microglobulin
- Method
  - Measurement - action
The value of concept model for AFP measurement in serum

Alpha-1-Fetoprotein measurement

- Tumor marker measurement
- Protein measurement
  - Component: Alpha fetoprotein
  - Method: Measurement - action

Alpha-1-fetoprotein measurement, serum

- Alpha-fetoprotein blood test
  - Component: Alpha fetoprotein
  - Method: Measurement - action
<table>
<thead>
<tr>
<th>Chinese Source Term</th>
<th>SNOMED CT Post-Coordination Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>甲胎蛋⽩白[定量]</td>
<td>Alpha Fetoprotein</td>
</tr>
<tr>
<td></td>
<td>Observable entity (observable entity): 370130000</td>
</tr>
<tr>
<td>糖抗原125</td>
<td>CA 125</td>
</tr>
<tr>
<td></td>
<td>Observable entity (observable entity): 704327008</td>
</tr>
<tr>
<td>糖抗原199</td>
<td>CA 199</td>
</tr>
<tr>
<td></td>
<td>Observable entity (observable entity): 704327008</td>
</tr>
</tbody>
</table>
Terminologists of Digital China Health

Fengxiang Chang
Ying Zhang
Fang Wang
Chenghuan Ding
Zhi Wang
Yishang Wang
Kuangyu Ma
谢谢 谢谢