Tutorial

authors:
Rahel Bröhan
Marie Steglich
Ruwen Böhm <ruwen.boehm@pharmakologie.uni-kiel.de>
Hans-Joachim Klein <hjk@is.informatik.uni-kiel.de>

version: 2015-12-07
1. Introduction

OpenVigil 2.0 ([http://www.is.informatik.uni-kiel.de:8503/OpenVigil/](http://www.is.informatik.uni-kiel.de:8503/OpenVigil/)) is a pharmacovigilance data analysis tool. It extends OpenVigil 1 ([http://www.uni-kiel.de/pharmacology/pvt/openvigil.php/](http://www.uni-kiel.de/pharmacology/pvt/openvigil.php/)) which is still maintained for exploring the raw data. Since OpenVigil 2 – unlike OpenVigil 1 – operates on cleaned data, it is the first choice for pharmacovigilance analyses.

The data currently used in OpenVigil 2.0 are taken from Adverse Event Reporting System (AERS) of the Food and Drug Administration (FDA) of the USA and – with respect to information on drugs – from Drugbank (drugbank.ca) and Drugs@FDA.

The advantage of the FDA source is a large amount of data due to the size of the reporting population. The disadvantage is that reports of AERS are often incomplete (e.g., missing patient demographic data) or wrong (e.g., non-professional reporter or biased reporting, see the OpenVigil cave-at documents). Nevertheless this data source can be used to generate hypotheses instead of conducting clinical trials which might be difficult to realize (e.g., the adverse event is very rare).

OpenVigil 2.0 is a data analysis tool which extracts, filters and analyses pharmacovigilance data (e.g., AERS) by different criteria. The following examples of the tutorial illustrate which queries can be realised by using OpenVigil 2.0.

---

1 Cave-at documents:

OpenVigil 1: [http://www.uni-kiel.de/pharmacology/pvt/caveat.html](http://www.uni-kiel.de/pharmacology/pvt/caveat.html)
2. Definitions

2.1. Pharmacovigilance

Pharmacovigilance is the science of drug safety. The observation of pharmaceutical products after the clinical trials leading to marketing authorization and the collection, monitoring and prevention of adverse effects belongs to this science. \(^1\)

In most jurisdictions it is mandatory for physicians, pharmacists and pharmaceutical companies to report adverse events.

2.2. “Drug” (as used by OpenVigil)

OpenVigil uses the term “drug” for a substance in a pharmaceutical product that is biologically active and responsible for the therapeutic effect. “Drug” must not be confused with other meanings like illicit drugs or a ready-made pharmaceutical product like a pill (see below).

Because OpenVigil was initially designed for the U.S. American pharmacovigilance data, drugs are named according to the U.S. Adopted Name (USAN) scheme. This differs from International Nonproprietary Name (INN):

<table>
<thead>
<tr>
<th>International Nonproprietary Name (INN)</th>
<th>U.S. Adopted Name (USAN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>glibenclamide</td>
<td>glyburide</td>
</tr>
<tr>
<td>acetylsalicylic acid</td>
<td>aspirin</td>
</tr>
<tr>
<td>metamizole</td>
<td>dipyrone</td>
</tr>
<tr>
<td>salbutamol</td>
<td>albuterol</td>
</tr>
<tr>
<td>paracetamol</td>
<td>acetaminophen</td>
</tr>
<tr>
<td>rifampicin</td>
<td>rifampin</td>
</tr>
<tr>
<td>suxamethonium</td>
<td>succinylcholine</td>
</tr>
<tr>
<td>glyceryl trinitrate</td>
<td>nitroglycerin</td>
</tr>
</tbody>
</table>

Since OpenVigil relies on external databases for mapping the drugnames to USAN, there is a risk of mismappings.

Note that there are also other drugnames like the British Adopted Name (BAN) which exist in the raw FDA data. BAN allows combining two drugs into one “drugname”, e.g., cotrimoxazole as a combination of trimethoprim and sulfamethoxazole.

---

\(^1\) http://en.wikipedia.org/wiki/Pharmacovigilance
2.3. “Pharmaproduct” (as used by OpenVigil)

OpenVigil uses “pharmaproduct” as notion for pharmaceutical products like a pill or liquid forms like a suspension or solution for injection which contains a **drug**(s) and **excipient**(s). Synonyms of the term “pharmaproduct” are thus

- medicine,
- medication,
- medicinal product,
- brand,
- brand name and
- pharmaceutical product.

To achieve correct results with OpenVigil 2.0 it is important to differentiate between the term “pharmaproduct” and the often colloquially synonymously used term “drug”.

2.4. Adverse event (AE) and Adverse drug reaction (ADR)

An **adverse event** (AE) is an event which occurs after the use of a pharmaceutical product. This does not automatically reflect a causal relationship. However, statistical, biological or clinical analysis of this association might reveal such a causal relationship. In this case it is called **adverse drug reaction** (ADR).

2.5. Structured Query Language (SQL)

The Structured Query Language (SQL) is used by OpenVigil to retrieve a certain dataset from a large database, e.g.

```
SELECT * FROM report LIMIT 10;
# get the first 10 reports from the REPORT table (=demographic data)
```

```
SELECT drugusage.route,COUNT(drugusage.route) FROM drugusage WHERE  
  drugusage.brandname='enbrel' GROUP BY drugusage.route  
# count which route of administration of the pharmaproduct “Enbrel®” was applied
```

As you can see, SQL is a domain specific language designed for storing, retrieving and modifying data in a relational database managed by a relational database management system (RDBMS). OpenVigil uses a SQL database to store the pharmacovigilance data. For complex queries which cannot be sufficiently phrased using the available graphical user interfaces (GUI), a generic SQL interface was added. Additionally, when using the GUI in OpenVigil 2.0 to construct a query, pressing the button “Show Query” will show the SQL query code(s) which resulted from your query. You can use this code to build a more complex query on top of it.

---

3. Examples

3.1. Individual Safety Reports (ISR)

**Problem:** Show all individual safety reports for a new drug (azilsartan medoxomil).

**Query construction:** Choose “drug” in “OpenVigil Search”; drugname is “azilsartan medoxomil”.

**Result:** A list of all reports; each single report can be accessed by clicking on the link in the ISR column.
Single Report:

In the ISR above some data (for example age, gender and weight of the patient) are missing.

In contrast to OpenVigil 1 (www.uni-kiel.de/pharmacology/pvt/) OpenVigil 2 filters ambiguous reports that contain misspelled names of drugs and pharmaproducts if they could not be corrected by using drug-databases (Drugbank, Drugs@FDA). Furthermore, OpenVigil 2 converts some attribute values like age, drug dosage and duration of therapy from free-text into a uniform format.
3.2. Interpretation of statistics used in OpenVigil 2.0

Problem: Is drug abuse an adverse reaction of loperamide?

Query construction: Choose “drug” in “OpenVigil Search”; drugname is “loperamide”; adverse event is “drug abuse”; data presentation and statistics are “Frequentist methods” (i.e., calculate a contingency table and various observed/expected ratios like PRR); choose an output format (e.g., HTML):

OpenVigil 2.0 counts the number of unique ISRs and not the number of patients (several ISRs can be connected to a single patient) nor the number of drug-usages.
The Chi-Squared value estimates whether observed values in this table differ from expected ones: A Chi-Square of 5 for a degree of freedom of 1 (= 2x2 table) tells us that the difference shown by the PRR exists with a probability of 97.5%.  

The other numbers are observed/expected-ratios:

The PRR (Proportional Reporting Ratio) in this case is 2.077. This tells us that drug abuse occurs twice as frequently for loperamide compared to all other drugs.

The ROR (Reports Odds Ratio) is 2.081, which means that the odds for drug abuse in case of using loperamide is twice the odds than for all other drugs. The lower bound of the confidence interval is 1.182; the upper bound is 3.664 (with a confidence level of 95% the true ROR value is in this confidence interval). Since the lower bound is > 1, we can assume with more than 95% probability that there is a disproportionality.

Details for observed/expected ratios like PRR and ROR can be found in the disproportionality analysis primer on the OpenVigil 2 website.

The result of this example might refer to the use of loperamide as an illicit drug. Loperamide is able to cross the blood-brain barrier but is normally immediately pumped out again by the p-glycoprotein (=ABCB1, MDR1). If loperamide is taken in combination with substances that inhibit p-glycoprotein like quinidine, loperamide has effects on the central nervous system.

Another explanation for the result is that loperamide is a drug used against diarrhoea. Drug addicts are often medicated with loperamide to prevent the diarrhoea which is a consequence of the drug withdrawal. People might have reported wrong data concerning loperamide to the AERS. For example, adverse event and indication might have been switched: Drug abuse is the reason why loperamide is used and not the consequence.

---

4 http://math.hws.edu/javamath/ryan/ChiSquare.html
5 https://people.richland.edu/james/lecture/m170/tbl-chi.html
6 http://en.wikipedia.org/wiki/Odds_ratio
8 http://en.wikipedia.org/wiki/Loperamide
3.3. **Query construction for the most reported adverse event connected to a drug/pharmaprodut**

**Problem:** What are the most reported adverse events connected to the drug amiodarone?

**Query construction:** Choose “drug“ in “OpenVigil Search“; drugname is “amiodarone“; no raw data shall be reported but a list of occurrences of each adverse event.

**Result:**

Most reported adverse event is “drug interaction“.
An explanation of the result might be that amiodarone inhibits a drug-metabolizing cytochrome P450 enzyme, isoform 3A4 (CYP3A4). Many drugs are metabolised by CYP3A4. An inhibition of CYP3A4 consequently increases the bioavailability of those drugs.

Remember that these are just raw counts that have to be normalized to other drugs (e.g., by using PRR, see example 5, or by using drug utilization data).
3.4. Query construction for a specific time interval

Problem: How many hypoglycaemic adverse events are reported for glibenclamide (USAN glyburide) in the year 2008? How many adverse events are reported in total?

Query construction: Choose “drug” in “OpenVigil Search”; drugname is “glyburide”; use the “Advanced search” to define the reporting date to the FDA (in this case the reporting date shall be within 2008); data presentation and statistics are “Frequency”. Output format is “Excel CSV” for further analysis and visualisation in a spreadsheet program.

Result: An Excel document with two columns – name and count of the events.

There are 93 ISRs with the adverse event “hypoglycaemia” reported for glibenclamide. 7009 adverse events have been reported in total.
3.5. Proportional Reporting Ratio (PRR) analysis of a drug or pharmaprouct

Problem: How likely is it that the reported adverse events are truly adverse drug reactions specific to the drug amiodarone?

Query construction: Choose “drug” in “OpenVigil Search”; drugname is “amiodarone”; data presentation and statistics are “Frequentist methods”. OpenVigil will compute and show a table with various values like measurements of disproportionality. As output format “Excel CSV” is chosen for further analysis and visualisation in a spreadsheet program.

NB: This calculation might take some time!

Result:

Excel CSV file imported into Excel:

Cave: If you cannot properly import numbers to your spreadsheet software, this might be due to the different symbols used for decimal marks. OpenVigil uses the U.S. american symbols, i.e., a point represents a decimal mark. For further information see:  
Use the columns “prr” and “chi-square” to create a graph: x-axis title is “PRR”; y-axis title is “Chi-Square”.

Changing the scale of both axes to logarithmic gives the final PRR graph:

The upper-right quadrant contains putative adverse drug reactions. Everything else is just an adverse event.

In the result list “drug interaction” (cf. example above) is reported with a PRR of 8.151 and a Chi-Squared value of 4042. Due to this drug interaction is very likely an adverse drug reaction of amiodarone.

However, prior knowledge of this CYP3A4 inhibition by amiodaron will influence reporting of these cases and thus skew the results.
3.6. Reverse PRR analysis of an adverse event

Problem: For which pharmaproducts/drugs is agranulocytosis reported as an adverse drug reaction?

Query construction: Adverse reaction is “agranulocytosis”; data presentation and statistics are “Frequentist methods” (Reverse PRR analysis of the adverse event “agranulocytosis”). “Excel CSV” is chosen as output format for further analysis and visualisation in a spreadsheet program.

Results:

The resulting list contains names of drugs and pharmaproducts.
Create a PRR graph like in the example above:

The upper-right quadrant contains drugs that likely have agranulocytosis as an adverse drug reaction, for example pirenzepine, a drug used in treatment of peptic ulcer\(^9\): PRR 178.020285; Chi-Square: 3086.672747; Pirenzepine is shown in the result list with 21 occurrences for agranulocytosis.

You can also choose “HTML” as output format of the query result. The query result is shown in a new window of the browser:

Tip: The result list can be sorted according to the values in a column by clicking on the arrows in the corresponding column header (for example data can be sorted in ascending order.)

In addition to this, the list can be sorted by two criteria (like for example rPRR in descending order and Chi-Squared value in ascending order) by holding down the shift key and clicking on a second arrow:

- **Drug**: The name of the drug.
- **Is ADHD**: Indicates if the drug is associated with ADHD.
- **rPRR**: The rPRR value.
- **Chi-square**: The Chi-square value.
- **This AE This Drug**: The AE count in this drug.
- **This AE Other Drugs**: The AE count in other drugs.
- **This AE Total**: The total AE count.
- **Other AE This Drug**: The other AE count in this drug.
- **Other AE Other Drugs**: The other AE count in other drugs.
- **Other AE Total**: The total other AE count.
- **All AE This Drug**: The total AE count in this drug.
- **All AE Other Drugs**: The total AE count in other drugs.
- **All AE Total**: The total AE count.

### Table

<table>
<thead>
<tr>
<th>Drug</th>
<th>Is ADHD</th>
<th>rPRR</th>
<th>Chi-sq</th>
<th>This AE This Drug</th>
<th>This AE Other Drugs</th>
<th>This AE Total</th>
<th>Other AE This Drug</th>
<th>Other AE Other Drugs</th>
<th>Other AE Total</th>
<th>All AE This Drug</th>
<th>All AE Other Drugs</th>
<th>All AE Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>delcore</td>
<td>Yes</td>
<td>97.161255</td>
<td>183.290398</td>
<td>3</td>
<td>2354</td>
<td>2357</td>
<td>39</td>
<td>2977079</td>
<td>2977118</td>
<td>42</td>
<td>2979433</td>
<td>29794</td>
</tr>
<tr>
<td>efexta</td>
<td>No</td>
<td>90.231165</td>
<td>20.09615</td>
<td>1</td>
<td>2356</td>
<td>2357</td>
<td>14</td>
<td>2977104</td>
<td>2977118</td>
<td>15</td>
<td>2979460</td>
<td>29794</td>
</tr>
<tr>
<td>clozoxin</td>
<td>Yes</td>
<td>9.994007</td>
<td>95.711271</td>
<td>13</td>
<td>2344</td>
<td>2357</td>
<td>1043</td>
<td>2973473</td>
<td>2977118</td>
<td>150</td>
<td>2977819</td>
<td>29794</td>
</tr>
<tr>
<td>promazine</td>
<td>No</td>
<td>9.791444</td>
<td>1.535077</td>
<td>1</td>
<td>2356</td>
<td>2357</td>
<td>129</td>
<td>2976989</td>
<td>2977118</td>
<td>130</td>
<td>2979345</td>
<td>29794</td>
</tr>
<tr>
<td>polaxamine</td>
<td>Yes</td>
<td>9.701201</td>
<td>30.571125</td>
<td>5</td>
<td>2352</td>
<td>2357</td>
<td>651</td>
<td>2976467</td>
<td>2977118</td>
<td>656</td>
<td>2978819</td>
<td>29794</td>
</tr>
<tr>
<td>haloperidol</td>
<td>Yes</td>
<td>9.666553</td>
<td>15.328158</td>
<td>3</td>
<td>2354</td>
<td>2357</td>
<td>392</td>
<td>2976726</td>
<td>2977118</td>
<td>395</td>
<td>2979080</td>
<td>29794</td>
</tr>
<tr>
<td>azoxyn</td>
<td>Yes</td>
<td>9.562326</td>
<td>75.429051</td>
<td>11</td>
<td>2340</td>
<td>2357</td>
<td>1433</td>
<td>2975663</td>
<td>2977118</td>
<td>1404</td>
<td>2978011</td>
<td>29794</td>
</tr>
<tr>
<td>adionoxyl pfs</td>
<td>Yes</td>
<td>9.47073</td>
<td>134.703706</td>
<td>19</td>
<td>2338</td>
<td>2357</td>
<td>2534</td>
<td>2974584</td>
<td>2977118</td>
<td>2553</td>
<td>2976922</td>
<td>29794</td>
</tr>
<tr>
<td>clarimazole</td>
<td>Yes</td>
<td>9.426092</td>
<td>51.753403</td>
<td>9</td>
<td>2340</td>
<td>2357</td>
<td>1072</td>
<td>2976046</td>
<td>2977118</td>
<td>1080</td>
<td>2978395</td>
<td>29794</td>
</tr>
<tr>
<td>nitro</td>
<td>No</td>
<td>9.420902</td>
<td>7.754556</td>
<td>2</td>
<td>2355</td>
<td>2357</td>
<td>268</td>
<td>2976830</td>
<td>2977118</td>
<td>270</td>
<td>2979205</td>
<td>29794</td>
</tr>
<tr>
<td>flagyl</td>
<td>Yes</td>
<td>9.201872</td>
<td>131.737911</td>
<td>19</td>
<td>2338</td>
<td>2357</td>
<td>2580</td>
<td>2974338</td>
<td>2977118</td>
<td>2599</td>
<td>2976876</td>
<td>29794</td>
</tr>
<tr>
<td>l-cysteine</td>
<td>Yes</td>
<td>9.287473</td>
<td>16.561537</td>
<td>3</td>
<td>2354</td>
<td>2357</td>
<td>408</td>
<td>2976710</td>
<td>2977118</td>
<td>411</td>
<td>2979064</td>
<td>29794</td>
</tr>
<tr>
<td>rifampin</td>
<td>Yes</td>
<td>9.226095</td>
<td>144.984481</td>
<td>21</td>
<td>2330</td>
<td>2357</td>
<td>2875</td>
<td>2974243</td>
<td>2977118</td>
<td>2896</td>
<td>2975579</td>
<td>29794</td>
</tr>
</tbody>
</table>
3.7. Query construction for different adverse events

Problem: What are the two most reported pharmaproducts with gastrointestinal haemorrhage as an adverse event?

Query construction: Choose “pharmaproduct” in “OpenVigil Search”; adverse events are “gastrointestinal haemorrhage”, “lower gastrointestinal haemorrhage”, “upper gastrointestinal haemorrhage” and “gastrointestinal ulcer haemorrhage”. Use the plus button to add more textfields. These conditions can be connected with operators (AND; all conditions met; OR: at least one condition met; XOR: exactly one condition met). Data presentation and statistics is “Frequency”. Output format of the query result is HTML.

Result:

<table>
<thead>
<tr>
<th>Occurrence</th>
<th>Type</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>2235</td>
<td>drug</td>
<td>metylacetic acid</td>
</tr>
<tr>
<td>1966</td>
<td>product</td>
<td>aspirin</td>
</tr>
<tr>
<td>1930</td>
<td>drug</td>
<td>dabigatran etexilate</td>
</tr>
<tr>
<td>1904</td>
<td>product</td>
<td>pradaxa</td>
</tr>
<tr>
<td>1579</td>
<td>drug</td>
<td>warfarin</td>
</tr>
<tr>
<td>949</td>
<td>drug</td>
<td>rivaroxib</td>
</tr>
<tr>
<td>951</td>
<td>product</td>
<td>ximox</td>
</tr>
<tr>
<td>913</td>
<td>drug</td>
<td>clopidogrel</td>
</tr>
<tr>
<td>776</td>
<td>drug</td>
<td>acetaminophen</td>
</tr>
<tr>
<td>754</td>
<td>drug</td>
<td>furosemide</td>
</tr>
<tr>
<td>678</td>
<td>product</td>
<td>pravix</td>
</tr>
<tr>
<td>674</td>
<td>drug</td>
<td>ibuprofen</td>
</tr>
<tr>
<td>672</td>
<td>product</td>
<td>coumadin</td>
</tr>
<tr>
<td>580</td>
<td>drug</td>
<td>simvastatin</td>
</tr>
<tr>
<td>544</td>
<td>drug</td>
<td>olmesapazole</td>
</tr>
<tr>
<td>544</td>
<td>drug</td>
<td>metoprolol</td>
</tr>
<tr>
<td>529</td>
<td>drug</td>
<td>oloroxib</td>
</tr>
<tr>
<td>519</td>
<td>drug</td>
<td>prednisone</td>
</tr>
</tbody>
</table>
The two most reported pharmaproducts with gastrointestinal haemorrhage as an adverse event are aspirin and pradaxa.
3.8. Structure Query Language (SQL)

Problem: The occurrence of gastrointestinal haemorrhage as an adverse event of the two most used acetylsalicylic acid-containing pharmaproducts shall be compared. A very complex query was constructed that cannot be created with the GUI of OpenVigil 2.0:

Query construction: The query can be written in SQL. A part of the database schema (full schema: see below this example) illustrates the query construction:

```
select count(drugusage.brandname), drugusage.brandname
from drugusage, pharmaproduct, product
where product.drugname = 'acetylsalicylic acid' and
      pharmaproduct.brandname = product.brandname and
      product.brandname = drugusage.brandname
group by drugusage.brandname
order by count(drugusage.brandname) desc
```
Results: Query result is a list with 31 pharma products (brand names).

<table>
<thead>
<tr>
<th>count</th>
<th>brandname</th>
</tr>
</thead>
<tbody>
<tr>
<td>73595</td>
<td>aspirin</td>
</tr>
<tr>
<td>647</td>
<td>ecotrin</td>
</tr>
<tr>
<td>455</td>
<td>florinal</td>
</tr>
<tr>
<td>453</td>
<td>bufferin</td>
</tr>
<tr>
<td>430</td>
<td>midol</td>
</tr>
<tr>
<td>164</td>
<td>coricidin</td>
</tr>
<tr>
<td>92</td>
<td>advil</td>
</tr>
<tr>
<td>65</td>
<td>asparin</td>
</tr>
<tr>
<td>60</td>
<td>8-hour bayer</td>
</tr>
<tr>
<td>58</td>
<td>anacin</td>
</tr>
<tr>
<td>50</td>
<td>no-actas aspirin</td>
</tr>
<tr>
<td>41</td>
<td>entrophan</td>
</tr>
<tr>
<td>27</td>
<td>norgesic</td>
</tr>
<tr>
<td>23</td>
<td>drisdan</td>
</tr>
<tr>
<td>20</td>
<td>solpria</td>
</tr>
<tr>
<td>16</td>
<td>no-actas</td>
</tr>
<tr>
<td>13</td>
<td>eurprin</td>
</tr>
<tr>
<td>13</td>
<td>aspren</td>
</tr>
<tr>
<td>8</td>
<td>acetosal</td>
</tr>
<tr>
<td>3</td>
<td>ratio-tocal</td>
</tr>
<tr>
<td>3</td>
<td>bayer extra strength aspirin for migraine pain</td>
</tr>
<tr>
<td>3</td>
<td>aceterine</td>
</tr>
<tr>
<td>2</td>
<td>acetol</td>
</tr>
<tr>
<td>1</td>
<td>colifar</td>
</tr>
<tr>
<td>1</td>
<td>emprin</td>
</tr>
</tbody>
</table>

For further analysis choose “Browse” and “Products” in OpenVigil 2.0:
Result is a list of pharmaceutical products ("pharmaproducts"): 

By clicking on a product, the drugs it consists of are shown:

In this example Bufferin® and Ecotrin® are compared to each other. Both pharmaproducts contain no other drugs except acetylsalicylic acid and appear to be used with a similar frequency, extrapolated from the number of reports in the database.
Choose “pharmaproduct” in “OpenVigil Search”; product name is “bufferin” (“ecotrin”); adverse event is “gastrointestinal haemorrhage”. Data presentation and statistics are “Frequentist methods”; output format of the query result is “HTML”.

Search results are two contingency tables:

Contingency table for Bufferin®:

<table>
<thead>
<tr>
<th></th>
<th>Drugs(s) of interest</th>
<th>All other drugs</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adverse event(s) of interest</td>
<td>6</td>
<td>12923</td>
<td>12929</td>
</tr>
<tr>
<td>All other adverse events</td>
<td>427</td>
<td>2966139</td>
<td>2966145</td>
</tr>
<tr>
<td>Σ</td>
<td>433</td>
<td>2979042</td>
<td>2979475</td>
</tr>
</tbody>
</table>

Chi-Square: 7.00987
PRR: 3 1.19289

ROR: 3.22514
ROR CI lower bound: 1.448907
ROR CI upper bound: 7.178878

Contingency table for Ecotrin®:
Comparing the PRR of bufferin (3.194307) and ecotrin (4.692033), it is obvious that gastrointestinal haemorrhage is very likely an adverse drug reaction to both pharmaproducts. Gastrointestinal haemorrhage occurs three times more frequently for bufferin than for all other drugs, while it occurs for ecotrin even four times more. The values for Chi-Squared confirm the results of the PRR (7.00987 for bufferin (the difference shown by the PRR exists with a probability of 99.995 %); 34.278924 for ecotrin).
The results of the two contingency tables can be merged in one table for further analysis (e.g., Fisher exact test, Chi-Squared test):

<table>
<thead>
<tr>
<th></th>
<th>Bufferin</th>
<th>Ecotrin</th>
<th>All other drugs</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gastrointestinal haemorrhage</td>
<td>6</td>
<td>13</td>
<td>12910</td>
<td>12929</td>
</tr>
<tr>
<td>All other adverse events</td>
<td>427</td>
<td>626</td>
<td>2965493</td>
<td>2966546</td>
</tr>
<tr>
<td>Σ</td>
<td>433</td>
<td>639</td>
<td>2978403</td>
<td>2979475</td>
</tr>
</tbody>
</table>
3.9. Compare OpenVigil 1 & 2 data (no. reports, PRR) to published data

Introduction: This example stresses the importance of carefully checking any results obtained. Common pitfalls are

- counting multiplicates,
- counting ambiguous reports and
- accidentally losing portion of the raw data.

These can happen at every time in the workflow. Therefore, it is important to know your data! Try different extraction conditions, check numbers for plausibility and browse result lists to manually screen the data.

Problem: Sakaeda et al. (Sakaeda T, Tamon A, Kadoyama K, Okuno Y. Data mining of the public version of the FDA Adverse Event Reporting System. Int. J. Med. Sci. 2013; 10(7):796-803. doi: 10.7150/ijms.6048, http://www.medsci.org/v10p0796.htm) report their results of data-mining AERS data from 2004 to 2009 for “warfarin” and other drugs and the adverse event “haematemesis” (see table below at the end of this example). The number of co-occurences (drug used, adverse event seen) was reported to be 268. A subsequent analysis of disproportionality did not reveal a statistical significant association.

Can we reproduce this data?

Query construction in OpenVigil 2: Enter “warfarin” as “drug” and “haematemesis” as adverse event, set the reporting date to between 2004 and 2009.

OpenVigil 2.0 can find 162 reports (out of 140 unique cases) and calculates – based on the counting of reports – a PRR of 3.109 and a ROR of 3.122. The latest OpenVigil 2.1 installation finds 166 reports (out of 143 unique cases) due to improved drugname mapping.

One first glance, both results appear way off: Too few reports and to few cases were found and the measurements of disproportionality indicate a rather strong association (i.e., haematemesis appears to be a real adverse reaction to warfarin). This in contrast to Sakaeda whose numbers do not fulfil Evans’ criteria (PRR > 2 for a signal, cf. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiol Drug Saf. 2001 Oct-Nov;10(6):483-6. http://www.ncbi.nlm.nih.gov/pubmed/11828828)

Discussion: OpenVigil 2 operates on cleaned and validated FDA data only. The drug “warfarin” is referred to in AERS data/marketed as

- warfarin
- Waran
- Jantoven
- Coumadin
- Lawarin
- Marevan
- Warfant
- coumarin derivative

and perhaps other names which we could not identify.
Drugs named something like “WARFARIN 5 MG” are currently discarded in OpenVigil 2 since the current version of OpenVigil 2 does not know what “5 MG” means. The misspelled “COUMADIN (WARFARIN SODIUM)” is not ambiguous for humans and should be mapped to warfarin, too. We are trying to improve that while at the same time keeping all drug-mapping unambiguous: Verbatim drugnames containing “BLIND” (like “BLINDED: WARFARIN SODIUM”) or ambiguous drug-names like “COUMADIN (CLOTRIMAZOLE)” must never be mapped to warfarin.

Finally, one has to decide whether “COUMARIN DERIVATE” should be included since drugs named like this or named “COUMARIN AND TROXERUTIN” or “ESBERIVEN (COUMARIN, HEPARIN SODIUM, MELILOT, RUTIN)” are probably not used to inhibit blood clotting and might contain no warfarin (a 4-hydroxy derivate if coumarin) at all.

The 162 cases in OpenVigil 2.0 are correct: You can look at the original free-text drugname and verify that only precise, unambiguous reports were considered.

However, OpenVigil 2.0 uses unique ISRs (162) for counting while unique CASEs (140) are probably the only reasonable way to count in this scenario. This mode of counting was added in OpenVigil 2.1.

Unfortunately, OpenVigil does currently not offer an automated check for multiplicates other than via CASE/ISR so the result list has to be screened manually.

**Raw data analysis – data importing and counting issues:**

Subsequently, we have also used GNU wc and OpenVigil 1 to explore the raw FDA AERS data and find out what Sakaeda might have been counting – because it’s not documented in the methods section of the publication: “Through an attempt to address these shortcomings, a novel system, named the CzeekV system, has been developed by Dr. Okuno in collaboration with Kyoto Constella Technologies Co., Ltd., Japan, “ (no source code provided) and “All drug names were unified into generic names by a text-mining approach, because FAERS permits the registering of arbitrary drug names, including trade names and abbreviations. Spelling errors were detected by a spell checker software, GNU Aspell, and carefully confirmed by working pharmacists.” (again no source code, and was really every free-text drugname looked at? we couldn’t do it!).

However, Sakaeda provides some numbers which we tried to check.

Sakaeda states that “the total number of reports used was 2,231,029”.

**AERS raw data is published quarterly. The lines in the DEMO AERS files from 2004Q1 to 2009Q4 were counted:**

```
wc DEMO0[4-9]*TXT
2234955
```

The result contains 24 header lines. Thus the real number of records is 2234931.
That’s 3,902 reports too much compared to Sakaeda. Some lines are discarded before importing them into SQL database due to syntax errors (i.e., wrong amount of items per line). The current importer of OpenVigil 1 just skips all non-matching data. The OpenVigil 2 import process provides an error correction mode and suggestions like merging two adjacent text lines. E.g., while OpenVigil 1 has discarded the two lines, OpenVigil 2 has merged them to one record. OpenVigil 1 stores these import failures in the database ([http://www.uni-kiel.de/pharmacology/pvt/openvigil.php?cd=if](http://www.uni-kiel.de/pharmacology/pvt/openvigil.php?cd=if)). However, the DEMO files in question had only one premature line break in DEMO09Q3 that results in two lines being discarded. So that’s still 3,901 to 3,900 reports more in the raw data compared to Sakaeda.

Within OpenVigil 2 there is currently no easy way to analyse certain data files only. Instead, we have to rely on date fields in the DEMO table that tell us whether a report falls into the period 2004 to 2009. Of note, future DEMO tables can contain reports from previous quarters. OpenVigil 1 offers the possibility to include only or exclude data from certain quarterly FDA AERS files.

DEMO contains 1,644,220 unique cases according to Sakaeda.

So we’ve counted total number of reports (containing duplicates), reports with unique ISR and reports with unique CASENO for the period where the time period is defined by either FDA_DT, MFR_DT or EVENT_DT for all data imported from DEMO04Q1 to DEMO09Q4 in OpenVigil 1:

```
SELECT COUNT(ISR), COUNT(DISTINCT ISR), COUNT(DISTINCT CASENO) FROM DEMO WHERE FDA_DT<"2009-12-31" AND FDA_DT>"2004-01-01" AND (DEMO.DSRC="DEMO04Q1" OR DEMO.DSRC="DEMO04Q2" OR DEMO.DSRC="DEMO04Q3" OR DEMO.DSRC="DEMO04Q4" OR DEMO.DSRC="DEMO05Q1" OR DEMO.DSRC="DEMO05Q2" OR DEMO.DSRC="DEMO05Q3" OR DEMO.DSRC="DEMO05Q4" OR DEMO.DSRC="DEMO06Q1" OR DEMO.DSRC="DEMO06Q2" OR DEMO.DSRC="DEMO06Q3" OR DEMO.DSRC="DEMO06Q4" OR DEMO.DSRC="DEMO07Q1" OR DEMO.DSRC="DEMO07Q2" OR DEMO.DSRC="DEMO07Q3" OR DEMO.DSRC="DEMO07Q4" OR DEMO.DSRC="DEMO08Q1" OR DEMO.DSRC="DEMO08Q2" OR DEMO.DSRC="DEMO08Q3" OR DEMO.DSRC="DEMO08Q4" OR DEMO.DSRC="DEMO09Q1" OR DEMO.DSRC="DEMO09Q2" OR DEMO.DSRC="DEMO09Q3" OR DEMO.DSRC="DEMO09Q4");
```

Out of curiosity, we have also counted all reports/cases minus the reports in the data files from 2004Q1 to 2005Q2 (see below for explanation).

<table>
<thead>
<tr>
<th>Data files and filtering</th>
<th>all reports</th>
<th>unique ISR</th>
<th>unique CASENO</th>
</tr>
</thead>
<tbody>
<tr>
<td>all files (2004-2012) and 2003-12-31 &gt; FDA_DT &lt; 2010-01-01</td>
<td>2234986</td>
<td>2231030</td>
<td>1645633</td>
</tr>
<tr>
<td>all reports in the quaterly files 2004-2009</td>
<td>2234929</td>
<td>2231036</td>
<td>1645605</td>
</tr>
<tr>
<td>only the quaterly files 2004-2009 and 2003-12-31 &gt; date &lt; 2010-01-01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FDA_DT</td>
<td>2234923</td>
<td>2231030</td>
<td>1645600</td>
</tr>
<tr>
<td>EVENT_DT</td>
<td>1655915</td>
<td>1653317</td>
<td>1184848</td>
</tr>
<tr>
<td>MFR_DT</td>
<td>2180288</td>
<td>2176768</td>
<td>1584290</td>
</tr>
<tr>
<td>FDA_DT minus data files DEMO04Q1 till DEMO05Q2</td>
<td>1805798</td>
<td>1803719</td>
<td>1331082</td>
</tr>
<tr>
<td>Sakaeda 2013</td>
<td>2231029</td>
<td>not provided</td>
<td>1644220</td>
</tr>
<tr>
<td>raw line count (minus headers)</td>
<td>2234931</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

These number differ, reflecting

- incomplete records (only ~ 70% of reports include the date of the event, EVENT_DT),
• numerous updates on cases (in ~5% of reports, an old ISR was reused, only at most ~70% of reports are unique cases) and
• data malformation (the total number of reports is different when comparing raw FDA data to the amount of data successfully imported into either OpenVigil 1 or Sakaeda’s system).

First raw data analysis in OpenVigil 1 using the GUI:
We have selected the professional wizard mode and entered “haematemesis” as adverse event (REAC.PT) and requested the reporting date to be within 2004 to 2009 (DEMO.FDA_DT). The above mentioned drugname, brandnames and other synonyms were subsequently used as part of the drugname (DRUG.DRUGNAME contains) and data was counted.
When we did this initially (see below concerning the problem we found) we counted these numbers:

Warfarin 148, Waran 3, Jantoven 1, Coumadin 109 (originally 110, but manual inspection of the list shows one overlap to warfarin since “WARFARIN 2.5 MG COUMADIN” was reported), Marevan 7 adding up to 268.
Thus, on first glance, we have found exactly as many “co-occurences” as Sakaeda.

Calculating the PRR is not automatically possible in OpenVigil 1.2.6 since the total number of reports containing one of the above listed terms needs to be added up while avoiding double counting.

SQL query construction in OpenVigil 1: We use the SQL code that was generated by the query above and fine-tune it to

```
SELECT DRUG.DRUGNAME, COUNT(DMO.ISR), COUNT(DISTINCT DMO.ISR), COUNT(DISTINCT DMO.CASENO) FROM DRUG,REAC,DMO WHERE ((DRUG.DRUGNAME LIKE "%WARAN%" OR DRUG.DRUGNAME LIKE "%WARFARIN%" OR DRUG.DRUGNAME LIKE "%COUMADIN%" OR DRUG.DRUGNAME LIKE "%JANTOVEN%" OR DRUG.DRUGNAME LIKE "%MAREVAN%") AND REAC.PT="HAEMATEMESIS" AND DMO.FDA_DT >= "2004-01-01" AND DMO.FDA_DT <= "2009-12-31") AND DRUG.ISR=REAC.ISR AND DRUG.ISR=DMO.ISR GROUP BY DRUG.DRUGNAME DESC;
```

The result is a list of ISRs and CASEs containing grouped by the different drugnames, adding up to 268 reports of which 256 have a unique ISR of which 212 have a unique CASENO:
Therefore, only 212 unique patients for warfarin (and generic) and the adverse event haematemesis appear to exist – but re-performing the query without grouping (no “GROUP BY DRUG.DRUGNAME DESC”) shows even less, just 202 distinct cases:

Obviously, some patients were on more than just one warfarin-containing drug and were thus listed several times in the output shown above.

The next step was to inspect the raw data to find any oddities:

It became apparent that no reports in 2004 and 2005 januar-june were included in this list. How could that be? We realized that the DEMO data prior to 2005Q3 were not imported properly into OpenVigil 1.2.3 at the time of the above presented analyses due to a change in the FDA data format in one data table. Re-performing the analysis with these data yields more reports (and cases):
There appear to be 413 reports from 299 distinct cases.

Hint: You can emulate losing data prior to 2005Q3 in OpenVigil 1 by adding

```sql
AND (DEMO.DSRC!="DEMO04Q1" AND DEMO.DSRC!="DEMO04Q2" AND DEMO.DSRC!="DEMO04Q3" AND DEMO.DSRC!="DEMO04Q4" AND DEMO.DSRC!="DEMO05Q1" AND DEMO.DSRC!="DEMO05Q2")
```

to the WHERE clause your SQL query like we did to obtain the screenshots above in spite of now using the complete dataset.

It is always important to look at the raw data before trusting any automated countings:

This resulting list has ideally to be completely scanned for multiplicates. E.g., we found the reports #5503640 and #5502179 which were both linked to different CASENO but have otherwise identical demographic data including date of death. Another example is #5064922 and #5655430. More examples might be there but we have not yet established a fast protocol to detect multiplicates. However, extrapolating from our findings here, we estimate that less than 1% are multiplicates.

Similar, one would need to run the above query without the adverse event and a third time with the adverse event but without the drugs to populate the 2x2 contingency table for disproportionality.
analysis. Before these numbers can be trusted, duplicates have to be eliminated (e.g., case 4004520 and 3909737 appear to be the same). Furthermore, the dataset in question has records like "[THERAPY UNSPECIFIED]" (76 records), "." (16 records) or "1 CONCOMITANT DRUG" (14 records) are impossible to map to a drugname and thus need a pre-defined way of dealing with. We’ll leave this as exercise to the reader. ;-) 

Results and comparison with Sakaeda 2013:

<table>
<thead>
<tr>
<th>Source</th>
<th>n (reports)</th>
<th>n (cases)</th>
<th>PRR</th>
<th>ROR (95%-CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenVigil 1 GUI without DEMO data prior to 2005Q3</td>
<td>268, maybe more</td>
<td>not available</td>
<td>not available</td>
<td>not available</td>
</tr>
<tr>
<td>OpenVigil 1 SQL without DEMO data prior to 2005Q3</td>
<td>251</td>
<td>202</td>
<td>not calculated</td>
<td>not calculated</td>
</tr>
<tr>
<td>OpenVigil 1 SQL (full LAERS data)</td>
<td>382</td>
<td>299, a few less because of multiplicates</td>
<td>not calculated</td>
<td>not calculated</td>
</tr>
<tr>
<td>OpenVigil 2.0 GUI (default install)</td>
<td>162</td>
<td>140</td>
<td>3.109*</td>
<td>3.122 (2.676; 3.642)</td>
</tr>
<tr>
<td>OpenVigil 2.1 GUI (additional manual drugname mapping)</td>
<td>166</td>
<td>143</td>
<td>3.141 (reports)</td>
<td>3.505 (cases)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.154 (reports)</td>
<td>3.522 (cases)</td>
</tr>
<tr>
<td>Sakaeda 2013</td>
<td>not reported</td>
<td>268</td>
<td>1.991</td>
<td>2.006 (1.778; 2.234)</td>
</tr>
</tbody>
</table>

*) all measurements of disproportionality were calculated on reports, not cases in OpenVigil 2.0. Congruence or marked disagreement are printed in bold letters.

Conclusions:

Using OpenVigil 1 is tedious work: You have to think yourself about which names and synonyms to use. Due to the constraints in the OpenVigil 1 implementation running currently at Kiel University, you cannot put everything into one big query. The output has to be manually checked to avoid duplicates.

Using OpenVigil 1 with SQL allows extraction of raw data which can further cleansed, e.g., of the 268 resp. 413 reports initially mentioned above, only at most 202 resp. 299 are unique cases.

OpenVigil 2 is much easier to use but offers just 140 resp. 143 of the putative 299 cases. However, here you can trust that only valid reports with an unambiguous mapping of the free-text drugname to a USAN drugname were included in the analysis. A reason for not finding the potential additional reports can be our drugname mapping system: Names like "WARFARIN 5 MG”, “WARFARIN (WARFARIN POTASSIUM)”, “WARFARIN 2.5 MG COUMADIN” are clear and understandable for human users but the drugname mapping system currently discards these verbatim "drugnames” to avoid potential mismapping.

There is no exact information available on how Sakaeda extracted the 268 cases and the other non-case-numbers needed for disproportionality analysis since the Japanese closed source system CzeekV by Kyoto Constella Technology was used. It is interesting to see that we can reproduce the number 268 when counting reports (including duplicates) and not using data prior to 2005Q3.

We can see that changes in the number of cases (268 vs 162) and non-cases (the remaining 3 fields of the 2x2 contingency table) can have a serious impact on signal generation (PRR 1.991 is smaller than 2 and does not thus yield a signal).
4. SQL-database schema:
5. References and resources

http://math.hws.edu/javamath/ryan/ChiSquare.html
http://en.wikipedia.org/wiki/Adverse_event
http://en.wikipedia.org/wiki/Loperamide
http://en.wikipedia.org/wiki/Odds_ratio
http://en.wikipedia.org/wiki/Pharmacovigilance
http://en.wikipedia.org/wiki/Pirenzepine
http://en.wikipedia.org/wiki/Proportional_reporting_ratio
http://en.wikipedia.org/wiki/SQL
https://people.richland.edu/james/lecture/m170/tbl-chi.html