Factors Influencing Background Incidence Rates: Systematic Empirical Evaluation Across an International Network of Observational Databases

Anna Ostropolets
Columbia University Medical Center
Acknowledgement

Thanks to all co-authors and collaborators:

Xintong Li, Rupa Makadia, Gowtham Rao, Peter R. Rijnbeek, Talita Duarte-Salles, Martijn Schuemie, Anthony G. Sena, Azza Shaoibi, Marc A. Suchard, Patrick B. Ryan, Daniel Prieto-Alhambra and George Hripcsak
Drug (vaccine) safety studies

- Observed rate of adverse event vs Expected rate of adverse event

  - FDA’s Adverse Event Reporting System
  - Adverse events following exposure in EHR and claims data
  - Background (baseline, historic) rates of adverse events in EHR and claims data
Background incidence rates

\[
\text{New cases} \quad \text{during specified time period} = \frac{\text{Background incidence rate}}{\text{Person-time of the at-risk population}}
\]

Background incidence rate

New cases

Time at risk (TAR)

Population

Estimates of baseline incidence of *stroke* in different studies: 4.6 – 679 per 100,000 person-years depending on the population, time-at-risk, data source.
Systematic experiment

15 adverse events (Brighton list)
2 new case identification strategies

2017-2020
8 age groups
3 condition groups
2 sexes
2 races

4 index dates
2 TAR starts
5 TAR durations
4 years
4 seasons

10 research questions, compute incidence rates and incidence rate ratios, random effect model meta-analysis
## Main findings

### Magnitude of influence

<table>
<thead>
<tr>
<th>Age</th>
<th>factor of 1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.g., pooled incidence rates of acute myocardial infarction in 6-17 yo is &lt;1 per 100,000 PY and ~1,330 per 100,000 in 85+ yo group</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data source</th>
<th>100</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Index date (anchoring)</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>* Date vs visit</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition subgroups*</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>* Patients with chronic conditions</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>2</th>
</tr>
</thead>
</table>

### Race | 2 |

### Season* | 1.5 |

### New case identification strategy (clean window) | 1.5 |

* Season and comparison of COVID-19 pandemic versus previous years
Selecting an **index date** in a cohort or case-crossover study

Exposed group/time

- Drug exposure date
- Vaccination date
- Procedure date

Unexposed group/time

- January 1st?
- A random date?
- A visit?
- Something else?
Anchoring on a visit increases incidence rates for short and medium time-at-risk intervals for all conditions

For short time-at-risk (0-1 day) anchoring on a visit is associated with up to a 100-fold increase in incidence when compared to anchoring on January 1st (IRR from meta-analysis 26.8 (95% CI 21.9-32.8)).

Comparison of anchoring on a random visit versus anchoring on January 1st in patients with a visit in the next year, incidence rate ratio.
# Main findings

## Magnitude of influence

<table>
<thead>
<tr>
<th>Factor</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (factor of 1,000)</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>2</td>
</tr>
<tr>
<td>Season*</td>
<td>1.5</td>
</tr>
<tr>
<td>New case identification strategy (clean window)</td>
<td>1.5</td>
</tr>
<tr>
<td>Index date (anchoring)</td>
<td>50</td>
</tr>
<tr>
<td>Condition subgroups*</td>
<td>4</td>
</tr>
<tr>
<td>Gender</td>
<td>2</td>
</tr>
</tbody>
</table>

E.g., pooled incidence rates of acute myocardial infarction in 6-17 yo is <1 per 100,000 PY and ~1,330 per 100,000 in 85+ yo group.
Key findings

• Age has the largest impact on incidence with incidence rates varying up to a factor of 1,000 across age groups.

• Anchoring has a great impact on incidence rates (and patient characteristics). Anchoring time-at-risk interval on any type of healthcare encounter yielded higher incidence when compared to anchoring on a random date, especially for the short time-at-risk.

• Temporal and seasonal trends, gender, race and clean window choice have moderate influence
Key implications

• As population characteristics have high impact on baseline rates, population used for background rates calculation should represent the population for observed rate calculations.

• As background rates don’t have a definitive point-in-time index date for time-at-risk interval, we must select an index date or event (anchor) that serves as a counterfactual for exposure (vaccination) based on the background knowledge and empirical assessment.