Standardization of Clinical Trials Data for Large-scale Incidence Calculations of Adverse Drug Events

OHDSI Community Call
July 16, 2024

Elise Ruan, MD, MPH
Clinical Trials Adverse Event Drug Data

- Much of knowledge regarding potential adverse drug events (ADE) comes from premarketing clinical trials.
- **Adverse event** is defined as any *unfavorable or unintended* sign, symptom, or disease *temporally* associated with the use of a drug, *without any judgement about causality* or relationship to the drug.
- Reported events (as well as most other data elements) are entered as unstructured text entries in the clinical trials database.
Objectives

1. Map clinical trials data elements to standardized concepts

- Drug Ingredient (RxNorm)
- Condition (SNOMED CT)
Objectives

1. Map clinical trials data elements to standardized concepts

2. Large-scale calculation of incidences for drug-condition pairs
## Objectives

1. Map clinical trials data elements to standardized concepts

   **Drug Ingredient (RxNorm)**

   **Condition (SNOMED CT)**

2. Large-scale calculation of incidences for drug-condition pairs

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Drug</th>
<th>Condition</th>
<th># Affected</th>
<th># At Risk</th>
<th>TAR</th>
<th>Time at Risk (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Objectives

1. Map clinical trials data elements to standardized concepts

2. Large-scale calculation of incidences for drug-condition pairs

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Drug</th>
<th>Condition</th>
<th># Affected</th>
<th># At Risk</th>
<th>TAR</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Drug</th>
<th>Condition</th>
<th>Pooled Incidence</th>
<th>Prediction Interval</th>
</tr>
</thead>
</table>
Objectives

1. Map clinical trials data elements to standardized concepts
2. Large-scale calculation of incidences for drug-condition pairs
3. Describe variance and interstudy heterogeneity of calculated incidences

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Drug</th>
<th>Condition</th>
<th># Affected</th>
<th># At Risk</th>
<th>TAR</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Drug</th>
<th>Condition</th>
<th>Pooled Incidence</th>
<th>Prediction Interval</th>
<th>Total Variance</th>
<th>Interstudy Heterogeneity</th>
</tr>
</thead>
</table>

Time at Risk (days)
Mapping of Drugs and Adverse Events

457,254 Trials 2000-6/2023 ———- 7,543 Trials ———- 4,098 Trials
- Had reported adverse event data
- Intervention type was drug
- FDA approved drug

- Had only 1 intervention
- Excluded result groups that contained “placebo”

What is AACT?

AACT is a publicly available relational database that contains all information (protocol and result data elements) about every study registered in ClinicalTrials.gov. Content is downloaded from ClinicalTrials.gov daily and loaded into AACT. The Clinical Trials Transformation Initiative (CTTI) enhanced AACT in October, 2016 to include the following features:

- Database content refreshed daily
- Database directly accessible in the cloud
- Static copies of the database available for download
- Open source tools freely available (postgresql, Ruby on Rails, Tableau Public)
- Source code available via Github
Mapping Results: Drug Concepts

All Trials (n= 7,543)

- 6,571 unique text entries
- 5,371 (82%) successfully mapped
- 1,200 (18%) not mapped
- 1,310 RxNorm (ingredient) concepts

Trials with 1 Intervention (n=4,098)

- 2,521 unique text entries
- 2,230 (88%) successfully mapped
- 291 (12%) not mapped
- 905 RxNorm (ingredient) concepts

Drug Ingredients with Most Trials

1. Cyclophosphamide
2. Pembrolizumab
3. Carboplatin
4. Dexamethasone
5. Paclitaxel
6. Nivolumab
7. Fludarabine
8. Cisplatin
9. Gemcitabine
10. Sodium Chloride
Mapping Results: Condition Concepts

**All Trials (n=7,543)**

- 29,769 unique text entries
- 25,006 (84%) successfully mapped
- 4,763 (16%) not mapped
- 9,134 SNOMED CT concepts

**Trials with 1 Intervention (n=4,098)**

- 18,922 unique text entries
- 16,339 (86%) successfully mapped
- 2,583 (14%) not mapped
- 7,411 SNOMED CT concepts

**Conditions with Most Trials**
1. Headache
2. Nausea
3. Diarrhea
4. Vomiting
5. Fatigue
6. Dizziness
7. Constipation
8. Fever
9. Abdominal Pain
10. Cough
### Data Schema

**interventions**
- **id**: int
- **nct_id**: string
- **intervention_type**: int
- **name**: string
- **description**: text

**reported_event**
- **id**: int
- **nct_id**: string
- **result_group_id**: int
- **ctgov_group_code**: string
- **time_frame**: text
- **event_type**: string
- **default_vocab**: string
- **default_assessment**: string
- **subjects_affected**: int
- **subjects_at_risk**: int
- **description**: text
- **event_count**: int
- **organ_system**: string
- **adverse_event_term**: string
- **frequency_threshold**: int
- **vocab**: string
- **assessment**: string

**time_frame**
- **time_frame**: text
- **tf_id**: int
- **tar_days**: double

**rep_ade**
- **ade_name**: text
- **ade_id**: int

**ade_codes**
- **sourcecode**: int
- **conceptid**: int
- **conceptname**: text

**drug_codes**
- **sourcecode**: int
- **conceptid**: int
- **conceptname**: text

Time frame mapped only for 1-intervention trials that had matched drug codes.
## Incidences & Prediction Intervals

<table>
<thead>
<tr>
<th>Drug</th>
<th>Condition</th>
<th>Trial</th>
<th>Reported</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tramadol</td>
<td>Headache</td>
<td>1</td>
<td>22</td>
<td>273</td>
</tr>
<tr>
<td>Tramadol</td>
<td>Nausea</td>
<td>1</td>
<td>57</td>
<td>273</td>
</tr>
<tr>
<td>Tramadol</td>
<td>Nausea</td>
<td>2</td>
<td>72</td>
<td>251</td>
</tr>
<tr>
<td>Dronabinol</td>
<td>Nausea</td>
<td>3</td>
<td>2</td>
<td>35</td>
</tr>
<tr>
<td>Dronabinol</td>
<td>Nausea</td>
<td>4</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>Prednisone</td>
<td>Heartburn</td>
<td>5</td>
<td>5</td>
<td>67</td>
</tr>
<tr>
<td>Prednisone</td>
<td>Heartburn</td>
<td>6</td>
<td>7</td>
<td>68</td>
</tr>
<tr>
<td>Drug</td>
<td>Condition</td>
<td>Trial</td>
<td>Reported</td>
<td>n</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------</td>
<td>-------</td>
<td>----------</td>
<td>----</td>
</tr>
<tr>
<td>Tramadol</td>
<td>Headache</td>
<td>1</td>
<td>22</td>
<td>273</td>
</tr>
<tr>
<td>Tramadol</td>
<td>Nausea</td>
<td>1</td>
<td>57</td>
<td>273</td>
</tr>
<tr>
<td>Tramadol</td>
<td>Nausea</td>
<td>2</td>
<td>72</td>
<td>251</td>
</tr>
<tr>
<td>Dronabinol</td>
<td>Nausea</td>
<td>3</td>
<td>2</td>
<td>35</td>
</tr>
<tr>
<td>Dronabinol</td>
<td>Nausea</td>
<td>4</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>Prednisone</td>
<td>Heartburn</td>
<td>5</td>
<td>5</td>
<td>67</td>
</tr>
<tr>
<td>Prednisone</td>
<td>Heartburn</td>
<td>6</td>
<td>7</td>
<td>68</td>
</tr>
</tbody>
</table>
## Incidences & Prediction Intervals

<table>
<thead>
<tr>
<th>Drug</th>
<th>Condition</th>
<th>Trial</th>
<th>Reported</th>
<th>n</th>
<th>Y</th>
<th>V</th>
<th>Incidence (Y): Reported/n</th>
<th>Variance (V): Y*(1-Y)/n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tramadol</td>
<td>Headache</td>
<td>1</td>
<td>22</td>
<td>273</td>
<td>0.08</td>
<td>0.0003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tramadol</td>
<td>Nausea</td>
<td>1</td>
<td>57</td>
<td>273</td>
<td>0.21</td>
<td>0.0006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tramadol</td>
<td>Nausea</td>
<td>2</td>
<td>72</td>
<td>251</td>
<td>0.29</td>
<td>0.0008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dronabinol</td>
<td>Nausea</td>
<td>3</td>
<td>2</td>
<td>35</td>
<td>0.06</td>
<td>0.0015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dronabinol</td>
<td>Nausea</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>0.50</td>
<td>0.0156</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prednisone</td>
<td>Heartburn</td>
<td>5</td>
<td>5</td>
<td>67</td>
<td>0.07</td>
<td>0.0010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prednisone</td>
<td>Heartburn</td>
<td>6</td>
<td>7</td>
<td>68</td>
<td>0.10</td>
<td>0.0013</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For pairs with only 1 study, just report incidence:

- **RMA* (Y, V)**

*For RMA used logit(Yi) for input and then inv.logit of pooled incidence and prediction intervals.*
# Incidences & Prediction Intervals

<table>
<thead>
<tr>
<th>drug_conceptid</th>
<th>drug_conceptname</th>
<th>ade_conceptid</th>
<th>ade_conceptname</th>
<th>pooled_incidence</th>
<th>predict_lower</th>
<th>predict_upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>701322</td>
<td>memantine</td>
<td>31967</td>
<td>Nausea</td>
<td>.12827</td>
<td>.0333714</td>
<td>.3854285</td>
</tr>
<tr>
<td>701322</td>
<td>memantine</td>
<td>75860</td>
<td>Constipation</td>
<td>.1641708</td>
<td>.0584904</td>
<td>.3830999</td>
</tr>
<tr>
<td>701322</td>
<td>memantine</td>
<td>81902</td>
<td>Urinary tract infectious disease</td>
<td>.0201333</td>
<td>.0108633</td>
<td>.0370178</td>
</tr>
<tr>
<td>701322</td>
<td>memantine</td>
<td>132797</td>
<td>Sepsis</td>
<td>.0276103</td>
<td>.0164179</td>
<td>.0460755</td>
</tr>
<tr>
<td>701322</td>
<td>memantine</td>
<td>133280</td>
<td>Alopecia</td>
<td>.0886829</td>
<td>.0123903</td>
<td>.4301427</td>
</tr>
<tr>
<td>701322</td>
<td>memantine</td>
<td>134736</td>
<td>Backache</td>
<td>.0911835</td>
<td>.0296742</td>
<td>.2476498</td>
</tr>
<tr>
<td>701322</td>
<td>memantine</td>
<td>135360</td>
<td>Syncope</td>
<td>.0065743</td>
<td>.0002753</td>
<td>.137233</td>
</tr>
<tr>
<td>701322</td>
<td>memantine</td>
<td>138525</td>
<td>Pain in limb</td>
<td>.0469607</td>
<td>.0081583</td>
<td>.2279072</td>
</tr>
<tr>
<td>701322</td>
<td>memantine</td>
<td>140214</td>
<td>Eruption</td>
<td>.0314289</td>
<td>.0091214</td>
<td>.1026403</td>
</tr>
<tr>
<td>701322</td>
<td>memantine</td>
<td>196523</td>
<td>Diarrhea</td>
<td>.0911387</td>
<td>.0189835</td>
<td>.3419538</td>
</tr>
<tr>
<td>701322</td>
<td>memantine</td>
<td>197672</td>
<td>Urinary incontinence</td>
<td>.006162</td>
<td>.0003358</td>
<td>.1026838</td>
</tr>
</tbody>
</table>
# Variance and Heterogeneity of Incidences

**RMA (logit(Yi), logit(Vi)) → on logit scale**

<table>
<thead>
<tr>
<th>Drug</th>
<th>Condition</th>
<th>Incidence (Pred Interval)</th>
<th>$\tau^2*$</th>
<th>$v_T^{**}$</th>
<th>Total Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tramadol</td>
<td>Nausea</td>
<td>-1.12 (-1.78, -0.46)</td>
<td>0.07</td>
<td>0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>Dronabinol</td>
<td>Nausea</td>
<td>-1.35 (-5.95, 3.25)</td>
<td>3.54</td>
<td>0.39</td>
<td>3.93</td>
</tr>
<tr>
<td>Prednisone</td>
<td>Heartburn</td>
<td>-2.31 (-2.91, -1.71)</td>
<td>0</td>
<td>0.19</td>
<td>0.19</td>
</tr>
</tbody>
</table>

$\tau^2 = \text{interstudy heterogeneity}$

$v_T = \text{typical sampling variance}$

Total Variance = $\tau^2 + v_T \rightarrow$ used for prediction interval

**Linear scale**

<table>
<thead>
<tr>
<th>Drug</th>
<th>Condition</th>
<th>Incidence (Pred Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tramadol</td>
<td>Nausea</td>
<td>0.25 (0.14, 0.39)</td>
</tr>
<tr>
<td>Dronabinol</td>
<td>Nausea</td>
<td>0.21 (0.003, 0.96)</td>
</tr>
<tr>
<td>Prednisone</td>
<td>Heartburn</td>
<td>0.09 (0.05, 0.15)</td>
</tr>
</tbody>
</table>

$\text{inv.logit()}$
## Variance and Heterogeneity - by Condition

### RMA (logit(Yi), logit(Vi)) → on logit scale

<table>
<thead>
<tr>
<th>Drug</th>
<th>Condition</th>
<th>$\tau^2*$</th>
<th>$\nu^*_T$</th>
<th>Total Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vincristine</td>
<td>Heartburn</td>
<td>0</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>Cyclophosphamide</td>
<td>Heartburn</td>
<td>0</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>Rituximab</td>
<td>Heartburn</td>
<td>0</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>Doxorubicin</td>
<td>Heartburn</td>
<td>0</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>Metformin</td>
<td>Heartburn</td>
<td>0.87</td>
<td>0.93</td>
<td>1.80</td>
</tr>
<tr>
<td>Prednisone</td>
<td>Heartburn</td>
<td>0</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>Lenalidomide</td>
<td>Heartburn</td>
<td>2.57</td>
<td>0.65</td>
<td>3.22</td>
</tr>
</tbody>
</table>

### Summary Table

<table>
<thead>
<tr>
<th>Condition</th>
<th>Median($\tau^2*$)</th>
<th>Median(TV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heartburn</td>
<td>0</td>
<td>0.19</td>
</tr>
<tr>
<td>Hemarthrosis</td>
<td>2.41</td>
<td>6.90</td>
</tr>
</tbody>
</table>
## Variance and Heterogeneity - Ex. Hemarthrosis

<table>
<thead>
<tr>
<th>Drug</th>
<th>Condition</th>
<th>(\tau^2)</th>
<th>(\nu_T)</th>
<th>Total Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitamin D3</td>
<td>Hemarthrosis</td>
<td>8.94</td>
<td>0.50</td>
<td>9.44</td>
</tr>
<tr>
<td>Empagliflozin</td>
<td>Hemarthrosis</td>
<td>3.36</td>
<td>1.00</td>
<td>4.37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition</th>
<th>Median((\tau^2))</th>
<th>Median((\nu_T))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemarthrosis</td>
<td>2.41</td>
<td>6.90</td>
</tr>
</tbody>
</table>
### Variance and Heterogeneity - Ex. Hemarthrosis

<table>
<thead>
<tr>
<th>Drug</th>
<th>Condition</th>
<th>Trial</th>
<th>Reported</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitamin D3</td>
<td>Hemarthrosis</td>
<td>1</td>
<td>2</td>
<td>22,374</td>
</tr>
<tr>
<td>Vitamin D3</td>
<td>Hemarthrosis</td>
<td>2</td>
<td>2</td>
<td>292</td>
</tr>
<tr>
<td>Empagliflozin</td>
<td>Hemarthrosis</td>
<td>3</td>
<td>1</td>
<td>157</td>
</tr>
<tr>
<td>Empagliflozin</td>
<td>Hemarthrosis</td>
<td>4</td>
<td>1</td>
<td>2,996</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Drug</th>
<th>Condition</th>
<th>tau2*</th>
<th>v_T**</th>
<th>Total Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitamin D3</td>
<td>Hemarthrosis</td>
<td>8.94</td>
<td>0.50</td>
<td>9.44</td>
</tr>
<tr>
<td>Empagliflozin</td>
<td>Hemarthrosis</td>
<td>3.36</td>
<td>1.00</td>
<td>4.37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition</th>
<th>Median(tau2)</th>
<th>Median(TV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemarthrosis</td>
<td>2.41</td>
<td>6.90</td>
</tr>
</tbody>
</table>
## Variance and Heterogeneity - by Drug

RMA (logit(Yi), logit(Vi)) → on logit scale

<table>
<thead>
<tr>
<th>Drug</th>
<th>Condition</th>
<th>tau²*</th>
<th>v₇**</th>
<th>Total Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tramadol</td>
<td>Nausea</td>
<td>0.07</td>
<td>0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>Tramadol</td>
<td>Constipation</td>
<td>0</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>Tramadol</td>
<td>Vomiting</td>
<td>0.15</td>
<td>0.03</td>
<td>0.18</td>
</tr>
<tr>
<td>Tramadol</td>
<td>Hypoxia</td>
<td>0.79</td>
<td>0.13</td>
<td>0.92</td>
</tr>
<tr>
<td>Tramadol</td>
<td>Injection Site Pain</td>
<td>0</td>
<td>0.17</td>
<td>0.07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Drug</th>
<th>Median(tau²)</th>
<th>Median(TV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tramadol</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>Dronabinol</td>
<td>8.77</td>
<td>9.53</td>
</tr>
</tbody>
</table>
### Variance and Heterogeneity - by Drug

RMA (logit(Yi), logit(Vi)) → on logit scale

<table>
<thead>
<tr>
<th>Drug</th>
<th>Condition</th>
<th>tau²*</th>
<th>v_T**</th>
<th>Total Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tramadol</td>
<td>Nausea</td>
<td>0.07</td>
<td>0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>Tramadol</td>
<td>Constipation</td>
<td>0</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>Tramadol</td>
<td>Vomiting</td>
<td>0.15</td>
<td>0.03</td>
<td>0.18</td>
</tr>
<tr>
<td>Tramadol</td>
<td>Hypoxia</td>
<td>0.79</td>
<td>0.13</td>
<td>0.92</td>
</tr>
<tr>
<td>Tramadol</td>
<td>Injection Site  Pain</td>
<td>0</td>
<td>0.17</td>
<td>0.07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Drug</th>
<th>Median(tau²)</th>
<th>Median(TV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tramadol</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>Dronabinol</td>
<td>8.77</td>
<td>9.53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Drug</th>
<th>Condition</th>
<th>Trial</th>
<th>Reported</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dronabinol</td>
<td>Nausea</td>
<td>3</td>
<td>2</td>
<td>35</td>
</tr>
<tr>
<td>Dronabinol</td>
<td>Nausea</td>
<td>4</td>
<td>8</td>
<td>16</td>
</tr>
</tbody>
</table>
Challenges & Limitations

- Low proportion of included studies
- Low number of studies per pair
- Combination drugs - for heartburn - 2 trials for R-CHOP - artificially brings down median

<table>
<thead>
<tr>
<th>Drug</th>
<th>Condition</th>
<th>tau2*</th>
<th>$\nu_T$**</th>
<th>Total Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vincristine</td>
<td>Heartburn</td>
<td>0</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>Cyclophosphamide</td>
<td>Heartburn</td>
<td>0</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>Rituximab</td>
<td>Heartburn</td>
<td>0</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>Doxorubicin</td>
<td>Heartburn</td>
<td>0</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>Metformin</td>
<td>Heartburn</td>
<td>0.87</td>
<td>0.93</td>
<td>1.80</td>
</tr>
<tr>
<td>Prednisone</td>
<td>Heartburn</td>
<td>0</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>Lenalidomide</td>
<td>Heartburn</td>
<td>2.57</td>
<td>0.65</td>
<td>3.22</td>
</tr>
</tbody>
</table>
Conclusions & Next Steps

- Majority (>80%) of standardized text entries can be mapped to a standardized concept in the OMOP CDM
- Standardizing how clinical trials data is entered would help support large-scale calculations
- Challenging to evaluate heterogeneity/variance in results

Next Steps:

- Compare with HowOften results
  - Pooled Incidences & Prediction Intervals
  - Variance & Heterogeneity

Thank you to George Hripcsak, Cindy Chen, Anna Ostropolets, Patrick Ryan, Seung In Seo