**INTRO:**

- Using the OMOP CDM and OHDSI analytical tools provide a rigorous framework to generate evidence from EMR data.
- Synthetic patient data converted to the OMOP CDM can be used as a training platform of methods without patient privacy concerns.
- A goal was to evaluate synthetic EMR generator Synthea to see how well it performs as a space for testing and developing methods in the OHDSI framework.
- Conducted example analysis on synthetic EMR dataset of Richmond VA and designed a training platform.
- Introduce OHDSI using the training platform to VCU Biostatistics Department.

**METHODS:**

**Learning Platform Process**

- Create modules of cancer risk for esophageal and breast cancer and generate 1.3 million patients in Richmond VA.
- Run ETL to convert to OMOP.
- Load CEM data to query negative controls.
- Generate ATLAS compatible cohorts using CAPR.
- Set up Case-Control and SCCS analyses in R.
- Design a series of teaching module Notebooks to demonstrate end-to-end process of conducting on OHDSI study.
- Compare Synthea results to 5% sample of SYNPUF (n=116K).

**Study Design**

<table>
<thead>
<tr>
<th>Design</th>
<th>Cohorts</th>
<th>Analytical Choices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case-Control</td>
<td>T: Alendronate; Match gender and age (caliper: 2); Washout: 180 days; Nesting: Osteoporosis</td>
<td>5 cases per control; 68 drug negative controls from CEM.</td>
</tr>
<tr>
<td>Self-Controlled Case Series</td>
<td>T: Alendronate; Nesting: Osteoporosis</td>
<td>Naive Period: 180 days; 68 drug negative controls from CEM.</td>
</tr>
</tbody>
</table>

**Lessons**

- Goal was to prepare the learning platform process and create some example modules ready for introduction of the platform to VCU before the end of the year.
- This clinical example was challenging because cancer is a difficult concept to define and may be rare in a study population.
- Synthea requires development through addition of clinical modules that reflect real world practice and maintenance by validating and improving existing modules.

**RESULTS:**

- Using synthetic data for training in the OHDSI framework, we found our example study to show no association between oral bisphosphonates and composite cancer outcome.
- Empirical calibration shows both methods are prone to systematic bias of the database. Synthea negative controls were harder to find.
- SYNPUF 5% was a better synthetic dataset for answering the example question.

**Next Steps**

- Work with clinical experts to improve and validate Synthea modules.
- Continue to improve the Synthea ETL process.
- On-going process of Introducing OHDSI to VCU Biostatistics Department.

- Overview of Learning Platform in October.
- Discussion of Empirical Calibration using negative controls in November.
- VCU to provide feedback on learning platform.

---

**Design**

<table>
<thead>
<tr>
<th>Design</th>
<th>RR</th>
<th>Conf. Int</th>
<th>Log(RR)</th>
<th>Se log(RR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synpuf5pct</td>
<td>CC</td>
<td>1.094</td>
<td>(0.664, 1.718)</td>
<td>0.0899</td>
</tr>
<tr>
<td></td>
<td>SCCS</td>
<td>0.963</td>
<td>(0.586, 1.507)</td>
<td>-0.0378</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Design</th>
<th>RR</th>
<th>Conf. Int</th>
<th>Log(RR)</th>
<th>Se log(RR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SyntheaVA1M</td>
<td>CC</td>
<td>1.25</td>
<td>(0.358, 3.41)</td>
<td>0.223</td>
</tr>
<tr>
<td></td>
<td>SCCS</td>
<td>0.9433</td>
<td>(0.280, 2.24)</td>
<td>-0.558</td>
</tr>
</tbody>
</table>