Building Data Capacity for Patient Centered Outcomes Research

Nigam Shah
Acknowledgements

Funding:
• Federal – NLM, NHLBI (Past: NIGMS, NHGRI, NINDS, NCI, FDA)
• Institutional – Dept. of Medicine, Population Health Sciences, Dean’s office, Stanford Hospital
• Fellowships – Med Scholars, Siebel Scholars Foundation, Stanford Graduate Fellowship, NSF, DoD
• Industry – Healogics, Janssen R&D, Oracle, Baidu USA, Amgen, Google, Apixio, CollabRx, Curai
• Philanthropic Gifts
A patient timeline view of data
Perform interval algebra

Patients with a history of myocardial infarction who have pneumonia.

Patient with T2DM who are uncontrolled with metformin.

<table>
<thead>
<tr>
<th>Allen Statements</th>
<th>Inverse Relations</th>
<th>Pictoral Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$ before $Y$</td>
<td>$Y$ after $X$</td>
<td><img src="image1" alt="Example" /></td>
</tr>
<tr>
<td>$X$ equals $Y$</td>
<td>$Y$ equals $X$</td>
<td><img src="image2" alt="Example" /></td>
</tr>
<tr>
<td>$X$ meets $Y$</td>
<td>$Y$ met by $X$</td>
<td><img src="image3" alt="Example" /></td>
</tr>
<tr>
<td>$X$ overlaps $Y$</td>
<td>$Y$ overlapped by $X$</td>
<td><img src="image4" alt="Example" /></td>
</tr>
<tr>
<td>$X$ contains $Y$</td>
<td>$Y$ during $X$</td>
<td><img src="image5" alt="Example" /></td>
</tr>
<tr>
<td>$X$ starts $Y$</td>
<td>$Y$ started by $X$</td>
<td><img src="image6" alt="Example" /></td>
</tr>
<tr>
<td>$X$ finishes $Y$</td>
<td>$Y$ finished by $X$</td>
<td><img src="image7" alt="Example" /></td>
</tr>
</tbody>
</table>

Mate et al, A Method for the Graphical Modeling of Relative Temporal Constraints (Preprint)
Patients with disorders of glucose metabolism.

Patient with treated with “beta blockers”.

Navigate knowledge graphs
### State phenotype definitions

<table>
<thead>
<tr>
<th>Patient Journey</th>
<th>CPT codes</th>
<th>ICD codes</th>
<th>Medications</th>
<th>Procedures</th>
<th>Lab tests</th>
<th>Clinical notes</th>
<th>Bedside monitors</th>
<th>Wearables</th>
<th>Gene Expression</th>
<th>Phone usage</th>
<th>Browsing history</th>
<th>Social media</th>
<th>Audio recordings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- CPT codes
- ICD codes
- Medications
- Procedures
- Lab tests
- Clinical notes
- Bedside monitors
- Wearables
- Gene Expression
- Phone usage
- Browsing history
- Social media
- Audio recordings
Technology for timelines to data frames in real time

Interval algebra
Knowledge graph
Phenotype criteria
ACE: the Advanced Cohort Engine for searching longitudinal patient records

Alison Callahan, Vladimir Polony, José D Posada, Juan M Banda, Saurabh Gombar, Nigam H Shah

Author Notes

Journal of the American Medical Informatics Association, ocab027, https://doi.org/10.1093/jamia/ocab027

Published: 13 March 2021

Article history ▼

//patients with cryptogenic stroke, where we don't have an obvious reason for the stroke

var stroke = INTERSECT(OR(ICD9=434, ICD9=436), NOT(OR(ICD9=393, ICD9=394, ICD9=397.1, ICD9=398, ICD9=246, ICD9=424.9, ICD9 =V43, ICD9=433.1, ICD9=431, ICD9=434.11, ICD9=434.01)), AGE (40 YEARS, 90 YEARS), VISIT TYPE="inpatient visit", NOT(TEXT ="heart valve prosthesis"), NOT(TEXT="disease of mitral valve"), NOT(TEXT="rheumatic heart disease"))
Closing thoughts

• We need to adopt the construct of a patient time (rather than tables, and data models)
• We need to upgrade the computational infrastructure to speed up “time to analysis ready data frame”
• We need a stronger focus on systems and software beyond “methods development”