Open Vigil

a pharmacovigilance data analysis tool

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

Contents

1. Introduction

OpenVigil 2.0 (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/) 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[1]).
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.

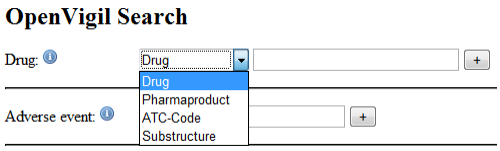
[1] Cave-at documents:
OpenVigil 1: http://www.uni-kiel.de/pharmacology/pvt/caveat.html

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):

Examples of differences between USAN and other drug names	
International Nonproprietary Name (INN)	**U.S. Adopted Name (USAN)**
glibenclamide	glyburide
acetylsalicylic acid	aspirin
metamizole	dipyrone
salbutamol	albuterol
paracetamol	acetaminophen
rifampicin	rifampin
suxamethonium	succinylcholine
glyceryl trinitrate	nitroglycerin

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]

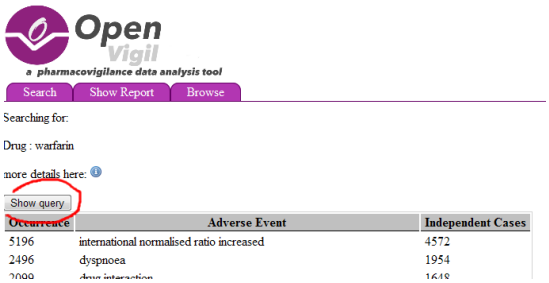
2.5. Structured Query Language (SQL)

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).[3]
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.



2 http://en.wikipedia.org/wiki/Adverse_event

3 http://en.wikipedia.org/wiki/SQL

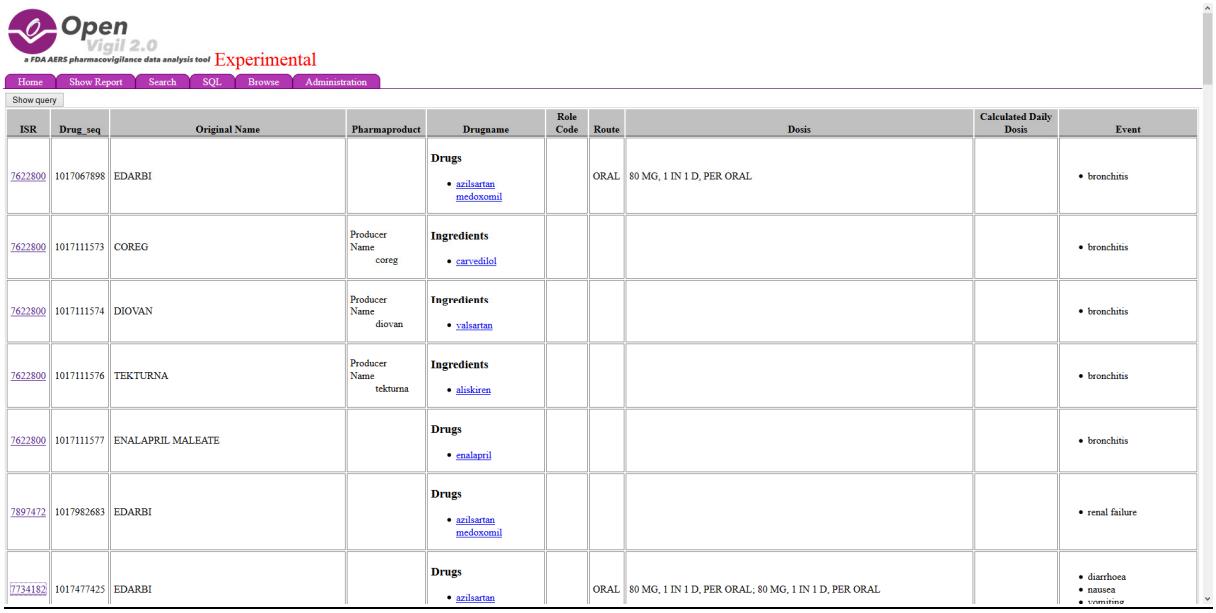
3. Examples

3.1. Individual Safety Reports (ISR)

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

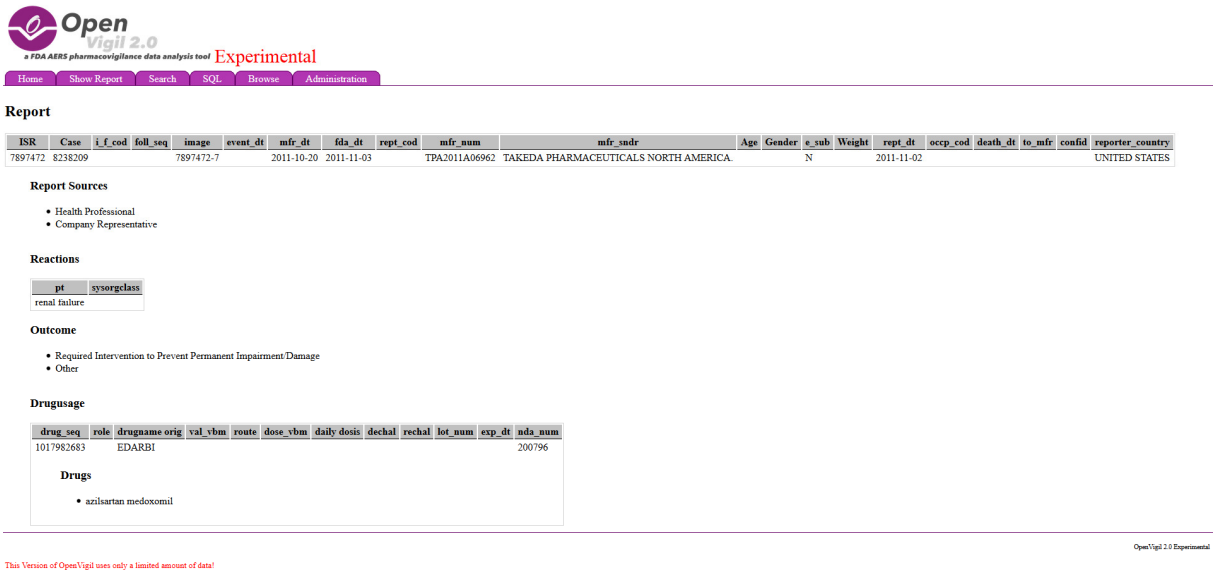
Query construction: Choose “drug“ in “OpenVigil Search“; drugname is “azilsartan medoxomil“.

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



ISR	Drug_seq	Original Name	Pharmaproduct	Drugname	Role Code	Route	Dosis	Calculated Daily Dosis	Event
7622800	1017067898	EDARBI		**Drugs** • azilsartan medoxomil		ORAL	80 MG, 1 IN 1 D, PER ORAL		• bronchitis
7622800	1017111573	COREG	Producer Name coreg	**Ingredients** • carvedilol					• bronchitis
7622800	1017111574	DIOVAN	Producer Name diovan	**Ingredients** • valsartan					• bronchitis
7622800	1017111576	TEKTURNA	Producer Name tekturna	**Ingredients** • aliskiren					• bronchitis
7622800	1017111577	ENALAPRIL MALEATE		**Drugs** • enalapril					• bronchitis
7897472	1017982683	EDARBI		**Drugs** • azilsartan medoxomil					• renal failure
7734182	1017477425	EDARBI		**Drugs** • azilsartan		ORAL	80 MG, 1 IN 1 D, PER ORAL; 80 MG, 1 IN 1 D, PER ORAL		• diarrhoea • nausea • vomiting

Single Report:



OpenVigil 2.0 a FDA AERS pharmacovigilance data analysis tool Experimental

Home | Show Report | Search | SQL | Browse | Administration

Report

ISR	Case	i_f_cod	foll_seq	image	event_dt	mfr_dt	fda_dt	rept_cod	mfr_num	mfr_sndr	Age	Gender	e_sub	Weight	rept_dt	occp_cod	death_dt	to_mfr	confid	reporter_country
7897472	8238209			7897472-7		2011-10-20	2011-11-03		TPA2011A06962	TAKEDA PHARMACEUTICALS NORTH AMERICA.			N		2011-11-02					UNITED STATES

Report Sources

- Health Professional
- Company Representative

Reactions

pt	sysorgclass
renal failure	

Outcome

- Required Intervention to Prevent Permanent Impairment/Damage
- Other

Drugusage

drug_seq	role	drugname orig	val_vbm	route	dose_vbm	daily dosis	dechal	rechal	lot_num	exp_dt	nda_num
1017982683		EDARBI									200796

Drugs

- azilsartan medoxomil

OpenVigil 2.0 Experimental

This Version of OpenVigil uses only a limited amount of data!

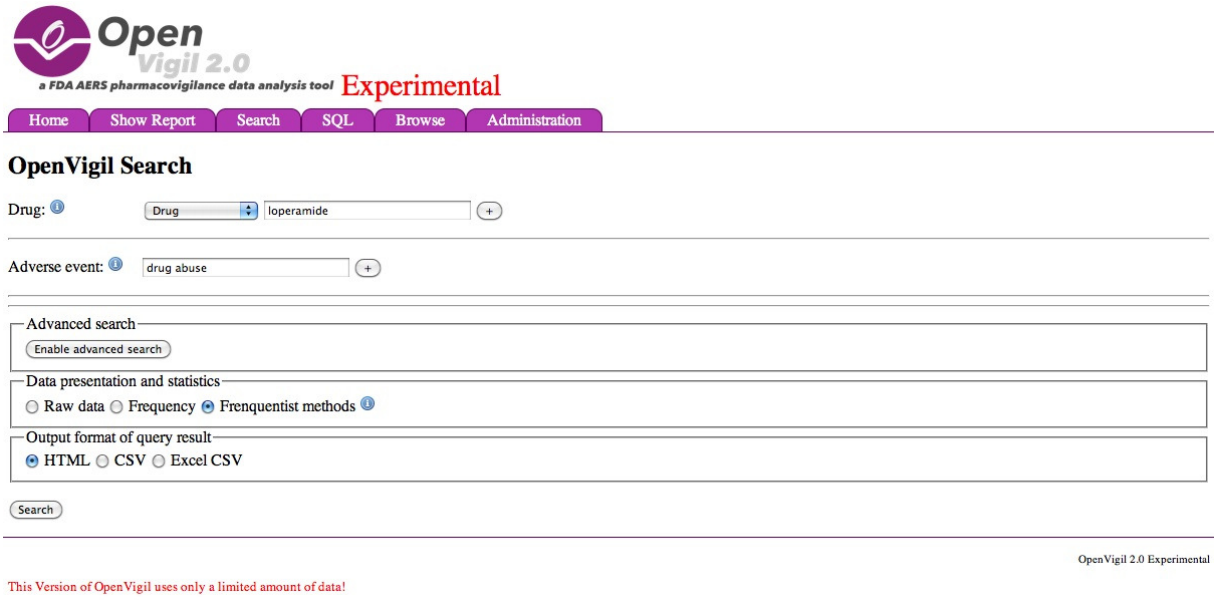
In the ISR above some data (for example age, gender and weight of the patient) are missing.

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.

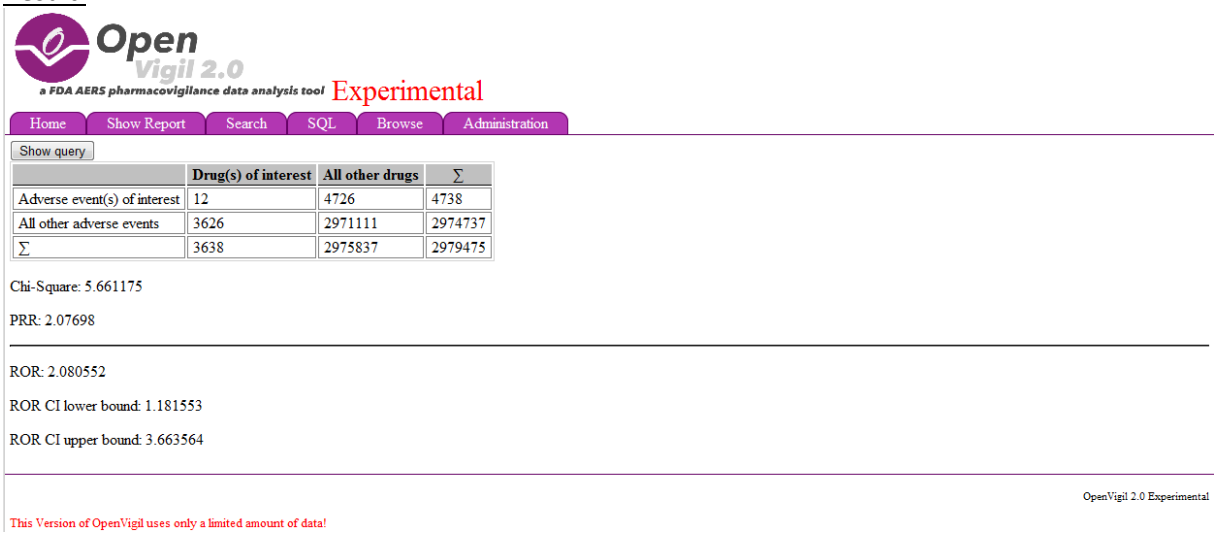
3.2. Interpretation of statistics used in OpenVigil 2.0

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):



Result:



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 %.[4] [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.[6]
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.[8]

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/pharmaproduct

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:

Occurrence	Adverse Event
638	drug interaction
609	dyspnoea
603	atrial fibrillation
475	asthenia
401	renal failure
382	hypotension
379	nausea
370	dizziness
322	bradycardia
320	hyperthyroidism
318	renal failure acute
303	pneumonia
303	cardiac arrest
302	fall
301	cardiac failure congestive
300	pain
283	fatigue
282	international normalised ratio increased
269	anxiety
261	vomiting
257	death
255	syncope
254	ventricular tachycardia
248	torsade de pointes
246	cardiac failure
236	pulmonary fibrosis
234	drug ineffective
225	electrocardiogram qt prolonged
222	oedema peripheral

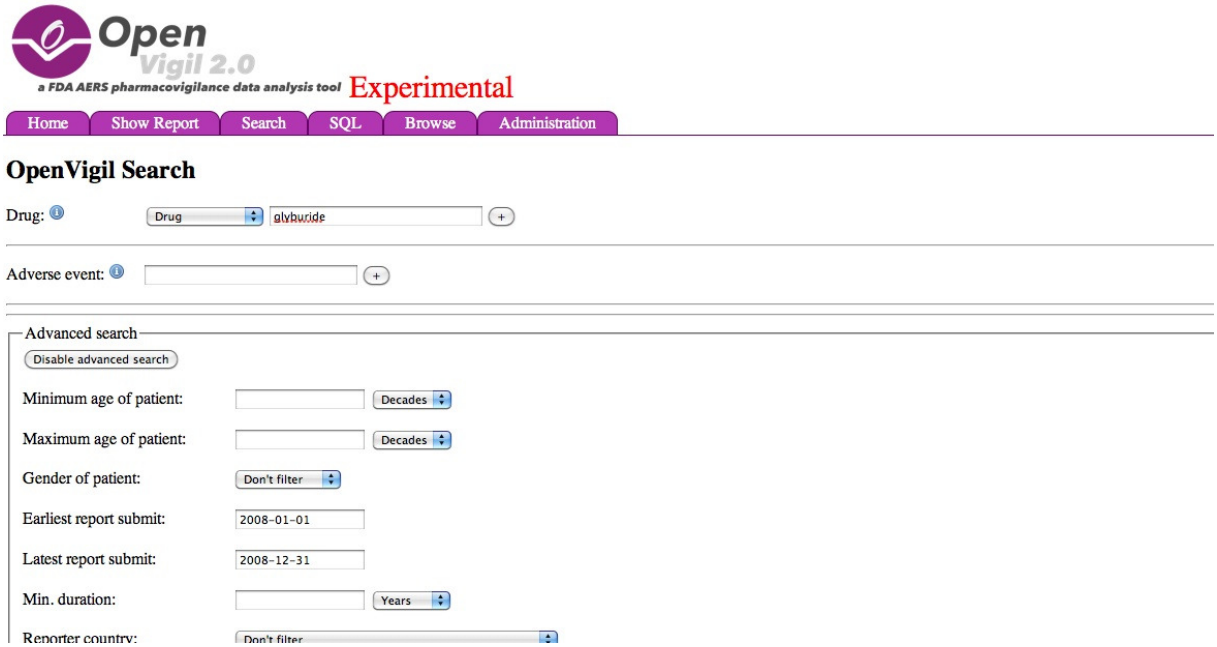
Most reported adverse event is “drug interaction“.
An explanation of the result might be that amiodarone inhibits a drug-metabolizing cytochrome P450 enzyme, isoform 3A4 (CYP3A4). Many drugs are metabolised by CYP3A4. An inhibition of CYP3A4 consequently increases the bioavailability of those drugs.

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).

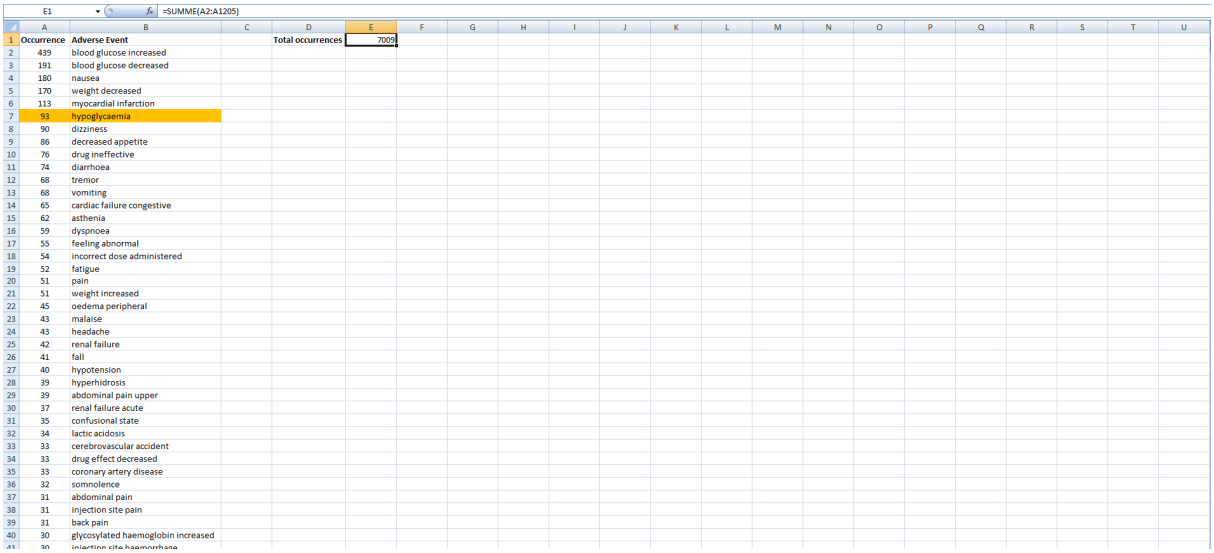
3.4. Query construction for a specific time interval

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.



E1 =SUMME(A2:A1205)

	A	B	C	D	E
1	Occurrence	Adverse Event		Total occurrences	7009
2	439	blood glucose increased			
3	191	blood glucose decreased			
4	180	nausea			
5	170	weight decreased			
6	113	myocardial infarction			
7	93	hypoglycaemia			
8	90	dizziness			
9	86	decreased appetite			
10	76	drug ineffective			
11	74	diarrhoea			
12	68	tremor			
13	68	vomiting			
14	65	cardiac failure congestive			
15	62	asthenia			
16	59	dyspnoea			
17	55	feeling abnormal			
18	54	incorrect dose administered			
19	52	fatigue			
20	51	pain			
21	51	weight increased			
22	45	oedema peripheral			
23	43	malaise			
24	43	headache			
25	42	renal failure			
26	41	fall			
27	40	hypotension			
28	39	hyperhidrosis			
29	39	abdominal pain upper			
30	37	renal failure acute			
31	35	confusional state			
32	34	lactic acidosis			
33	33	cerebrovascular accident			
34	33	drug effect decreased			
35	33	coronary artery disease			
36	32	somnolence			
37	31	abdominal pain			
38	31	injection site pain			
39	31	back pain			
40	30	glycosylated haemoglobin increased			

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 pharmaproduct

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:

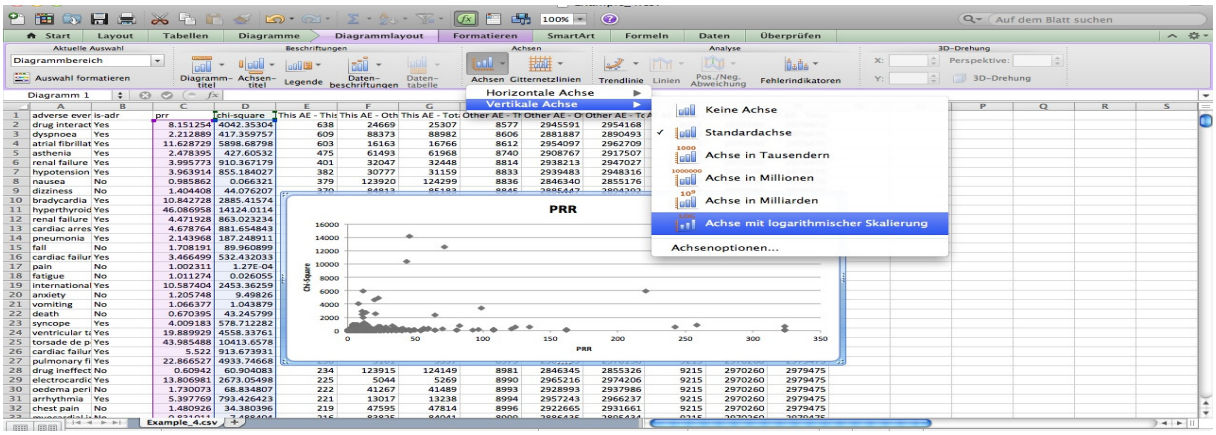
Excel CSV file imported into Excel:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	adverse event	is-adr	prr	chi-square	This AE - This	This AE - Oth	This AE - Tot	Other AE - Th	Other AE - O	Other AE - Tc	All AE - This I	All AE - Othe	All AE - Total	
2	drug interaction	Yes	8.151254	4042.35304	638	24669	25307	8577	2945591	2954168	9215	2970260	2979475	
3	dyspnoea	Yes	2.212889	417.359757	609	88373	88982	8606	2881887	2890493	9215	2970260	2979475	
4	atrial fibrillation	Yes	11.628729	5898.68798	603	16163	16766	8612	2954097	2962709	9215	2970260	2979475	
5	asthenia	Yes	2.478395	427.60532	475	61493	61968	8740	2908767	2917507	9215	2970260	2979475	
6	renal failure	Yes	3.995773	910.367179	401	32047	32448	8814	2938213	2947027	9215	2970260	2979475	
7	hypotension	Yes	3.963914	855.184027	382	30777	31159	8833	2939483	2948316	9215	2970260	2979475	
8	nausea	No	0.985862	0.066321	379	123920	124299	8836	2846340	2855176	9215	2970260	2979475	
9	dizziness	No	1.404408	44.076207	370	84813	85183	8845	2885447	2894292	9215	2970260	2979475	
10	bradycardia	Yes	10.842728	2885.41574	322	9280	9602	8893	2960980	2969873	9215	2970260	2979475	
11	hyperthyroidism	Yes	46.086958	14124.0114	320	1925	2245	8895	2968335	2977230	9215	2970260	2979475	
12	renal failure acute	Yes	4.471928	863.023234	318	22674	22992	8897	2947586	2956483	9215	2970260	2979475	
13	cardiac arrest	Yes	4.678764	881.654843	303	20636	20939	8912	2949624	2958536	9215	2970260	2979475	
14	pneumonia	Yes	2.143968	187.248911	303	45392	45695	8912	2924868	2933780	9215	2970260	2979475	
15	fall	No	1.708191	89.960899	302	56861	57163	8913	2913399	2922312	9215	2970260	2979475	
16	cardiac failure congestive	Yes	3.466499	532.432033	301	27774	28075	8914	2942486	2951400	9215	2970260	2979475	
17	pain	No	1.002311	1.27E-04	300	96475	96775	8915	2873785	2882700	9215	2970260	2979475	
18	fatigue	No	1.011274	0.026055	283	90199	90482	8932	2880061	2888993	9215	2970260	2979475	
19	international normalised ratio increased	Yes	10.587404	2453.36259	282	8330	8612	8933	2961930	2970863	9215	2970260	2979475	
20	anxiety	No	1.205748	9.49826	269	71865	72134	8946	2898395	2907341	9215	2970260	2979475	
21	vomiting	No	1.066377	1.043879	261	78875	79136	8954	2891385	2900339	9215	2970260	2979475	
22	death	No	0.670395	43.245799	257	123693	123950	8958	2846567	2855525	9215	2970260	2979475	
23	syncope	Yes	4.009183	578.712282	255	20310	20565	8960	2949950	2958910	9215	2970260	2979475	
24	ventricular tachycardia	Yes	19.889929	4558.33761	254	3875	4129	8961	2966385	2975346	9215	2970260	2979475	
25	torsade de pointes	Yes	43.985488	10413.6578	248	1575	1823	8967	2968685	2977652	9215	2970260	2979475	
26	cardiac failure	Yes	5.522	913.673931	246	14158	14404	8969	2956102	2965071	9215	2970260	2979475	
27	pulmonary fibrosis	Yes	22.866527	4933.74668	236	3101	3337	8979	2967159	2976138	9215	2970260	2979475	
28	drug ineffective	No	0.60942	60.904083	234	123915	124149	8981	2846345	2855326	9215	2970260	2979475	
29	electrocardiogram qt prolonged	Yes	13.806981	2673.05498	225	5044	5269	8990	2965216	2974206	9215	2970260	2979475	
30	oedema peripheral	No	1.730073	68.834807	222	41267	41489	8993	2928993	2937986	9215	2970260	2979475	
31	arrhythmia	Yes	5.397769	793.426423	221	13017	13238	8994	2957243	2966237	9215	2970260	2979475	
32	chest pain	No	1.480926	34.380396	219	47595	47814	8996	2922665	2931661	9215	2970260	2979475	
33	myocardial infarction	No	0.831011	7.488404	216	83825	84041	8999	2886435	2895434	9215	2970260	2979475	

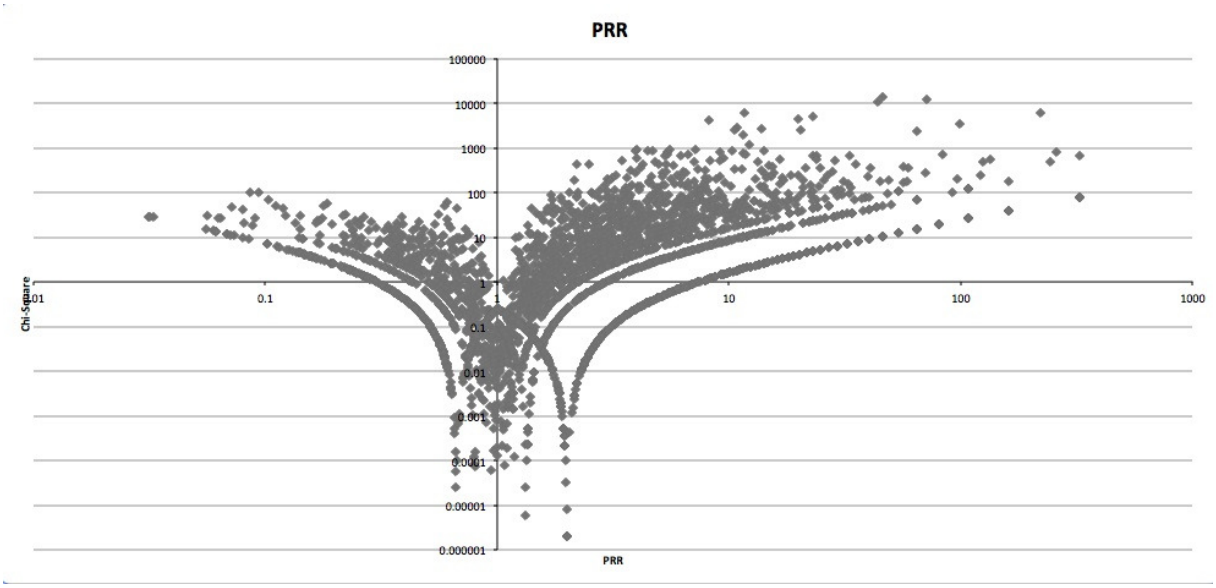
result-1.csv

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:
http://en.wikipedia.org/wiki/Decimal_mark

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:

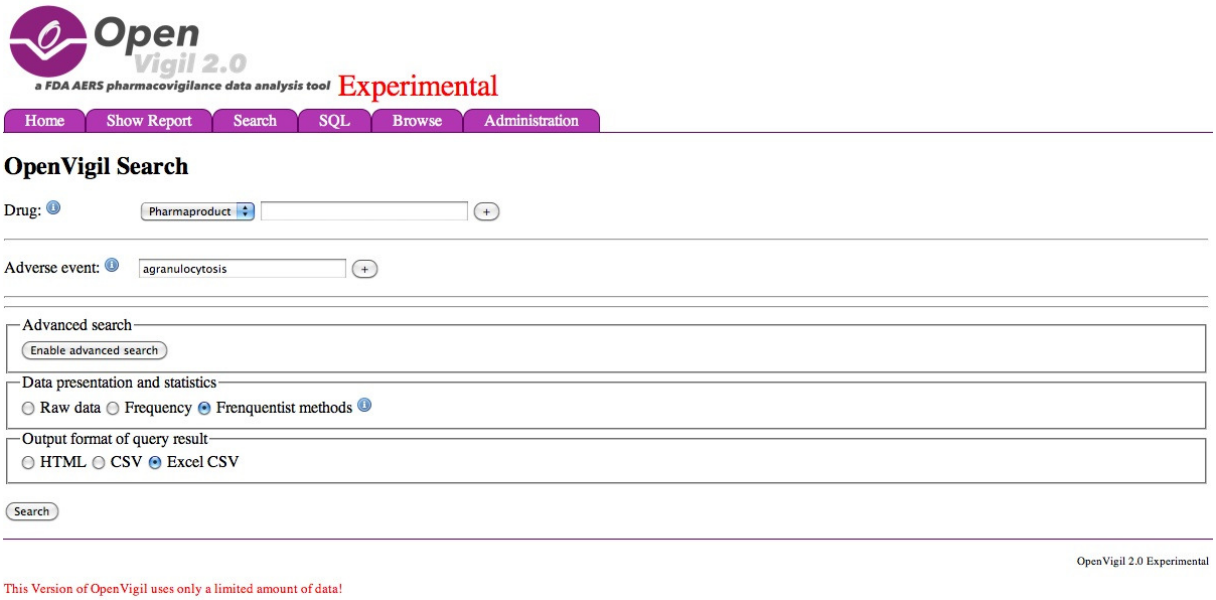


The upper-right quadrant contains putative adverse drug reactions. Everything else is just an adverse event.
In the result list “drug interaction” (cf. example above) is reported with a PRR of 8.151 and a Chi-Squared value of 4042. Due to this drug interaction is very likely an adverse drug reaction of amiodarone.
However, prior knowledge of this CYP3A4 inhibtion 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.



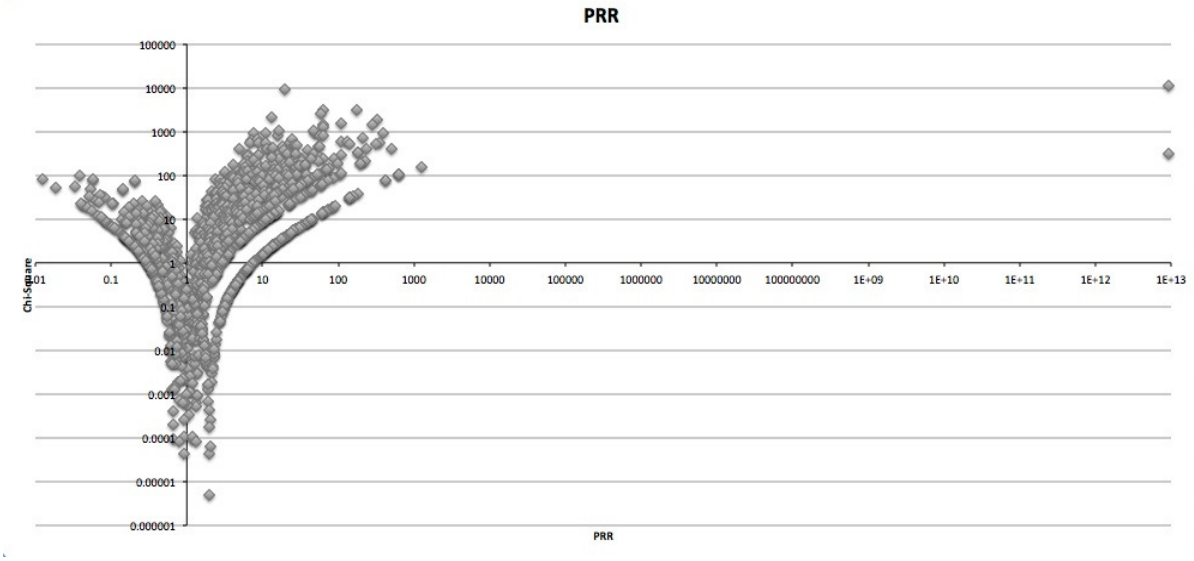
Results:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	drug	is-adr	rPRR	chi-square	This AE - This	This AE - Oth	This AE - Tot	Other AE - Th	Other AE - O	Other AE - Tc	All AE - This I	All AE - Othe	All AE - Total	
2	clozaril	Yes	13.125255	2085.01309	188	2169	2357	18092	2959026	2977118	18280	2961195	2979475	
3	seroquel	No	1.297905	4.736305	72	2285	2357	70069	2907049	2977118	70141	2909334	2979475	
4	nexium	No	1.745795	20.262043	65	2292	2357	47028	2930090	2977118	47093	2932382	2979475	
5	aspirin	No	1.082183	0.32295	63	2294	2357	73532	2903586	2977118	73595	2905880	2979475	
6	neupogen	Yes	16.072748	795.933344	58	2299	2357	4558	2972560	2977118	4616	2974859	2979475	
7	zyprexa	Yes	2.377284	36.374641	47	2310	2357	24972	2952146	2977118	25019	2954456	2979475	
8	depakote	Yes	4.649434	125.333382	45	2312	2357	12225	2964893	2977118	12270	2967205	2979475	
9	bactrim	Yes	11.360637	402.489563	44	2313	2357	4892	2972226	2977118	4936	2974539	2979475	
10	mabthera	Yes	8.876147	291.24795	43	2314	2357	6119	2970999	2977118	6162	2973313	2979475	
11	cellcept	Yes	8.863834	283.788748	42	2315	2357	5985	2971133	2977118	6027	2973448	2979475	
12	risperdal	Yes	2.405451	32.449801	41	2316	2357	21529	2955589	2977118	21570	2957905	2979475	
13	lasix	Yes	2.247713	25.221799	38	2319	2357	21354	2955764	2977118	21392	2958083	2979475	
14	lyrica	No	0.907665	0.237241	34	2323	2357	47314	2929804	2977118	47348	2932127	2979475	
15	plavix	No	1.549532	6.137338	34	2323	2357	27715	2949403	2977118	27749	2951726	2979475	
16	keppra	Yes	3.881745	67.781827	33	2324	2357	10738	2966380	2977118	10771	2968704	2979475	
17	diovan	Yes	2.006109	15.234706	32	2325	2357	20148	2956970	2977118	20180	2959295	2979475	
18	tegretol	Yes	4.131985	72.868292	32	2325	2357	9782	2967336	2977118	9814	2969661	2979475	
19	abilify	No	1.843502	11.198829	31	2326	2357	21240	2955878	2977118	21271	2958204	2979475	
20	taxotere	Yes	5.978926	123.286983	31	2326	2357	6549	2970569	2977118	6580	2972895	2979475	
21	predonine	Yes	17.964587	441.791956	29	2328	2357	2039	2975079	2977118	2068	2977407	2979475	
22	leponex	Yes	22.758492	551.261238	28	2329	2357	1554	2975564	2977118	1582	2977893	2979475	
23	prograf	Yes	4.27685	64.503615	27	2330	2357	7974	2969144	2977118	8001	2971474	2979475	
24	rituxan	Yes	6.156996	111.186906	27	2330	2357	5539	2971579	2977118	5566	2973909	2979475	
25	haldol	Yes	6.45958	114.185299	26	2331	2357	5084	2972034	2977118	5110	2974365	2979475	
26	neoral	Yes	5.747376	97.006121	26	2331	2357	5714	2971404	2977118	5740	2973735	2979475	
27	fazaclo odt	Yes	63.792743	1411.95498	25	2332	2357	495	2976623	2977118	520	2978955	2979475	
28	gleevec	No	1.76934	7.646049	25	2332	2357	17847	2959271	2977118	17872	2961603	2979475	
29	rocephin	Yes	13.511942	274.545282	25	2332	2357	2337	2974781	2977118	2362	2977113	2979475	
30	remicade	No	0.407248	20.626132	24	2333	2357	74437	2902681	2977118	74461	2905014	2979475	
31	enbrel	No	0.206598	72.707892	23	2334	2357	140617	2836501	2977118	140640	2838835	2979475	
32	fosamax	No	0.697894	2.750064	23	2334	2357	41627	2935491	2977118	41650	2937825	2979475	

result-3.csv

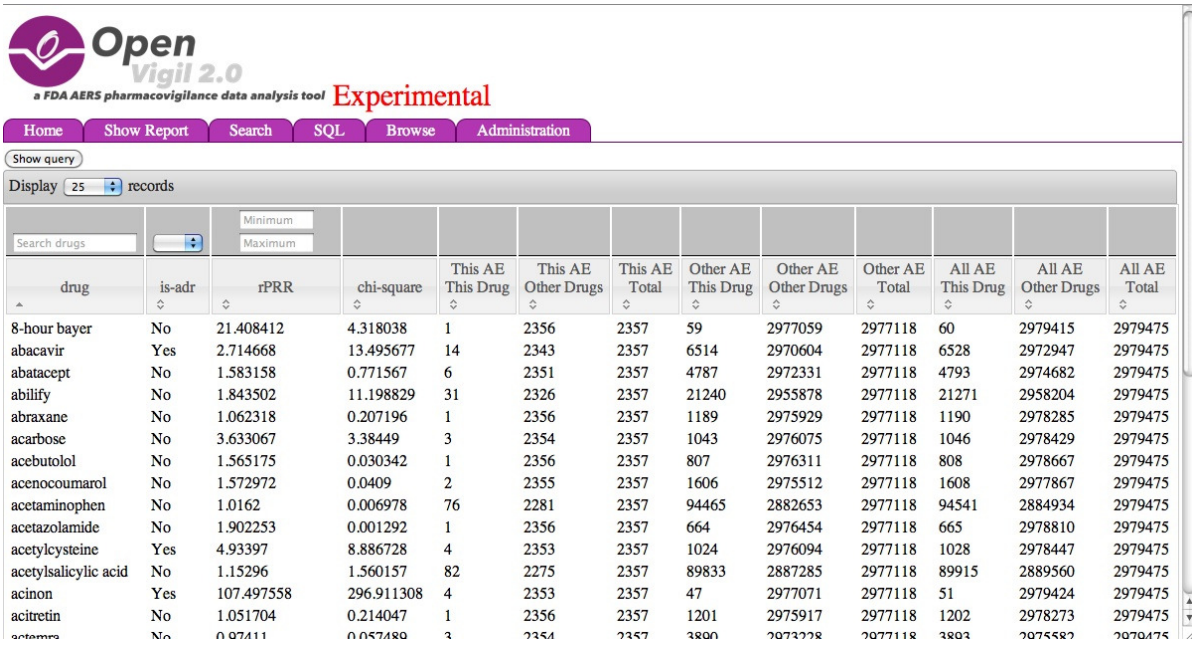
The resulting list contains names of drugs and pharmaproducts.

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:

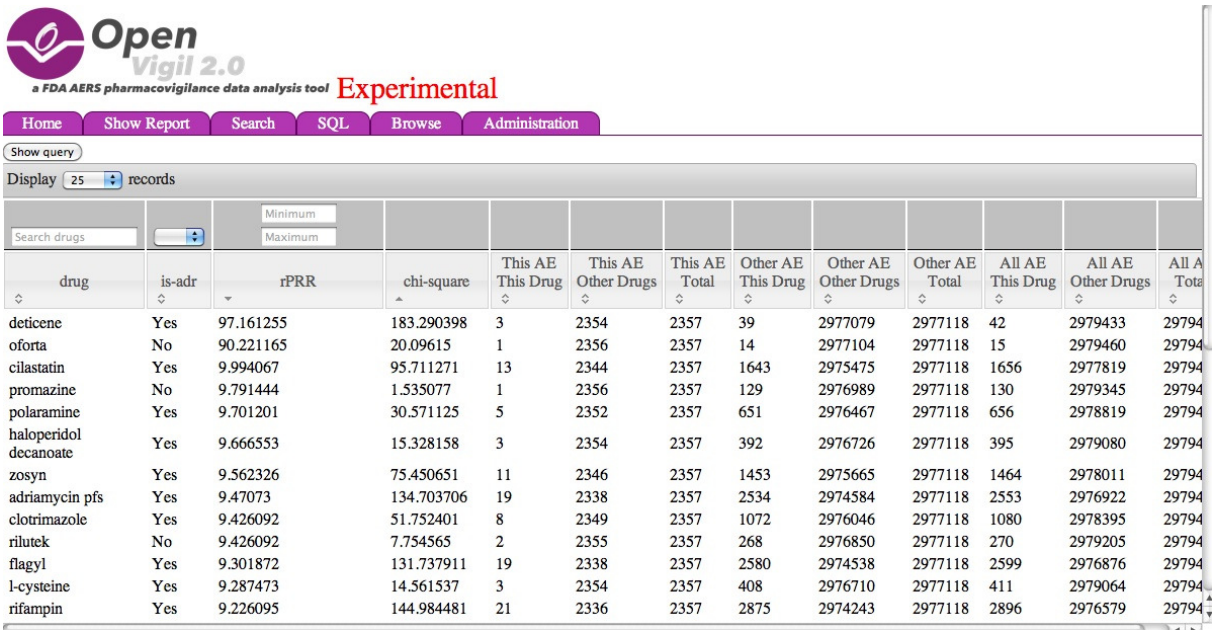


drug	is-adr	rPRR	chi-square	This AE This Drug	This AE Other Drugs	This AE Total	Other AE This Drug	Other AE Other Drugs	Other AE Total	All AE This Drug	All AE Other Drugs	All AE Total
8-hour bayer	No	21.408412	4.318038	1	2356	2357	59	2977059	2977118	60	2979415	2979475
abacavir	Yes	2.714668	13.495677	14	2343	2357	6514	2970604	2977118	6528	2972947	2979475
abatacept	No	1.583158	0.771567	6	2351	2357	4787	2972331	2977118	4793	2974682	2979475
abilify	No	1.843502	11.198829	31	2326	2357	21240	2955878	2977118	21271	2958204	2979475
abraxane	No	1.062318	0.207196	1	2356	2357	1189	2975929	2977118	1190	2978285	2979475
acarbose	No	3.633067	3.38449	3	2354	2357	1043	2976075	2977118	1046	2978429	2979475
acebutolol	No	1.565175	0.030342	1	2356	2357	807	2976311	2977118	808	2978667	2979475
acenocoumarol	No	1.572972	0.0409	2	2355	2357	1606	2975512	2977118	1608	2977867	2979475
acetaminophen	No	1.0162	0.006978	76	2281	2357	94465	2882653	2977118	94541	2884934	2979475
acetazolamide	No	1.902253	0.001292	1	2356	2357	664	2976454	2977118	665	2978810	2979475
acetylcysteine	Yes	4.93397	8.886728	4	2353	2357	1024	2976094	2977118	1028	2978447	2979475
acetylsalicylic acid	No	1.15296	1.560157	82	2275	2357	89833	2887285	2977118	89915	2889560	2979475
acinon	Yes	107.497558	296.911308	4	2353	2357	47	2977071	2977118	51	2979424	2979475
acitretin	No	1.051704	0.214047	1	2356	2357	1201	2975917	2977118	1202	2978273	2979475

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.)

9 http://en.wikipedia.org/wiki/Pirenzepine

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:

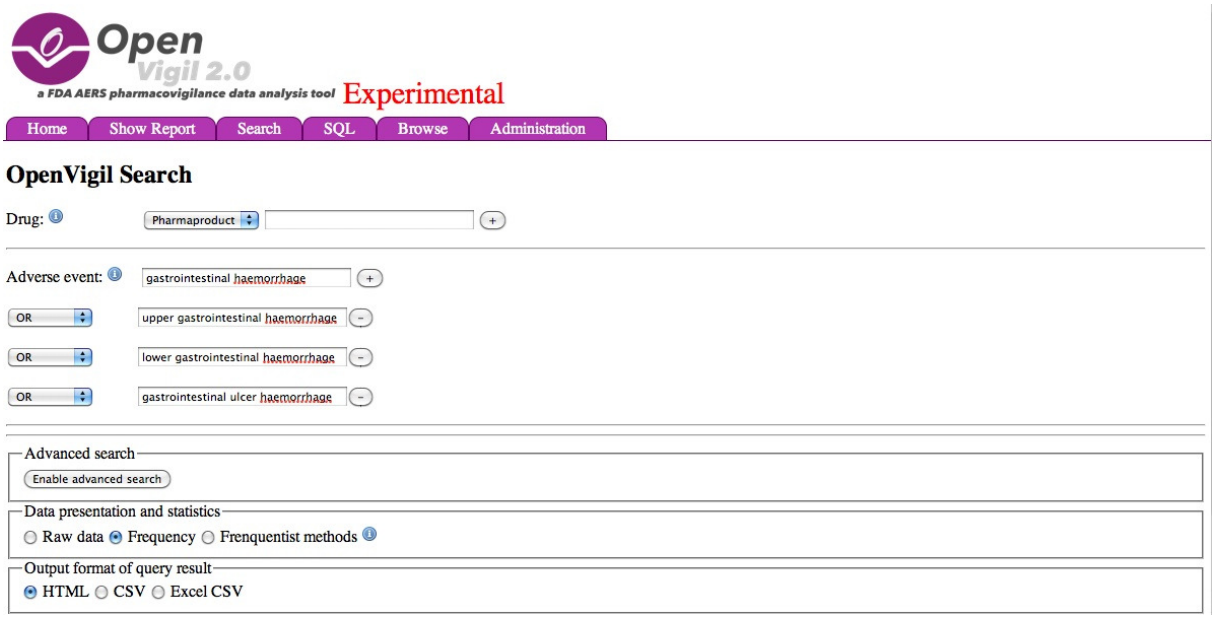


drug	is-adr	rPRR	chi-square	This AE This Drug	This AE Other Drugs	This AE Total	Other AE This Drug	Other AE Other Drugs	Other AE Total	All AE This Drug	All AE Other Drugs	All A Tota
deticene	Yes	97.161255	183.290398	3	2354	2357	39	2977079	2977118	42	2979433	29794
oforta	No	90.221165	20.09615	1	2356	2357	14	2977104	2977118	15	2979460	29794
cilastatin	Yes	9.994067	95.711271	13	2344	2357	1643	2975475	2977118	1656	2977819	29794
promazine	No	9.791444	1.535077	1	2356	2357	129	2976989	2977118	130	2979345	29794
polaramine	Yes	9.701201	30.571125	5	2352	2357	651	2976467	2977118	656	2978819	29794
haloperidol decanoate	Yes	9.666553	15.328158	3	2354	2357	392	2976726	2977118	395	2979080	29794
zosyn	Yes	9.562326	75.450651	11	2346	2357	1453	2975665	2977118	1464	2978011	29794
adriamycin pfs	Yes	9.47073	134.703706	19	2338	2357	2534	2974584	2977118	2553	2976922	29794
clotrimazole	Yes	9.426092	51.752401	8	2349	2357	1072	2976046	2977118	1080	2978395	29794
rilutek	No	9.426092	7.754565	2	2355	2357	268	2976850	2977118	270	2979205	29794
flagyl	Yes	9.301872	131.737911	19	2338	2357	2580	2974538	2977118	2599	2976876	29794
l-cysteine	Yes	9.287473	14.561537	3	2354	2357	408	2976710	2977118	411	2979064	29794
rifampin	Yes	9.226095	144.984481	21	2336	2357	2875	2974243	2977118	2896	2976579	29794

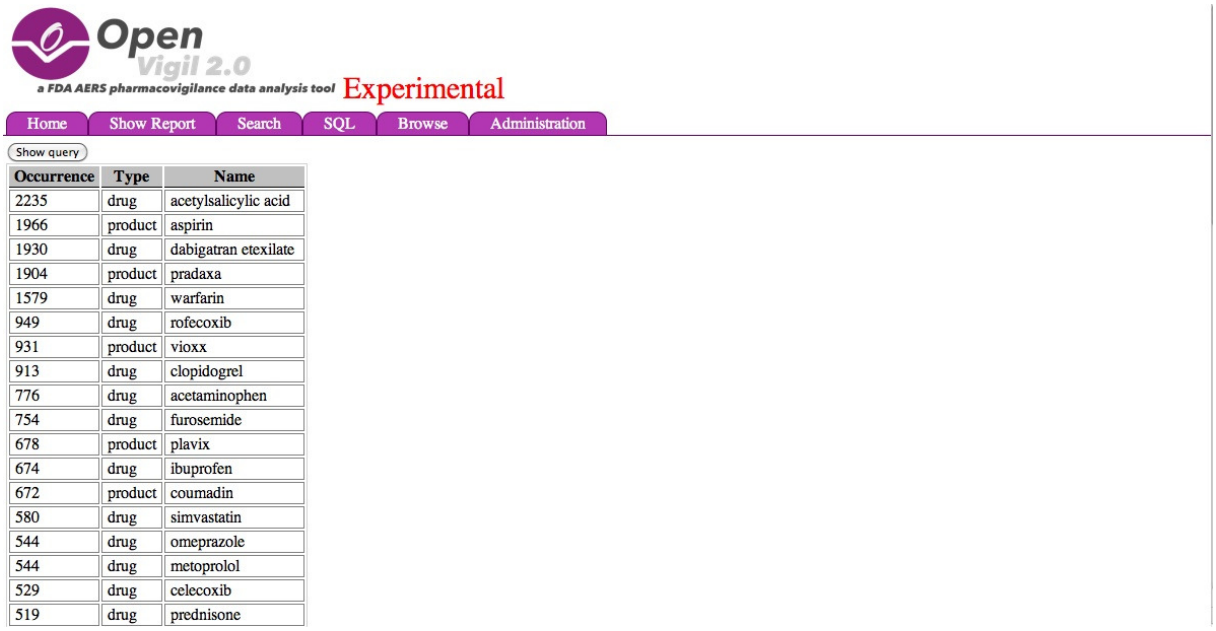
3.7. Query construction for different adverse events

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:



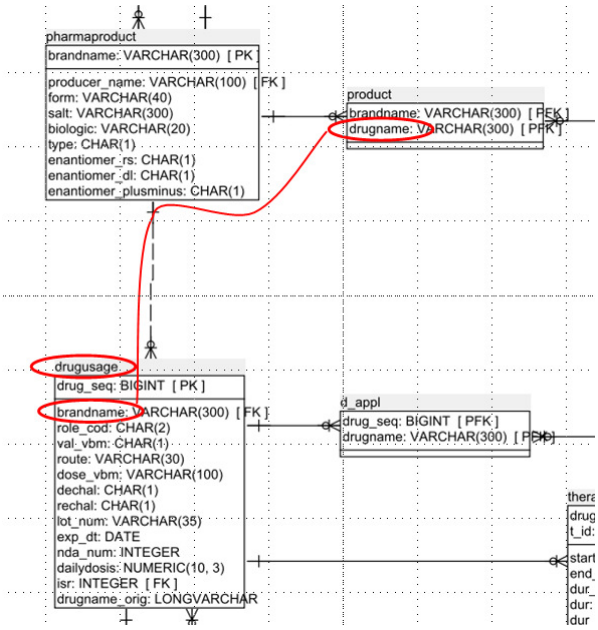
Occurrence	Type	Name
2235	drug	acetylsalicylic acid
1966	product	aspirin
1930	drug	dabigatran etexilate
1904	product	pradaxa
1579	drug	warfarin
949	drug	rofecoxib
931	product	vioxx
913	drug	clopidogrel
776	drug	acetaminophen
754	drug	furosemide
678	product	plavix
674	drug	ibuprofen
672	product	coumadin
580	drug	simvastatin
544	drug	omeprazole
544	drug	metoprolol
529	drug	celecoxib
519	drug	prednisone

The two most reported pharmaproducts with gastrointestinal haemorrhage as an adverse event are aspirin and pradaxa.

3.8. Structure Query Language (SQL)

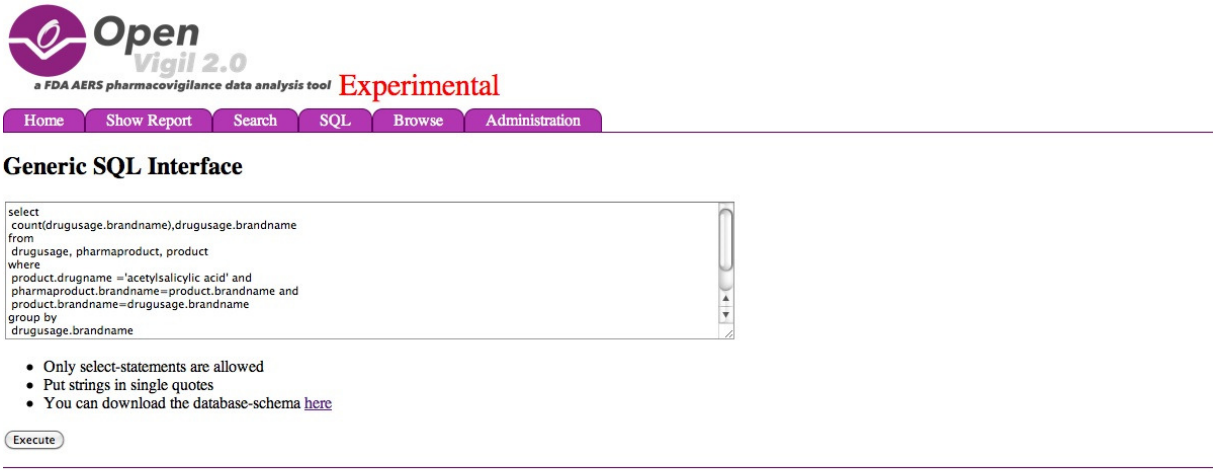
Problem: The occurrence of gastrointestinal haemorrhage as an adverse event of the two most used acetylsalicylic acid-containing pharmaproducts shall be compared. A very complex query was constructed that cannot be created with the GUI of OpenVigil 2.0:

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:



Query construction in SQL:

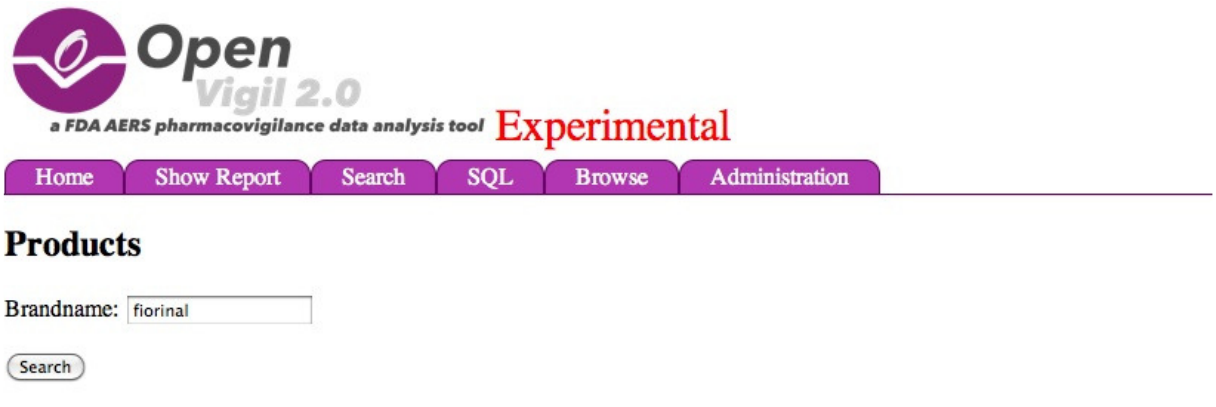
```
select
      count(drugusage.brandname),drugusage.brandname
from
      drugusage, pharmaproduct, product
where
      product.drugname ='acetylsalicylic acid' and
      pharmaproduct.brandname=product.brandname and
      product.brandname=drugusage.brandname
group by
      drugusage.brandname
order by
      count(drugusage.brandname) desc
```



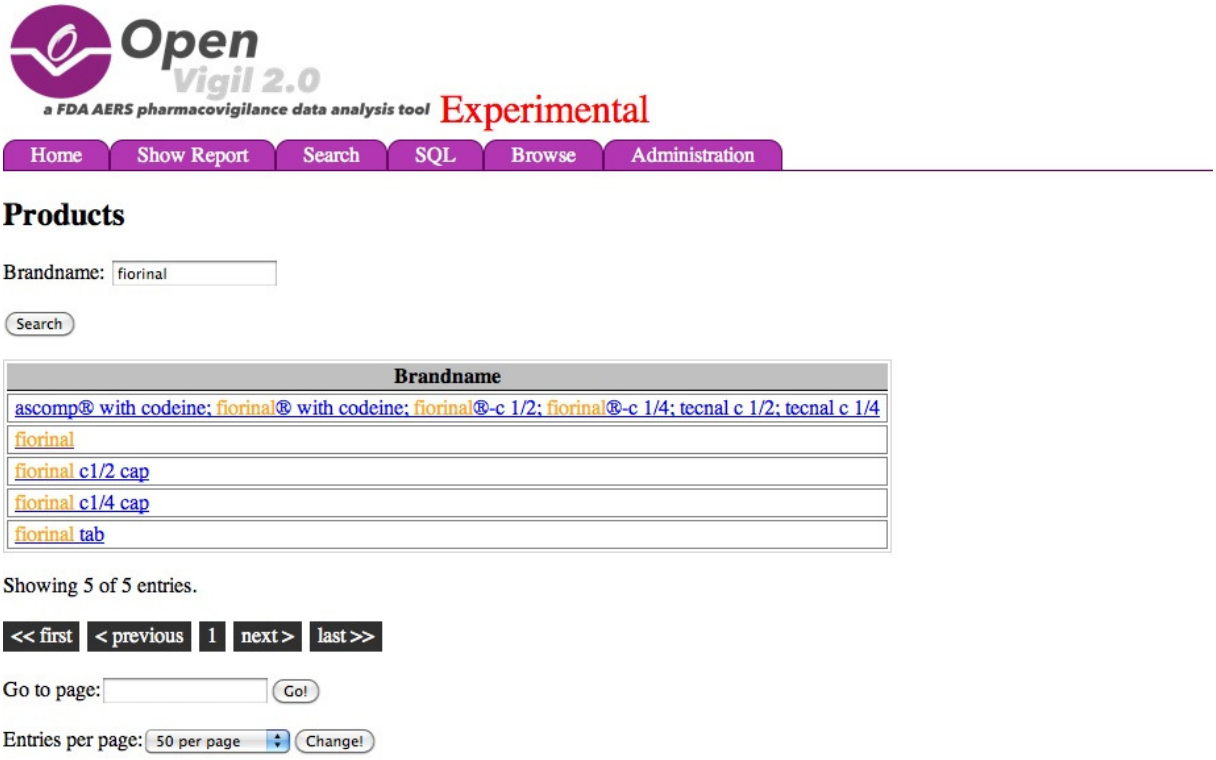
Results: Query result is a list with 31 pharmaproducts (brand names).

count	brandname
73595	aspirin
647	ecotrin
455	fiorinal
453	bufferin
430	midol
164	coricidin
92	adiro
65	aspirine
60	8-hour bayer
58	anacin
50	nu-seals aspirin
41	entrophen
27	norgesic
23	dristan
20	solprin
16	nu-seals
13	easprin
13	aspro
8	acetosal
3	ratio-tecnal
3	bayer extra strength aspirin for migraine pain
3	acenterine
2	acetol
1	colfarit
1	empirin

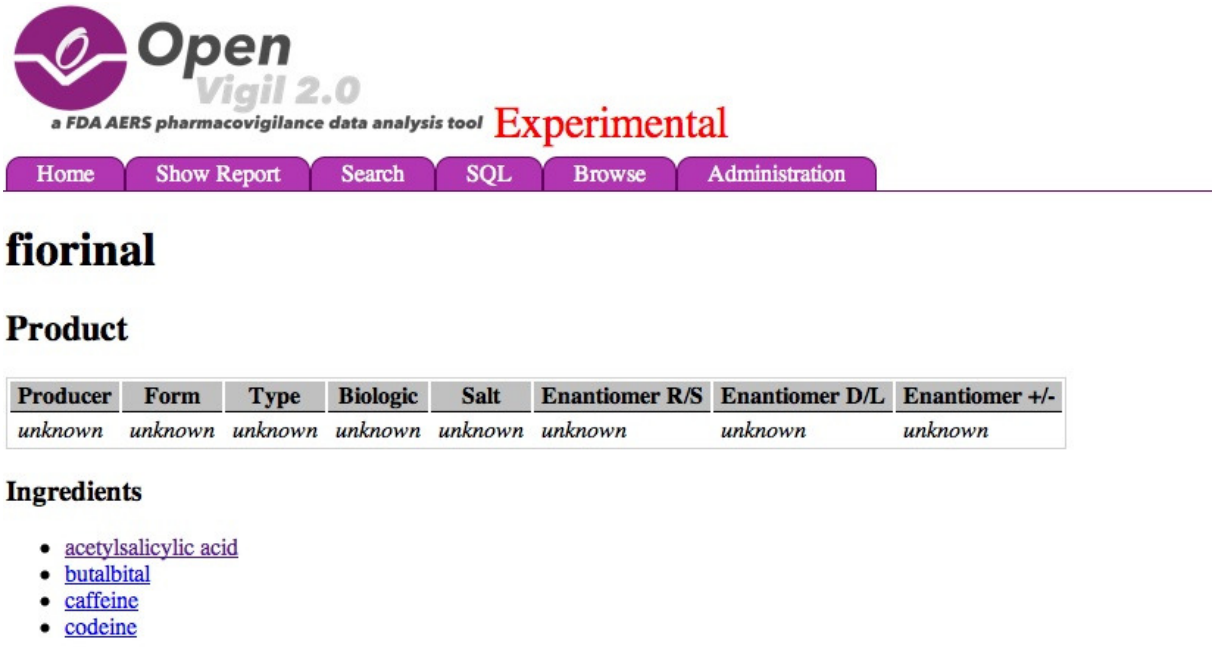
For further analysis choose “Browse” and “Products” in OpenVigil 2.0:



Result is a list of pharmaceutical products (“pharmaproducts”):

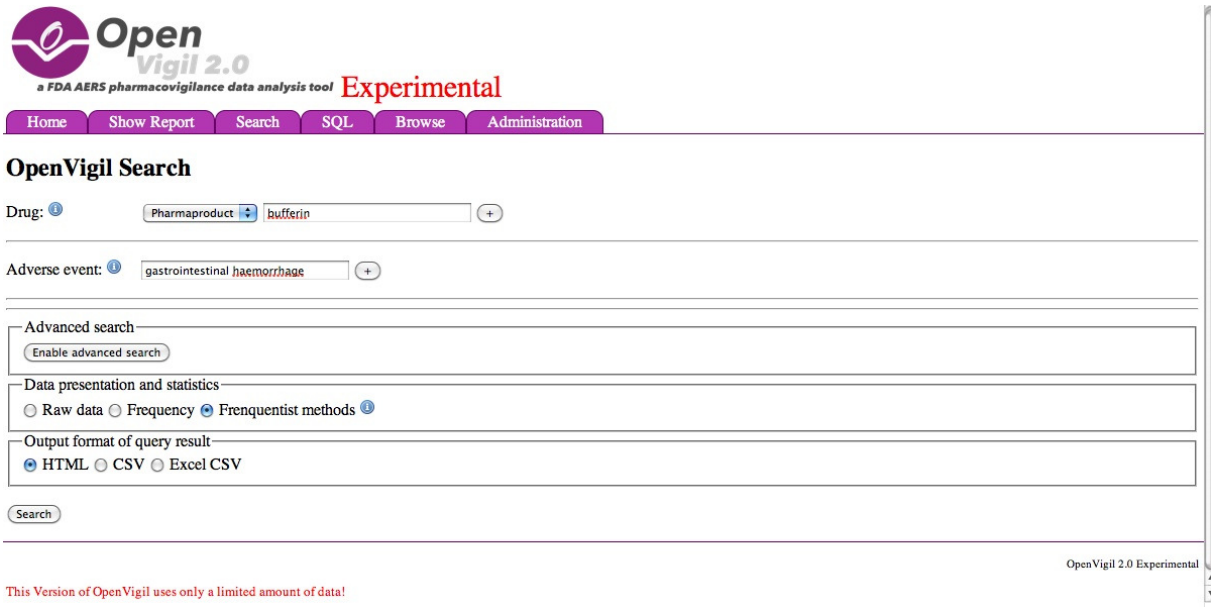


By clicking on a product, the drugs it consists of are shown:



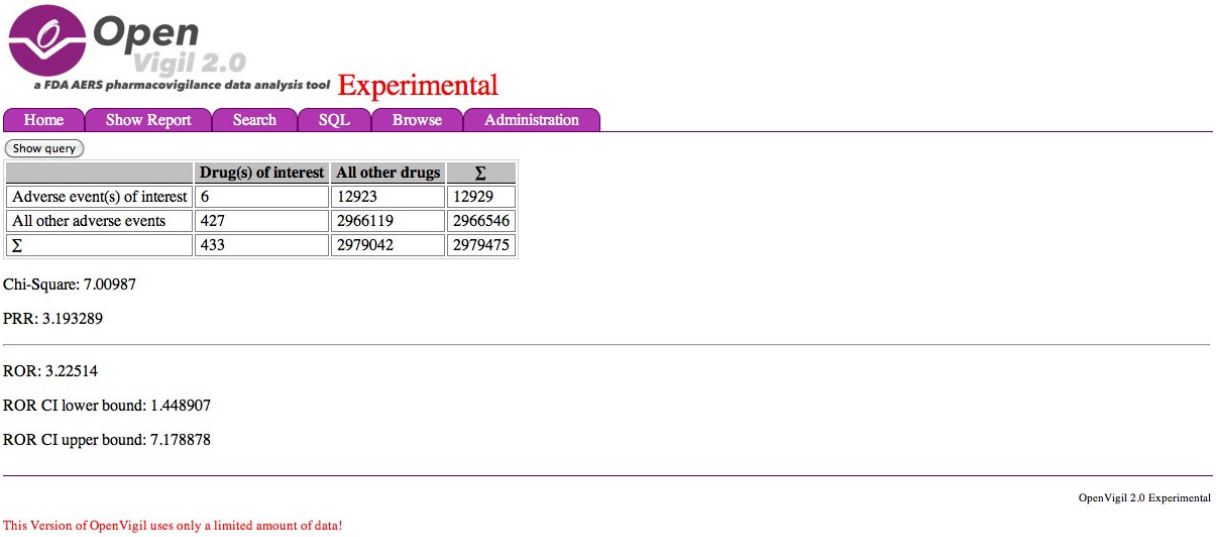
In this example Bufferin® and Ecotrin® are compared to each other. Both pharmaproducts contain no other drugs except acetylsalicylic acid and appear to be used with a similar frequency, extrapolated from the number of reports in the database.

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®:



Contingency table for Ecotrin®:

Open Vigil 2.0
a FDA AERS pharmacovigilance data analysis tool Experimental

Home | Show Report | Search | SQL | Browse | Administration

Show query

	Drug(s) of interest	All other drugs	Σ
Adverse event(s) of interest	13	12916	12929
All other adverse events	626	2965920	2966546
Σ	639	2978836	2979475

Chi-Square: 34.278924

PRR: 4.688321

ROR: 4.768705

ROR CI lower bound: 2.768951

ROR CI upper bound: 8.212693

OpenVigil 2.0 Experimental

This Version of OpenVigil uses only a limited amount of data!

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 %)[10]; 34.278924 for ecotrin).

1 0 https://people.richland.edu/james/lecture/m170/tbl-chi.html

The results of the two contingency tables can be merged in one table for further analysis (e.g., Fisher exact test, Chi-Squared test):

	Bufferin	**Ecotrin**	**All other drugs**	**Σ**
Gastrointestinal haemorrhage	6	13	12910	12929
All other adverse events	427	626	2965493	2966546
Σ	433	639	2978403	2979475

3.9. Compare OpenVigil 1 & 2 data (no. reports, PRR) to published data

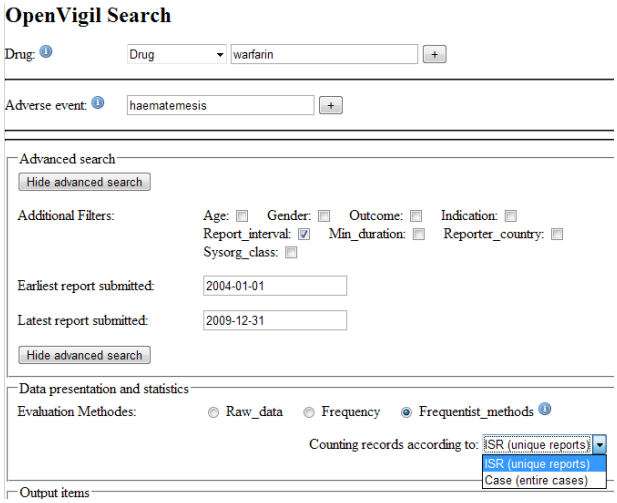
Introduction: This example stresses the importance of carefully checking any results obtained. Common pitfalls are

- 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?

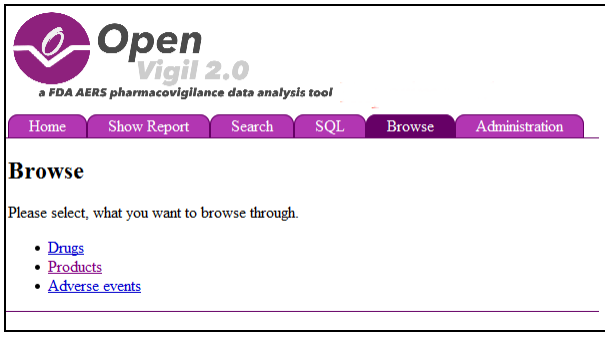
Query construction in OpenVigil 2: Enter "warfarin" as "drug" and "haematemesis" as adverse event, set the reporting date to between 2004 and 2009.
OpenVigil 2.0 can find 162 reports (out of 140 unique cases) and calculates – based on the counting of reports – a PRR of 3.109 and a ROR of 3.122. The latest 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.



Hint: You can also use OpenVigil 2 to learn more about drugs and pharmaproducts. Select ➔ Browse and ➔ Drugs to see a list of drugnames. Clicking on drug shows you the associated pharmaproducts (=brandnames).

Drugs named something like “WARFARIN 5 MG” are currently discarded in OpenVigil 2 since the the current version of OpenVigil 2 does not know what “5 MG” means. The misspelled “COUMADIN (WA***FR***ARIN 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 adjecent 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). 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).

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

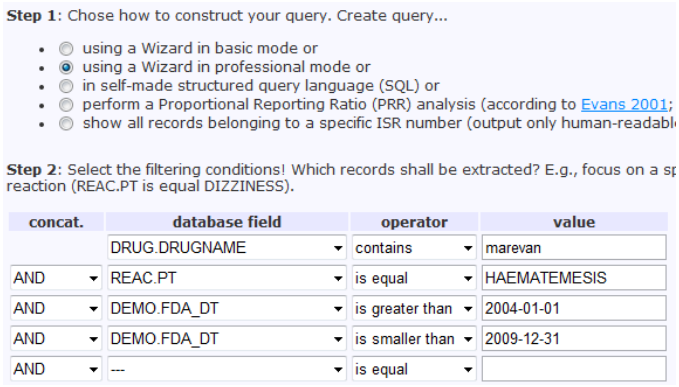
These number differ, reflecting

- 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(DEMO.ISR),COUNT(DISTINCT
DEMO.ISR),COUNT(DISTINCT DEMO.CASENO) FROM DRUG,REAC,DEMO 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
DEMO.FDA_DT >= "2004-01-01" AND DEMO.FDA_DT <= "2009-12-31") AND
DRUG.ISR=REAC.ISR AND DRUG.ISR=DEMO.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:

DRUGNAME	COUNT(DEMO.ISR)	COUNT(DISTINCT DEMO.ISR)	COUNT(DISTINCT DEMO.CASENO)
WARFARIN SODIUM	110	104	97
WARFARIN POTASSIUM	1	1	1
WARFARIN 5 MG	2	2	2
WARFARIN 2.5 MG COUMADIN	1	1	1
WARFARIN 2 MG	1	1	1
WARFARIN (WARFARIN)	1	1	1
WARFARIN (WARFARIN POTASSIUM)	1	1	1
WARFARIN 5 MG	1	1	1
WARFARIN 4MG	1	1	1
WARFARIN	29	29	25
WARAN (TABLETS)	1	1	1
WARAN	2	2	2
MAREVAN	7	7	6
JANTOVEN	1	1	1
COUMADIN	109	103	71

SQL query: SELECT DRUG.DRUGNAME,COUNT(DEMO.ISR),COUNT(DISTINCT DEMO.ISR),COUNT(DISTINCT DEMO.CASENO) FROM DRUG,REAC,DEMO 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 DEMO.FDA_DT >= "2004-01-01" AND DEMO.FDA_DT <= "2009-12-31") AND DRUG.ISR=REAC.ISR AND DRUG.ISR=DEMO.ISR AND (DEMO.DSRC!="DEMO04Q1" AND DEMO.DSRC!="DEMO04Q2" AND DEMO.DSRC!="DEMO04Q3" AND DEMO.DSRC!="DEMO04Q4" AND DEMO.DSRC!="DEMO05Q1" AND DEMO.DSRC!="DEMO05Q2") GROUP BY DRUG.DRUGNAME DESC;

DRUGNAME	#ISR	#uniqueISR	#uniqueCASE	%uniqueISR	%uniqueCASE
WARFARIN SODIUM	110	104	97	*94.55*	*93.27*
WARFARIN POTASSIUM	1	1	1	*100.00*	*100.00*
WARFARIN 5 MG	2	2	2	*100.00*	*100.00*
WARFARIN 2.5 MG COUMADIN	1	1	1	*100.00*	*100.00*
WARFARIN 2 MG	1	1	1	*100.00*	*100.00*
WARFARIN (WARFARIN)	1	1	1	*100.00*	*100.00*
POTASSIUM)	1	1	1	*100.00*	*100.00*
WARFARIN 5 MG	1	1	1	*100.00*	*100.00*
WARFARIN 4MG	1	1	1	*100.00*	*100.00*
WARFARIN	29	29	25	*100.00*	*86.21*
WARAN (TABLETS)	1	1	1	*100.00*	*100.00*
WARAN	2	2	2	*100.00*	*100.00*
MAREVAN	7	7	6	*100.00*	*85.71*
JANTOVEN	1	1	1	*100.00*	*100.00*
COUMADIN	109	103	71	*94.50*	*68.93*
SUMS	**268**	**256**	**212**	**95.52**	**82.81**

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:

DRUGNAME	COUNT(DEMO.ISR)	COUNT(DISTINCT DEMO.ISR)	COUNT(DISTINCT DEMO.CASENO)
COUMADIN	268	251	202

SQL query: SELECT DRUG.DRUGNAME,COUNT(DEMO.ISR),COUNT(DISTINCT DEMO.ISR),COUNT(DISTINCT DEMO.CASENO) FROM DRUG,REAC,DEMO 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 DEMO.FDA_DT >= "2004-01-01" AND DEMO.FDA_DT <= "2009-12-31") AND DRUG.ISR=REAC.ISR AND DRUG.ISR=DEMO.ISR AND (DEMO.DSRC!="DEMO04Q1" AND DEMO.DSRC!="DEMO04Q2" AND DEMO.DSRC!="DEMO04Q3" AND DEMO.DSRC!="DEMO04Q4" AND DEMO.DSRC!="DEMO05Q1" AND DEMO.DSRC!="DEMO05Q2")

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:

DRUGNAME	ISR	CASENO	FDA_DT
COUMADIN	4708034	5832365	2005-07-07
WARFARIN	4714857	5837903	2005-07-15
WARFARIN SODIUM	4714932	5829243	2005-07-13
WARFARIN	4727393	5851405	2005-07-27
WARFARIN SODIUM	4727406	5802551	2005-07-26
WARFARIN	4735657	5855347	2005-08-01
WARFARIN 5 MG	4742362	5861329	2005-08-10
WARFARIN	4746558	5837903	2005-08-17
COUMADIN	4748562	5865667	2005-08-18
WARFARIN (WARFARIN)	4751572	5868903	2005-08-22

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):

DRUGNAME	#ISR	#uniqueISR	#uniqueCASE	%uniqueISR	%uniqueCASE
WARFARIN SODIUM	170	159	145	*93.53*	*91.19*
WARFARIN SODIUIM (WARFRIN SODIUIM)	1	1	1	*100.00*	*100.00*
WARFARIN POTASSIUM	1	1	1	*100.00*	*100.00*
WARFARIN 5MG PO	1	1	1	*100.00*	*100.00*
WARFARIN 5 MG TAB	1	1	1	*100.00*	*100.00*
WARFARIN 5 MG BMS	1	1	1	*100.00*	*100.00*
WARFARIN 5 MG	3	3	3	*100.00*	*100.00*
WARFARIN 2.5 MG COUMADIN	1	1	1	*100.00*	*100.00*
WARFARIN 2 MG	1	1	1	*100.00*	*100.00*
WARFARIN (WARFARIN)	3	3	3	*100.00*	*100.00*
WARFARIN (WARFARIN POTASSIUM)	1	1	1	*100.00*	*100.00*
WARFARIN 5 MG	1	1	1	*100.00*	*100.00*
WARFARIN 4MG	1	1	1	*100.00*	*100.00*
WARFARIN 1MG	1	1	1	*100.00*	*100.00*
WARFARIN (WARFARIN POTASSIUM)	1	1	1	*100.00*	*100.00*
WARFARIN 1MG	1	1	1	*100.00*	*100.00*
WARFARIN	44	44	35	*100.00*	*79.55*
WARFARIIN SODIUM (WARFARIN SODIUM)	1	1	1	*100.00*	*100.00*
WARAN (TABLETS)	1	1	1	*100.00*	*100.00*
WARAN	2	2	2	*100.00*	*100.00*
MAREVAN	10	10	8	*100.00*	*80.00*
JANTOVEN	1	1	1	*100.00*	*100.00*
COUMADIN	165	152	104	*92.12*	*68.42*
SUMS	**413**	**389**	**316**	**94.19**	**81.23**

COUNT(DEMO.ISR)	COUNT(DISTINCT DEMO.ISR)	COUNT(DISTINCT DEMO.CASENO)
413	382	299

SQL query: SELECT COUNT(DEMO.ISR),COUNT(DISTINCT DEMO.ISR),COUNT(DISTINCT DEMO.CASENO) FROM DRUG,REAC,DEMO 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 DEMO.FDA_DT >= "2004-01-01" AND DEMO.FDA_DT <= "2009-12-31") AND DRUG.ISR=REAC.ISR AND DRUG.ISR=DEMO.ISR;

There appear to be 413 reports from 299 distinct cases.

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

```
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:

DRUGNAME	ISR	CASENO	FDA_DT	GNDR_COD	AGE	MFR_NUM
COUMADIN	4289963	4084394	2004-02-09	M	68	CA-ROCHE-356746
COUMADIN	4289963	4084394	2004-02-09	M	68	CA-ROCHE-356746
COUMADIN	4297398	4092980	2004-02-17	F	0	
COUMADIN	6437446	6630850	2009-11-12	F	31	US-BAXTER-2008BH004256
COUMADIN	6497918	7150125	2009-12-15	M	73	US-BAXTER-2009BH015627
WARFARIN	6509640	7230303	2009-12-21	F	64	

SQL query: SELECT DRUG.DRUGNAME,DEMO.ISR,DEMO.CASENO,DEMO.FDA_DT,DEMO.GNDR_COD,DEMO.AGE,DEMO.MFR_NUM FROM DRUG,REAC,DEMO 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 DEMO.FDA_DT >= "2004-01-01" AND DEMO.FDA_DT <= "2009-12-31") AND DRUG.ISR=REAC.ISR AND DRUG.ISR=DEMO.ISR;

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:

Source	**n (reports)**	**n (cases)**	**PRR**	**ROR (95%-CI)**
OpenVigil 1 GUI without DEMO data prior to 2005Q3	**268**, maybe more	not available	not available	not available
OpenVigil 1 SQL without DEMO data prior to 2005Q3	251	202	not calculated	not calculated
OpenVigil 1 SQL (full LAERS data)	382	299, a few less because of multiplicates	not calculated	not calculated
OpenVigil 2.0 GUI (default install)	162	140	3.109*	3.122 (2.676; 3.642)
OpenVigil 2.1 GUI (additional manual drugname mapping)	166	143	3.141 (reports) **3.505** (cases)	3.154 (reports) 3.522 (cases)
Sakaeda 2013	not reported	**268**	**1.991**	2.006 (1.778; 2.234)
*) all measurements of disproportionality were calculated on reports, not cases in OpenVigil 2.0. Congruence or marked disagreement are printed in **bold letters**.				

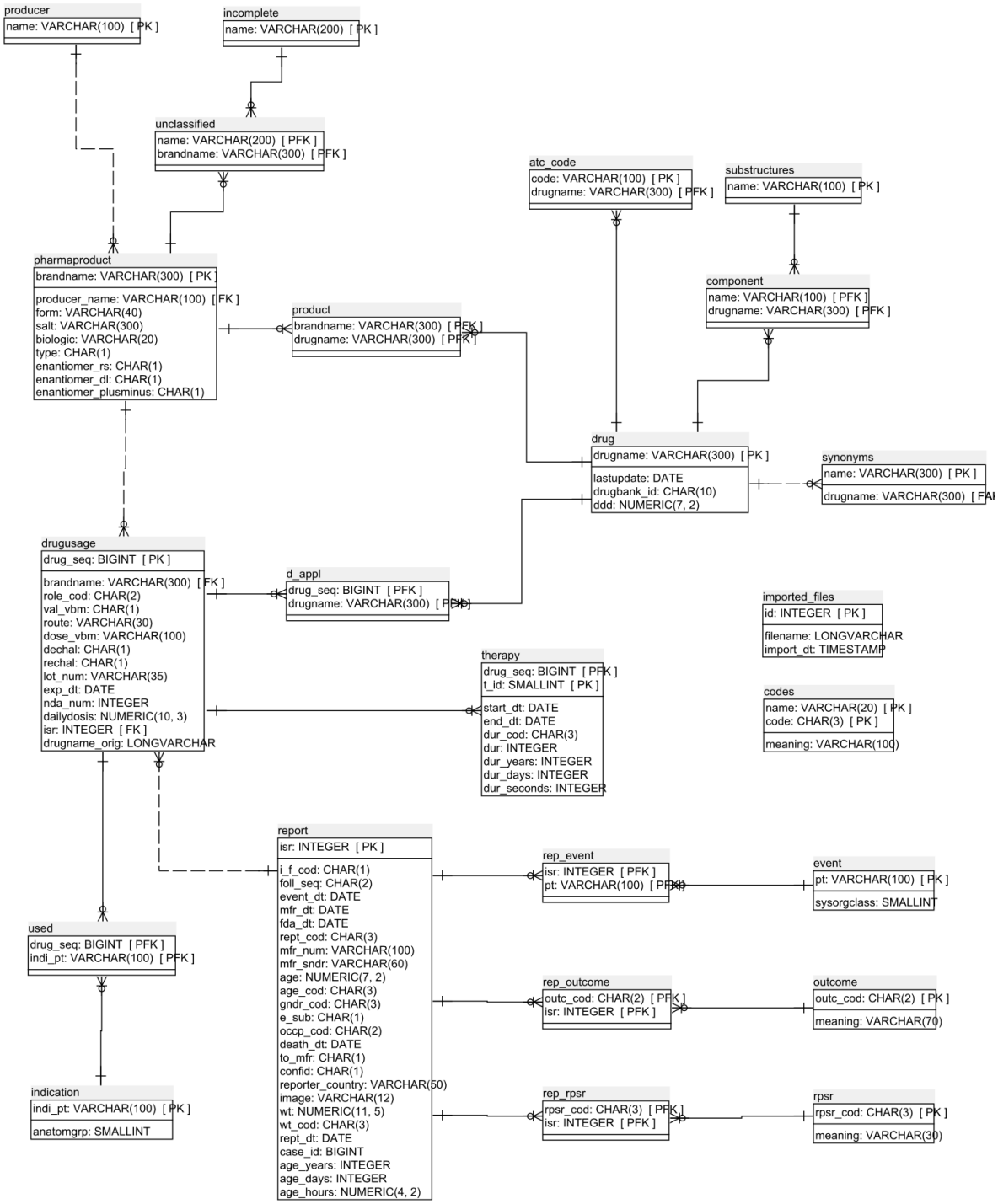
Conclusions:

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 thus not 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