Synthea and SyntheticMass

Jason Walonoski

7 March 2017
What

- **Synthetic Patient Simulation**
  - Synthea is an open-source synthetic patient generator that simulates the medical history of synthetic patients.

- **High Quality Health Records**
  - Our mission is to output high-quality synthetic, realistic but not real, patient data and associated health records covering every aspect of health.

- **Freely Available**
  - The resulting data is free from legal, cost, privacy, and security restrictions for a variety of secondary uses in academia, research, industry, and government where realistic data is sufficient.
Why

- **High demand for EHR datasets**
  - Non-clinical or secondary uses including software development, testing, clinical training, policy analysis, where realistic (but not real) data is sufficient

- **Lack of Access**
  - EHR datasets are difficult to obtain

- **Costs and Demand**
  - Anonymized records are being bought and sold by federal and state health departments, hospitals, health insurers, pharmacists, general practitioners, government lobby groups, law firms, charities, marketers

- **Risks**
  - Real patient records carry privacy, confidentiality, consent, policy, and legal risks that effectively prevent use

- **Not Anonymous**
  - Deidentified and anonymized records have been successfully reidentified
Synthea Architecture

- Clinical Care Maps
- Incidence & Prevalence Statistics
- Disease Modules (State Machines)
- Synthea Architecture
- Synthetic World Population
- Exporters
- FHIR STU3
- CCDA
- SNOMED, LOINC, RxNorm
- Census Data Demographics
- Configuration
### Diseases

<table>
<thead>
<tr>
<th>Top 10 Reasons Patients Visit PCP</th>
<th>Top 10 Years of Life Lost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Routine infant/child health check</td>
<td>Ischemic Heart Disease</td>
</tr>
<tr>
<td>2. Essential Hypertension</td>
<td>Lung Cancer</td>
</tr>
<tr>
<td>3. Diabetes Mellitus</td>
<td>Alzheimer’s Disease</td>
</tr>
<tr>
<td>4. Normal Pregnancy</td>
<td>COPD</td>
</tr>
<tr>
<td>5. Respiratory Infections (Pharyngitis, Bronchitis, Sinusitis)</td>
<td>Cerebrovascular Disease</td>
</tr>
<tr>
<td>6. General Adult Medical Examination</td>
<td>Road Injuries</td>
</tr>
<tr>
<td>7. Disorders of Lipoid Metabolism</td>
<td>Self-Harm</td>
</tr>
<tr>
<td>8. Ear Infections (Otitis Media)</td>
<td>Diabetes Mellitus</td>
</tr>
<tr>
<td>9. Asthma</td>
<td>Colorectal Cancer</td>
</tr>
<tr>
<td>10. Urinary Tract Infections</td>
<td>Drug Use Disorders (limited to Opioids)</td>
</tr>
</tbody>
</table>
Appendicitis
Disease Models (as of February 2017)
SyntheticMass and Synthea

**Population**

Number of Residents

<table>
<thead>
<tr>
<th>Region Type</th>
<th>Cities and Towns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
<td>Synthetic data generated from Synthea</td>
</tr>
<tr>
<td>Total Population</td>
<td>1,006,324</td>
</tr>
<tr>
<td>Mean</td>
<td>2,867</td>
</tr>
<tr>
<td>Max</td>
<td>Boston Cities and Towns: 99,261</td>
</tr>
<tr>
<td>Min</td>
<td>Gosnold Cities and Towns: 19</td>
</tr>
</tbody>
</table>
### Patient Record

- **Family name**: Boehm218
- **Given name**: Alison619
- **Address**: 56822 Candice Orchard Apt. 325 Boston, MA 02293
- **City, State**: Boston, MA
- **Postal Code**: 02293

**Download Patient Data (FHIR JSON) | Download Patient Data (CCDA XML) | Send Data via Direct Message**

#### Demographics
<table>
<thead>
<tr>
<th>Category</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female Population</td>
<td>50.9% (148 of 351)</td>
</tr>
<tr>
<td>Male Population</td>
<td>49.1% (204 of 351)</td>
</tr>
<tr>
<td>Diabetes Prevalence</td>
<td>9.1% (37 of 351)</td>
</tr>
<tr>
<td>Opioid Addiction Prevalence</td>
<td>0.8% (104 of 351)</td>
</tr>
<tr>
<td>Heart Disease Prevalence</td>
<td>8.0% (8 of 351)</td>
</tr>
</tbody>
</table>

#### Observations

<table>
<thead>
<tr>
<th>Condition</th>
<th>Date of Onset</th>
<th>Date Resolved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypertension</td>
<td>21-Jun-1971</td>
<td>n/a</td>
</tr>
<tr>
<td>Asthma</td>
<td>21-Jun-1971</td>
<td>n/a</td>
</tr>
<tr>
<td>Sinusitis (disorder)</td>
<td>23-Mar-2010</td>
<td>14-Apr-2010</td>
</tr>
<tr>
<td>Chronic sinusitis (disorder)</td>
<td>07-Apr-2010</td>
<td>14-Apr-2010</td>
</tr>
<tr>
<td>Acute bronchitis (disorder)</td>
<td>14-Apr-2014</td>
<td>04-May-2014</td>
</tr>
</tbody>
</table>

#### Vaccinations

- **Name**: Hermiston857, Lexie730
  - **Gender**: male
  - **DOB**: 19-Jul-1933
- **Name**: Senger118, Kiera63
  - **Gender**: male
  - **DOB**: 14-Sep-1941
- **Name**: Miller704, Kelton341
  - **Gender**: male
  - **DOB**: 18-Jun-1946
- **Name**: Boehm218, Alison619
  - **Gender**: male
  - **DOB**: 11-Jul-1951
- **Name**: Watsica873, Numa391
  - **Gender**: female
  - **DOB**: 22-Nov-1951
Synthea Vision for the Future

Top 10 Years of Life Lost
Top 10 Reasons for PCP Visit
Long Tail of Diseases and Global Burden of Disease (GBD)
Who (external collaborators)

- **Jeff Eastman** (MiHIN)
- **Tom Gallagher** (University of Montana)
- **Kuda Dube** (Massey University [NZ])
  - Joel Waldock
- **Scott McLachlan** (Massey University [NZ])
- **Mark Braunstein** (Georgia Tech)
  - Jaya Rao, Daniel Sahu, Lichen Shen
- **James Agnew** (HAPI)
- **Sen Yang** (Rutgers)
- **Ida Sim** (UCSF, Open mHealth)
- **Aristotle Mannan** (Boswell)
- **Sona Vasudevan** (Georgetown School of Medicine)

**Contributed**
- pregnancy and birth modules

**Collaborating on**
- JAMIA Paper

**Building**
- Authoring Tools

**Submitted bug fixes and using data with apps**

**Synthea Grant application to NIH**

**Homeless and community health centers**

**Disease Model Validation**
Upcoming Engagements

- **FHIR Datathon Workshop @ AMIA's Translational Science Summit**
  - March 26 in San Francisco, CA

- **2017 Health Datapalooza**
  - April 27-28 at the Washington Hilton in Washington DC.
  - Presenting “Synthetic Patient Generation” as part of the “Patient Privacy Blockchain, Encryption, and Synthetic Data” session
Open Source Software Reuse

- **Health Data Standards**
  - Exporting C-CDA

- **Crucible**
  - FHIR libraries for Ruby
  - [https://github.com/fhir-crucible/fhir_client](https://github.com/fhir-crucible/fhir_client)

- **Standard Health Record (SHR)**
  - Patient Identification elements added as extensions
  - [https://github.com/standardhealth](https://github.com/standardhealth)

- **Intervention Engine**
  - Go FHIR server
  - [https://github.com/synthetichealth/gofhir](https://github.com/synthetichealth/gofhir)
Resources

- **Contact me:** Jason Walonoski
  - [jwalonoski@mitre.org](mailto:jwalonoski@mitre.org)

- **Synthea**
  - [https://github.com/synthetichealth/synthea](https://github.com/synthetichealth/synthea)

- **SyntheticMass**
  - Browse: [https://syntheticmass.mitre.org](https://syntheticmass.mitre.org)
  - FHIR: [https://syntheticmass.mitre.org/fhir/metadata](https://syntheticmass.mitre.org/fhir/metadata)

- **Fortnightly Community Teleconference**
  - [https://docs.google.com/document/d/1AabSpo8Nd2ynFH43C9amKgN68rIhRc-v8uN-nk2IrpU/edit?usp=sharing](https://docs.google.com/document/d/1AabSpo8Nd2ynFH43C9amKgN68rIhRc-v8uN-nk2IrpU/edit?usp=sharing)