OpenVigil 1 – Tutorial

Ruwen Böhm <ruwen.boehm@pharmakologie.uni-kiel.de>
Version 2015-12-02

1. Introduction

OpenVigil 1 (http://www.uni-kiel.de/pharmacology/pvt/openvigil.php/) is a pharmacovigilance data analysis tool. It is superseded by OpenVigil 2 (http://www.is.informatik.uni-kiel.de:8503/OpenVigil/) which is faster and more suited for data anaylsis since it operates on cleaned data. OpenVigil 1 is thus now depreciated for pharmacovigilance analyses but still maintained for exploring the raw data.

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.

Pharmacovigilance data source

The data currently used in OpenVigil 1 are taken from Adverse Event Reporting System (AERS) of the Food and Drug Administration (FDA) of the USA.

1 http://en.wikipedia.org/wiki/Pharmacovigilance

The advantage of the FDA source is a large amount of data due to the size of the reporting population. The disadvantage on the other hand is that reports in the 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).

Import errors are recorded in
http://www.uni-kiel.de/pharmacology/pvt/openvigil.php?cd=if.

FNAME	DT	NERR_PARSER	NERR_SQL	TERR_PARSER	TERR_SQL
DEMO09Q3	2014-09-09 14:20:05	2	0	Parser error: item counts (expected 23/found 11) at line 53920, file ascii/DEMO09Q3.TXT do not match: 6307398$6784218$I$$6307398-8$20040501$20081001$20090810$PER$US-JNJFOC-20081000321$JOHNSON + JOHNSON PHARMACEUTICAL Parser error: item counts (expected 23/found 13) at line 53921, file ascii/DEMO09Q3.TXT do not match: 13YRFN$$$20090713$CN$$$$UNITED STATES	

Fig. 1: Example of errors found in the import log

Citing OpenVigil

If referring to data extracted by OpenVigil, you must provide information on the version of OpenVigil used and the nature of the pharmacovigilance data (e.g., which files, were there errors?). OpenVigil 1 provides a quick overview of the installation that you are using: http://www.uni-kiel.de/pharmacology/pvt/openvigil127.php?cd=vs (cf. fig. 2). You should save this page and publish it with the results, for example as supplemental material.

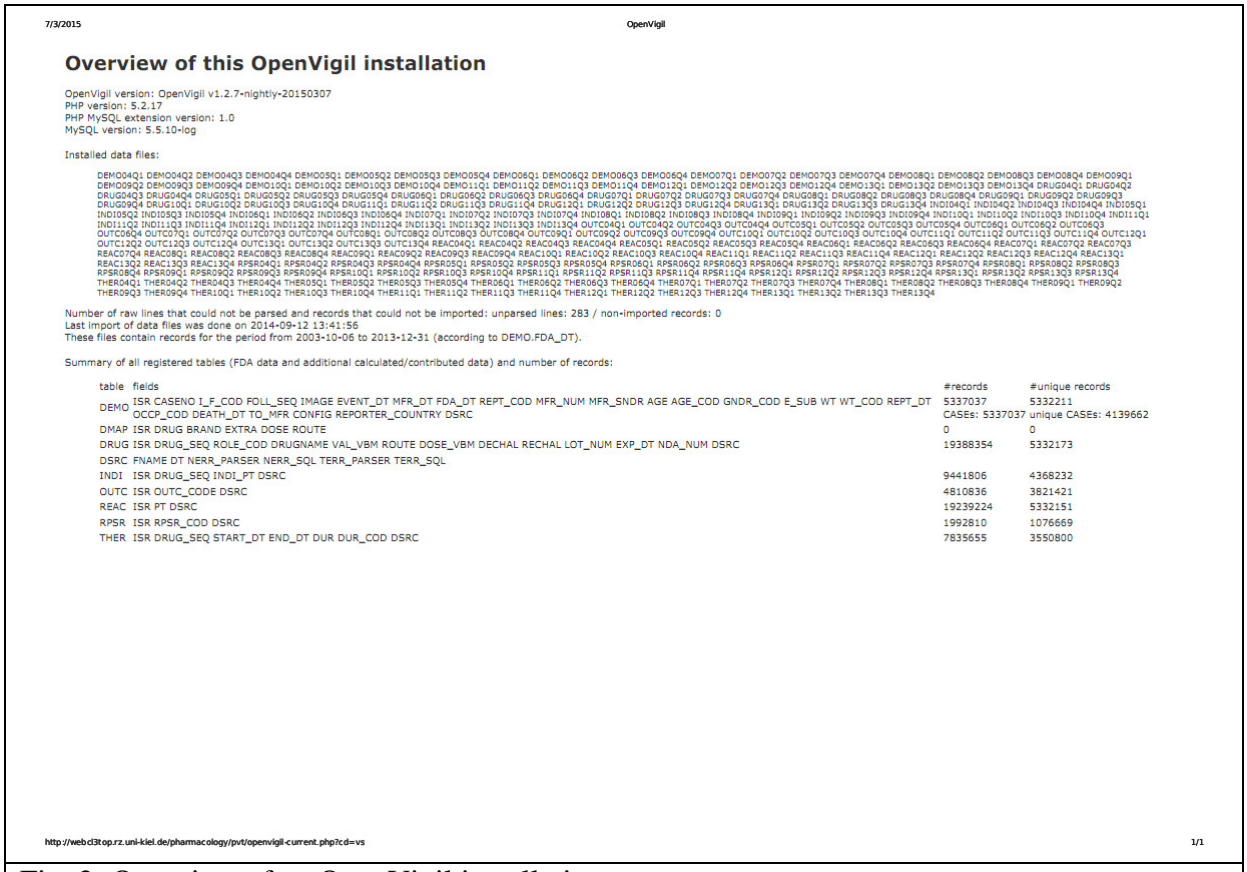


Fig. 2: Overview of an OpenVigil installation
Information on OpenVigil version and imported files and the resulting database are presented in a concise way.

"Drugname" (as used by OpenVigil 1)

OpenVigil 1 uses the data field drugname which was conceived by the FDA to hold a text string that describes the medication used in this report. It is not easily parseable by computer software. It may contain references to unknown or blinded study drugs. The last example in this tutorial shows some common problems and pitfalls.
„Drugname“ is different from the term “drug” which we use for a substance in a pharmaceutical product that is biologically active and responsible for the therapeutic effect. “Drug”, in turn, must not be confused with other meanings like illicit drugs or a ready-made pharmaceutical product like a pill, denoted by its brandname.
Because OpenVigil uses the U.S. American pharmacovigilance data, most 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

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.

Note that there are also ambigous reports like „WARFARIN BLINDED“ or „UNKNOWN“ that can never be resolved to something meaningful.
OpenVigil 1.2.7 introduces experimental drug-mapping via RXNORM. However, RXNORM will also map „WARFARIN BLINDED“ to „WARFARIN“, so be very, very careful!
OpenVigil 2 does only drugname-mapping of unambigous reports and is thus safer to use.
See the last example for the various pitfalls you can step into!

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

Graphical user interface

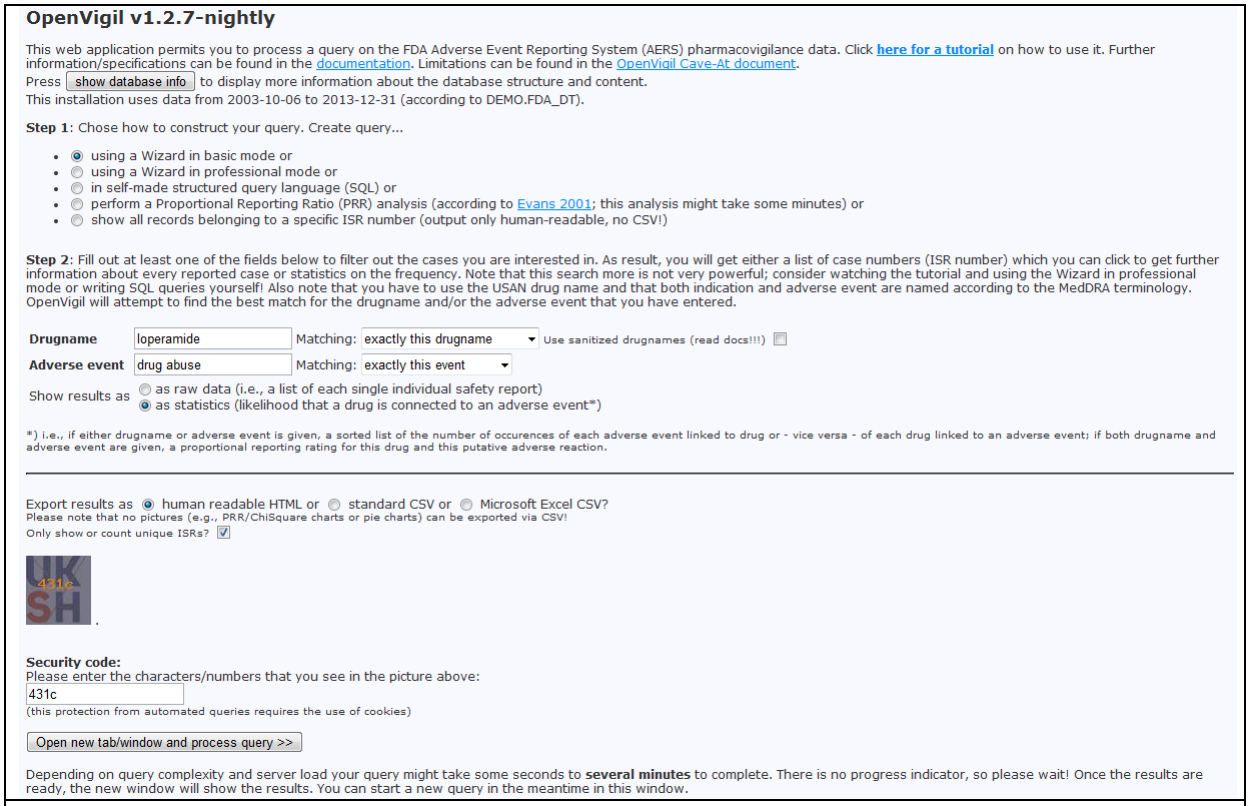


Fig 2: Screenshot for the OpenVigil 1.27 installation at university of Kiel
The first paragraph describes the data source: which reports were imported?
The user can chose between different search modes which will be explained below.
Most input like drugnames can be matched exactly (fast) or as substring of the database entry (slow, error prone).
Finally, the output can be raw data or statistics, as HTML or CSV file.
At the university of Kiel, we needed to protect the application with a simple CAPTCHA to stop automated spamming attempts.

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 DEMO LIMIT 10
# get the first 10 reports from the DEMO table (=demographic
  data)
```

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 1 to construct a query, the last line of your resultssheet will show the SQL query code which was generated for your query. You can use this code to build a more complex query on top of it.

Further reading

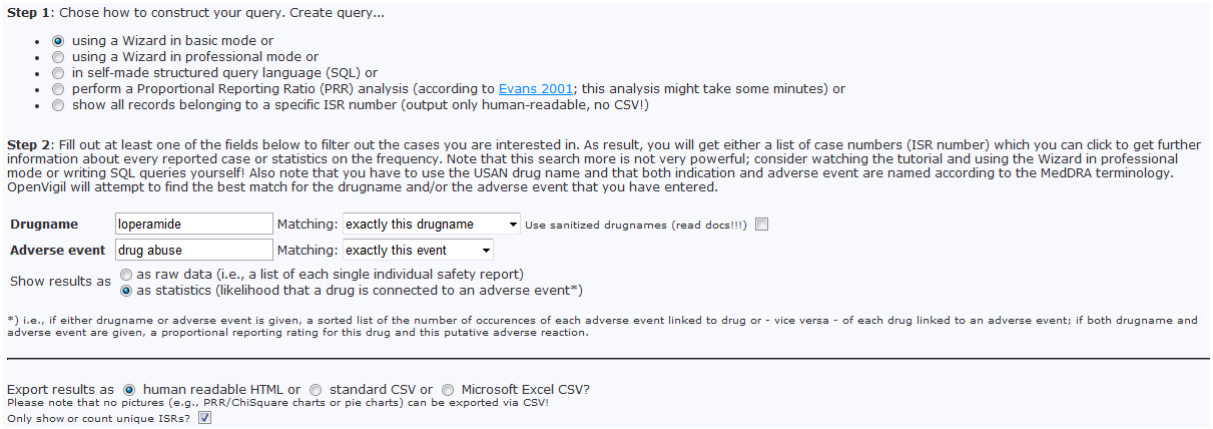
- Primer on disproportionality analysis
- OpenVigil 2 tutorial
- OpenVigil 2 data import flow

2. Examples

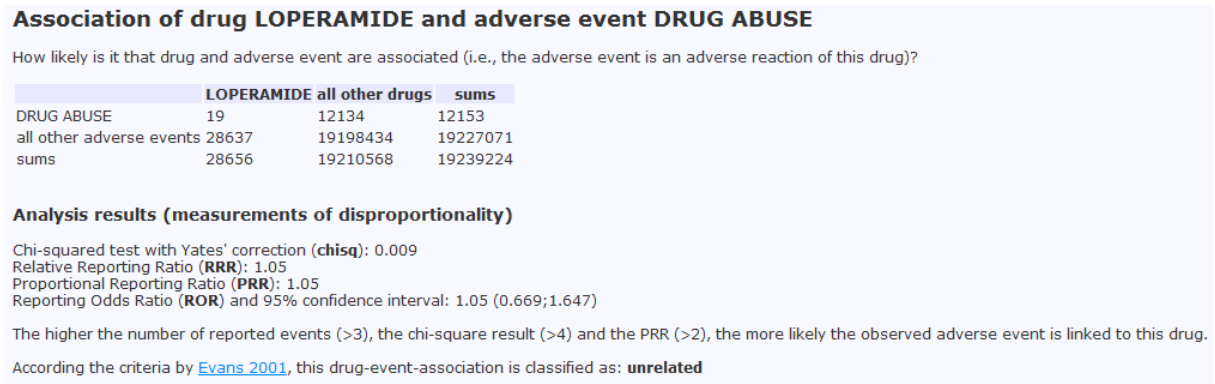
Basic disproportionality analysis: Is drug abuse an ADR of loperamide?

Problem: Is drug abuse an adverse reaction of the opioid loperamide?

Query construction: Wizard in basic mode; drugname is loperamide; adverse event is abuse; no raw data shall be reported but a statistical analysis (measurements of disproportionality) instead:



Results:



No, there are only few reports (19). The 2x2 contingency table does not show disproportionality (chisq 0.009). Other measurements of disproportionality indicate that is adverse event is not reported more frequently than with other drugs. Based on the number of

reports, chisq and PRR, OpenVigil checks whether Evans' criteria are fulfilled, i.e., whether the adverse event should be considered a true adverse drug reaction.

Cave #1: These numbers will change depending whether you use EXACT match or SIMILAR match and whether you use drugname-mapping or not. For meaningful results, use OpenVigil 2 instead of OpenVigil 1!

For example, if you use "SIMILAR" match for the drugname, you get

Association of drug LOPERAMIDE and adverse event DRUG ABUSE

How likely is it that drug and adverse event are associated (i.e., the adverse event is an adverse reaction of this drug)?

	LOPERAMIDE	all other drugs	sums
DRUG ABUSE	30	12123	12153
all other adverse events	57336	19169735	19227071
sums	57366	19181858	19239224

Analysis results (measurements of disproportionality)

Chi-squared test with Yates' correction (**chisq**): 0.912
Relative Reporting Ratio (**RRR**): 0.828
Proportional Reporting Ratio (**PRR**): 0.827
Reporting Odds Ratio (**ROR**) and 95% confidence interval: 0.827 (0.578;1.184)

The higher the number of reported events (>3), the chi-square result (>4) and the PRR (>2), the more likely the observed adverse event is linked to this drug.

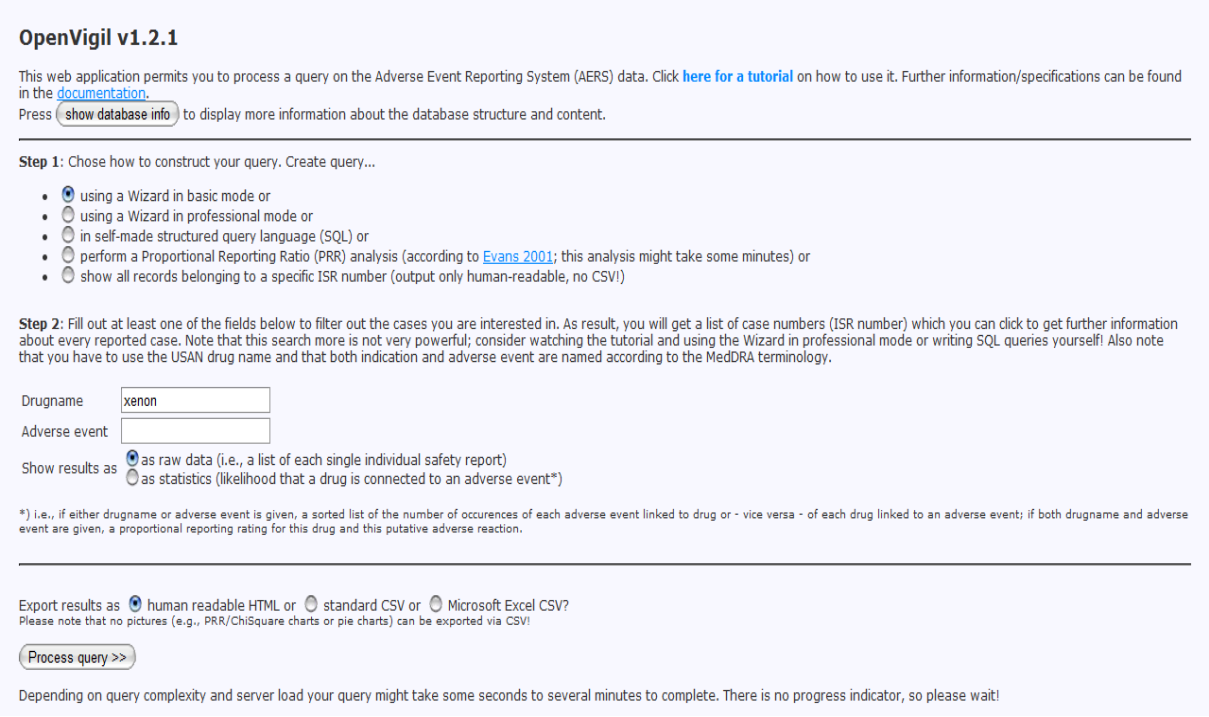
According the criteria by Evans 2001, this drug-event-association is classified as: **probably unrelated**

Cave #2: OpenVigil 1 counts all ISRs. The "Only show or count unique ISRs?" checkbox will not work in all circumstances and is of little use for post-2012Q4-data. Again, if you want meaningful numbers, use OpenVigil 2!

Raw data extraction: List all reports for a drug

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

Query constructions: Wizard in basic mode; drugname is “xenon”.



OpenVigil v1.2.1

This web application permits you to process a query on the Adverse Event Reporting System (AERS) data. Click **here for a tutorial** on how to use it. Further information/specifications can be found in the documentation.
Press show database info to display more information about the database structure and content.

Step 1: Chose how to construct your query. Create query...

- using a Wizard in basic mode or
- using a Wizard in professional mode or
- in self-made structured query language (SQL) or
- perform a Proportional Reporting Ratio (PRR) analysis (according to Evans 2001; this analysis might take some minutes) or
- show all records belonging to a specific ISR number (output only human-readable, no CSV!)

Step 2: Fill out at least one of the fields below to filter out the cases you are interested in. As result, you will get a list of case numbers (ISR number) which you can click to get further information about every reported case. Note that this search more is not very powerful; consider watching the tutorial and using the Wizard in professional mode or writing SQL queries yourself! Also note that you have to use the USAN drug name and that both indication and adverse event are named according to the MedDRA terminology.

Drugname: xenon
Adverse event:
Show results as: as raw data (i.e., a list of each single individual safety report) / as statistics (likelihood that a drug is connected to an adverse event*)

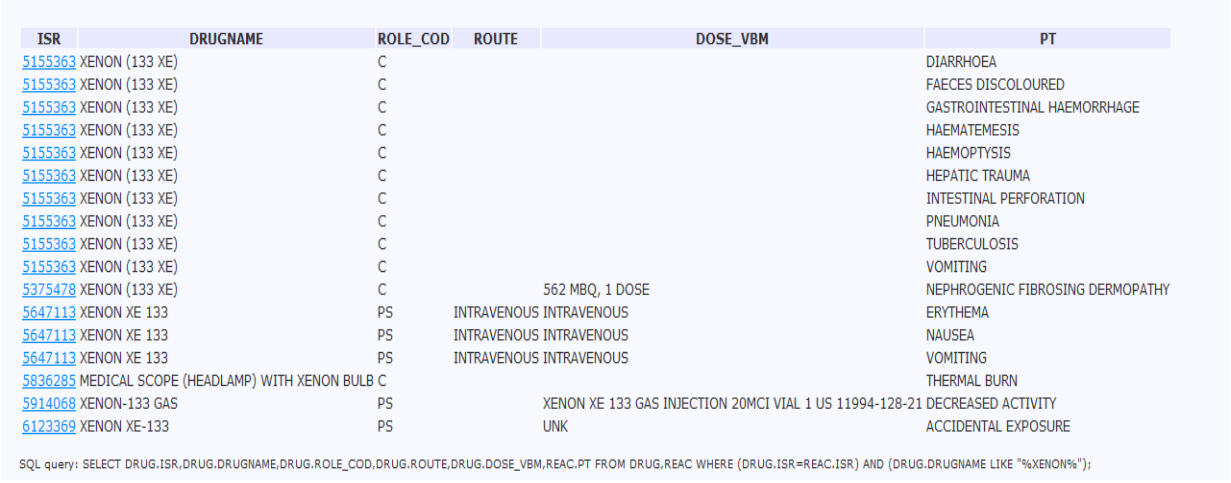
*) i.e., if either drugname or adverse event is given, a sorted list of the number of occurences of each adverse event linked to drug or - vice versa - of each drug linked to an adverse event; if both drugname and adverse event are given, a proportional reporting rating for this drug and this putative adverse reaction.

Export results as human readable HTML or standard CSV or Microsoft Excel CSV?
Please note that no pictures (e.g., PRR/ChiSquare charts or pie charts) can be exported via CSV!

Process query >>

Depending on query complexity and server load your query might take some seconds to several minutes to complete. There is no progress indicator, so please wait!

Results: A list of all reports; each single reported can be further analysed by clicking on the link in the ISR column.



ISR	DRUGNAME	ROLE_COD	ROUTE	DOSE_VBM	PT
5155363	XENON (133 XE)	C			DIARRHOEA
5155363	XENON (133 XE)	C			FAECES DISCOLOURED
5155363	XENON (133 XE)	C			GASTROINTESTINAL HAEMORRHAGE
5155363	XENON (133 XE)	C			HAEMATEMESIS
5155363	XENON (133 XE)	C			HAEMOPTYSIS
5155363	XENON (133 XE)	C			HEPATIC TRAUMA
5155363	XENON (133 XE)	C			INTESTINAL PERFORATION
5155363	XENON (133 XE)	C			PNEUMONIA
5155363	XENON (133 XE)	C			TUBERCULOSIS
5155363	XENON (133 XE)	C			VOMITING
5375478	XENON (133 XE)	C		562 MBQ, 1 DOSE	NEPHROGENIC FIBROSING DERMOPATHY
5647113	XENON XE 133	PS	INTRAVENOUS	INTRAVENOUS	ERYTHEMA
5647113	XENON XE 133	PS	INTRAVENOUS	INTRAVENOUS	NAUSEA
5647113	XENON XE 133	PS	INTRAVENOUS	INTRAVENOUS	VOMITING
5836285	MEDICAL SCOPE (HEADLAMP) WITH XENON BULB	C			THERMAL BURN
5914068	XENON-133 GAS	PS		XENON XE 133 GAS INJECTION 20MCI VIAL 1 US 11994-128-21	DECREASED ACTIVITY
6123369	XENON XE-133	PS		UNK	ACCIDENTAL EXPOSURE

SQL query: SELECT DRUG.ISR,DRUG.DRUGNAME,DRUG.ROLE_COD,DRUG.ROUTE,DRUG.DOSE_VBM,REAC.PT FROM DRUG,REAC WHERE (DRUG.ISR=REAC.ISR) AND (DRUG.DRUGNAME LIKE "%XENON%");

Individual Safety Report #5647113

The following is an extract from the different tables in the database that contain information to this individual safety report (ISR). E.g., the table DRUG contains information of all drugs given to the subject; the table REAC lists all observed adverse events. A description of the contents of the other tables and the meanings of the fields can be found in the chapter B (data element description) of this FDA document.

Showing all records in every data table concerning case #5647113...

Matching records in table "DEMO"

ISR	CASENO	I_F_COD	FOLL_SEQ	IMAGE	EVENT_DT	MFR_DT	FDA_DT	REPT_COD	MFR_NUM	MFR_SNDR	AGE	AGE_COD	GNDR_COD	E_SUB	WT	WT_COD	REPT_DT	OCCP_C
5647113	6575814	I	0	5647113-4	2008-02-14	0000-00-00	2008-02-28	DIR			25	YR		N	0		2008-02-28	

1 records found in this table.

Matching records in table "DRUG"

ISR	DRUG_SEQ	ROLE_COD	DRUGNAME	VAL_VBM	ROUTE	DOSE_VBM	DECHAL	RECHAL	LOT_NUM	EXP_DT	NDA_NUM
5647113	1009590245	PS	XENON XE 133	1	INTRAVENOUS	INTRAVENOUS	Y	D		0000-00-00	

1 records found in this table.

Matching records in table "INDI"

ISR	DRUG_SEQ	INDI_PT

There were no matching records in this table.

Matching records in table "OUTC"

ISR	OUTC_CODE

There were no matching records in this table.

Matching records in table "REAC"

ISR	PT
5647113	ERYTHEMA
5647113	NAUSEA
5647113	VOMITING

3 records found in this table.

Matching records in table "RPSR"

ISR	RPSR_COD

There were no matching records in this table.

Matching records in table "THER"

ISR	DRUG_SEQ	START_DT	END_DT	DUR	DUR_COD
5647113	1009590245	2008-02-14	2008-02-14	0	

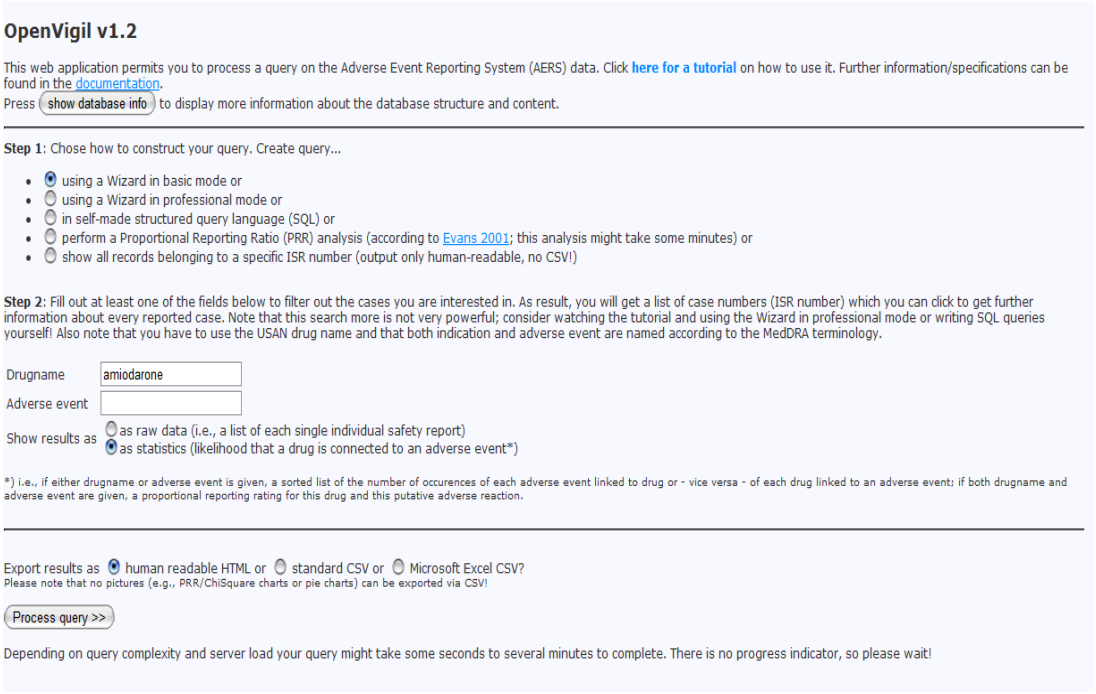
1 records found in this table.

Note the strange route of administration. Occasionally, single individual safety reports contain wrong data or a lot of data is missing. However, statistical analyses of large numbers of ISR are often quite correct and helpful.

Frequency: List AE sorted by frequency for a drug

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

Query construction: Wizard in basic mode; amiodarone as drugname; no raw data shall be reported but a list of occurrences of each adverse event.



OpenVigil v1.2

This web application permits you to process a query on the Adverse Event Reporting System (AERS) data. Click **here for a tutorial** on how to use it. Further information/specifications can be found in the documentation.

Press [show database info] to display more information about the database structure and content.

Step 1: Chose how to construct your query. Create query...

- (•) using a Wizard in basic mode or
- ( ) using a Wizard in professional mode or
- ( ) in self-made structured query language (SQL) or
- ( ) perform a Proportional Reporting Ratio (PRR) analysis (according to Evans 2001; this analysis might take some minutes) or
- ( ) show all records belonging to a specific ISR number (output only human-readable, no CSV!)

Step 2: Fill out at least one of the fields below to filter out the cases you are interested in. As result, you will get a list of case numbers (ISR number) which you can click to get further information about every reported case. Note that this search more is not very powerful; consider watching the tutorial and using the Wizard in professional mode or writing SQL queries yourself! Also note that you have to use the USAN drug name and that both indication and adverse event are named according to the MedDRA terminology.

Drugname: amiodarone

Adverse event:

Show results as: ( ) as raw data (i.e., a list of each single individual safety report) (•) as statistics (likelihood that a drug is connected to an adverse event*)

*) i.e., if either drugname or adverse event is given, a sorted list of the number of occurences of each adverse event linked to drug or - vice versa - of each drug linked to an adverse event; if both drugname and adverse event are given, a proportional reporting rating for this drug and this putative adverse reaction.

Export results as (•) human readable HTML or ( ) standard CSV or ( ) Microsoft Excel CSV?
Please note that no pictures (e.g., PRR/ChiSquare charts or pie charts) can be exported via CSV!

[Process query >>]

Depending on query complexity and server load your query might take some seconds to several minutes to complete. There is no progress indicator, so please wait!

Result:

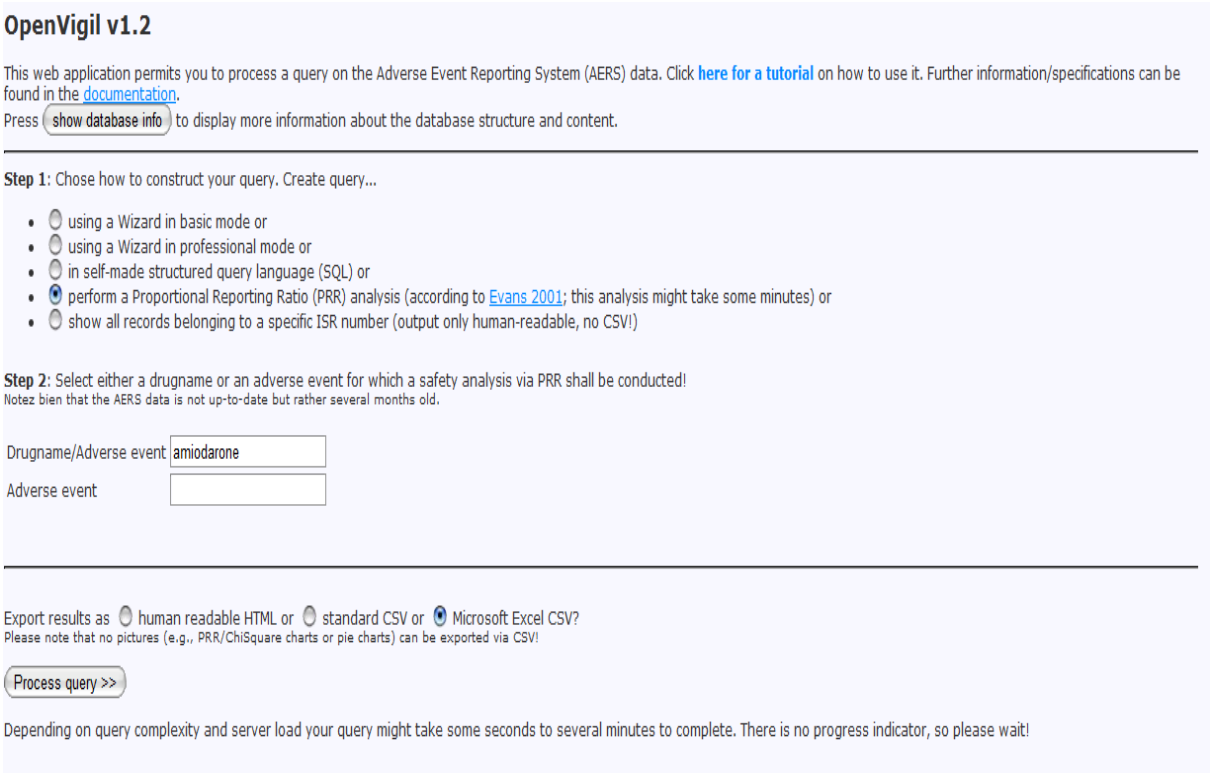
PT	COUNT(REAC.PT)
ATRIAL FIBRILLATION	1015
DYSPNOEA	852
DRUG INTERACTION	701
ASTHENIA	675
DIZZINESS	573
HYPOTENSION	570
CARDIAC FAILURE CONGESTIVE	555
NAUSEA	510
PNEUMONIA	504
FATIGUE	500
RENAL FAILURE ACUTE	493
VENTRICULAR TACHYCARDIA	476
FALL	447
BRADYCARDIA	436
INTERNATIONAL NORMALISED RATIO INCREASED	408
RENAL FAILURE	401
PLEURAL EFFUSION	400
ANAEMIA	394
OEDEMA PERIPHERAL	390
PAIN	366
VOMITING	358
CARDIAC ARREST	351
DRUG INEFFECTIVE	348
CONDITION AGGRAVATED	342
CARDIAC FAILURE	323
SYNCOPE	318
CHEST PAIN	302

Most reported is atrial fibrillation. Remember that these are just raw counts that have to be normalized to other drugs.

Disproportionality Analysis for all AEs and one drug

Problem: How likely is it that the results shown in example #3 are truly adverse reactions specific to amiodarone usage?

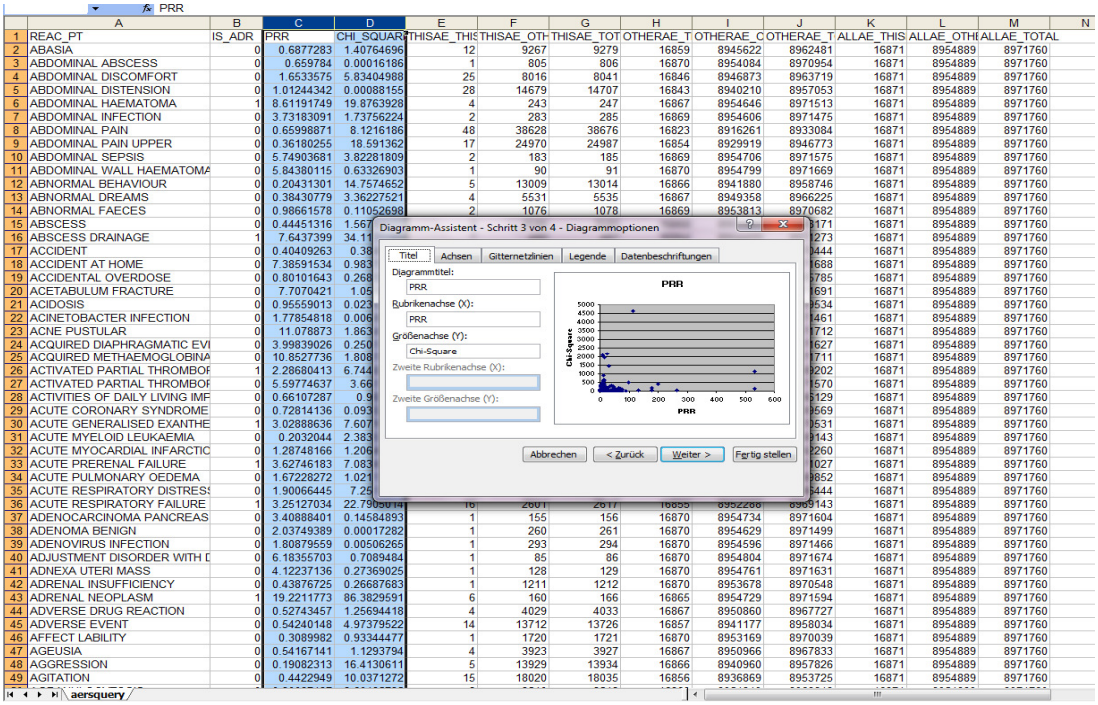
Query construction: Proportional reporting ratio (PRR) analysis (a type of disproportionality analysis) of amiodarone as drugname; export of the resulting list to Microsoft Excel for further analysis and visualisation.



Result:

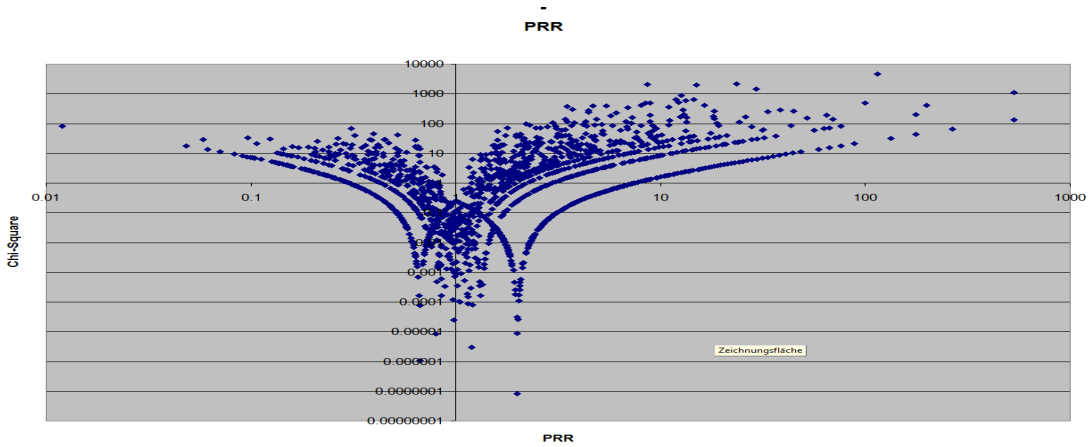
(This calculation might take several minutes!)

The list can be imported into Excel as horizontalised table:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	REAC_PT	IS_ADR	PRR	CHI_SQUARE	THISAE_THIS	THISAE_OTH	THISAE_TOT	OTHERAE_T	OTHERAE_C	OTHERAE_T	ALLAE_THIS	ALLAE_OTHE	ALLAE_TOTAL	
2	ABASIA	0	0.6877283	1.40764696	12	9267	9279	16859	8945622	8962481	16871	8954889	8971760	
3	ABDOMINAL ABSCESS	0	0.659784	0.00016186	1	805	806	16870	8954084	8970954	16871	8954889	8971760	
4	ABDOMINAL DISCOMFORT	0	1.6533575	5.83404988	25	8016	8041	16846	8946873	8963719	16871	8954889	8971760	
5	ABDOMINAL DISTENSION	0	1.01244342	0.00088155	28	14679	14707	16843	8940210	8957053	16871	8954889	8971760	
6	ABDOMINAL HAEMATOMA	1	8.61191749	19.8763928	4	243	247	16867	8954646	8971513	16871	8954889	8971760	
7	ABDOMINAL INFECTION	0	3.73183091	1.73756224	2	283	285	16869	8954606	8971475	16871	8954889	8971760	
8	ABDOMINAL PAIN	0	0.65998871	8.1216186	48	38628	38676	16823	8916261	8933084	16871	8954889	8971760	
9	ABDOMINAL PAIN UPPER	0	0.36180255	18.591362	17	24970	24987	16854	8929919	8946773	16871	8954889	8971760	
10	ABDOMINAL SEPSIS	0	5.74903681	3.82281809	2	183	185	16869	8954706	8971575	16871	8954889	8971760	
11	ABDOMINAL WALL HAEMATOMA	0	5.84380115	0.63326903	1	90	91	16870	8954799	8971669	16871	8954889	8971760	
12	ABNORMAL BEHAVIOUR	0	0.20431301	14.7574652	5	13009	13014	16866	8941880	8958746	16871	8954889	8971760	
13	ABNORMAL DREAMS	0	0.38430779	3.36227521	4	5531	5535	16867	8949358	8966225	16871	8954889	8971760	
14	ABNORMAL FAECES	0	0.98661578	0.11052698	2	1076	1078	16869	8953813	8970682	16871	8954889	8971760	
15	ABSCESS	0	0.44451316	1.567						171	16871	8954889	8971760	
16	ABSCESS DRAINAGE	1	7.6437399	34.11						273	16871	8954889	8971760	
17	ACCIDENT	0	0.40409263	0.38						444	16871	8954889	8971760	
18	ACCIDENT AT HOME	0	7.38591534	0.983						688	16871	8954889	8971760	
19	ACCIDENTAL OVERDOSE	0	0.80101643	0.268						785	16871	8954889	8971760	
20	ACETABULUM FRACTURE	0	7.7070421	1.05						691	16871	8954889	8971760	
21	ACIDOSIS	0	0.95559013	0.023						534	16871	8954889	8971760	
22	ACINETOBACTER INFECTION	0	1.77854818	0.006						461	16871	8954889	8971760	
23	ACNE PUSTULAR	0	11.078873	1.863						712	16871	8954889	8971760	
24	ACQUIRED DIAPHRAGMATIC EVI	0	3.99839026	0.250						627	16871	8954889	8971760	
25	ACQUIRED METHAEMOGLOBINA	0	10.8527736	1.808						711	16871	8954889	8971760	
26	ACTIVATED PARTIAL THROMBOF	1	2.28680413	6.744						202	16871	8954889	8971760	
27	ACTIVATED PARTIAL THROMBOF	0	5.59774637	3.66						570	16871	8954889	8971760	
28	ACTIVITIES OF DAILY LIVING IMF	0	0.66107287	0.9						129	16871	8954889	8971760	
29	ACUTE CORONARY SYNDROME	0	0.72814136	0.093						569	16871	8954889	8971760	
30	ACUTE GENERALISED EXANTHE	1	3.02888636	7.607						531	16871	8954889	8971760	
31	ACUTE MYELOID LEUKAEMIA	0	0.2032044	2.383						143	16871	8954889	8971760	
32	ACUTE MYOCARDIAL INFARCTIC	0	1.28748166	1.206						260	16871	8954889	8971760	
33	ACUTE PRERENAL FAILURE	1	3.62746183	7.083						027	16871	8954889	8971760	
34	ACUTE PULMONARY OEDEMA	0	1.67228272	1.021						852	16871	8954889	8971760	
35	ACUTE RESPIRATORY DISTRESS	0	1.90066445	7.25						444	16871	8954889	8971760	
36	ACUTE RESPIRATORY FAILURE	1	3.25127034	22.7905014	16	2601	2617	16855	8952288	8969143	16871	8954889	8971760	
37	ADENOCARCINOMA PANCREAS	0	3.40888401	0.14584893	1	155	156	16870	8954734	8971604	16871	8954889	8971760	
38	ADENOMA BENIGN	0	2.03749389	0.00017282	1	260	261	16870	8954629	8971499	16871	8954889	8971760	
39	ADENOVIRUS INFECTION	0	1.80879559	0.00506265	1	293	294	16870	8954596	8971466	16871	8954889	8971760	
40	ADJUSTMENT DISORDER WITH D	0	6.18355703	0.7089484	1	85	86	16870	8954804	8971674	16871	8954889	8971760	
41	ADNEXA UTERI MASS	0	4.12237136	0.27369025	1	128	129	16870	8954761	8971631	16871	8954889	8971760	
42	ADRENAL INSUFFICIENCY	0	0.43876725	0.26687683	1	1211	1212	16870	8953678	8970548	16871	8954889	8971760	
43	ADRENAL NEOPLASM	1	19.2211773	86.3829591	6	160	166	16865	8954729	8971594	16871	8954889	8971760	
44	ADVERSE DRUG REACTION	0	0.52743457	1.25694418	4	4029	4033	16867	8950860	8967727	16871	8954889	8971760	
45	ADVERSE EVENT	0	0.54240148	4.97379522	14	13712	13726	16857	8941177	8958034	16871	8954889	8971760	
46	AFFECT LABILITY	0	0.3089982	0.93344477	1	1720	1721	16870	8953169	8970039	16871	8954889	8971760	
47	AGEUSIA	0	0.54167141	1.1293794	4	3923	3927	16867	8950966	8967833	16871	8954889	8971760	
48	AGGRESSION	0	0.19082313	16.4130611	5	13929	13934	16866	8940960	8957826	16871	8954889	8971760	
49	AGITATION	0	0.4422949	10.0371272	15	18020	18035	16856	8936869	8953725	16871	8954889	8971760	

Changing the scale to logarithmic gives the final PRR graph:



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

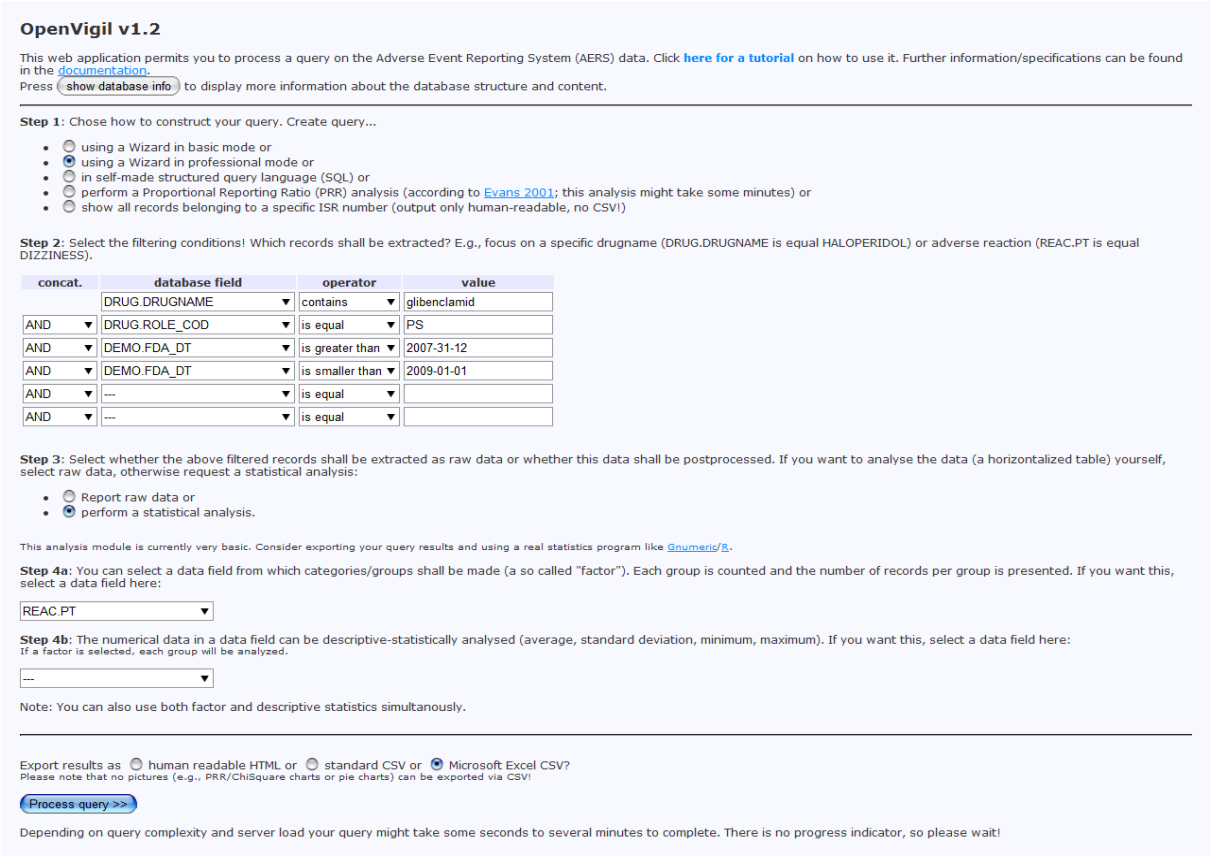
In the raw data, we can see that atrial fibrillation is very likely an adverse reaction:

REAC.PT	IS_ADR	PRR	CHI_SQUARE
ATRIAL FIBRILLATION	1	8.65757592	2059.51764

Extract data with special filters, e.g., date range

Problem: How many hypoglycaemic adverse events are reported for glibenclamide (USAN glyburide) in the year 2008 with glibenclamide as primary suspect for this event? How many in total?

Query construction: Wizard in professional mode; DRUG.DRUGNAME contains glibenclamid (catching glibenclamide, too; a second search for glyburide is needed); DRUG.ROLE_COD is PS (primary suspect); reporting date to the FDA (DEMO.FDA_DT) shall be within 2008 (>2007-12-31, <2009-01-01)
Output shall be counts (=statistical analysis) of the different adverse events (REAC.PT) as Microsoft Excel file.



Results: An Excel document with two columns: Name of the events and count of events.

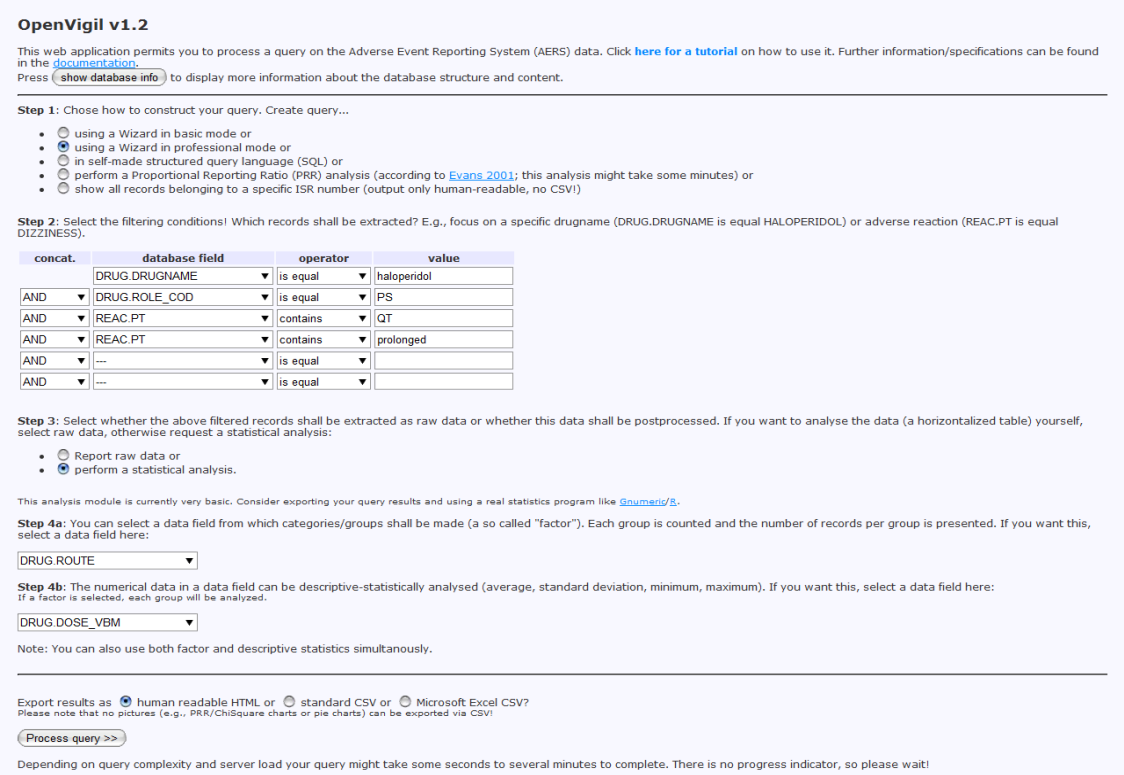
B42 =SUMME(B2:B40)

	A	B	C
1	PT	COUNT(REAC.PT)	
2	METABOLIC ACIDOSIS	6	
3	RENAL FAILURE ACUTE	5	
4	LACTIC ACIDOSIS	3	
5	HYPOTENSION	2	
6	INTENTIONAL OVERDOSE	2	
7	CONFUSIONAL STATE	2	
8	DRUG TOXICITY	2	
9	STILLBIRTH	2	
10	GRANULOMATOUS LIVER DISEASE	2	
11	COMPLETED SUICIDE	2	
12	DEHYDRATION	2	
13	CARDIAC ARREST	2	
14	HAEMODIALYSIS	2	
15	TARDIVE DYSKINESIA	1	
16	PANCREATITIS ACUTE	1	
17	ATRIAL FIBRILLATION	1	
18	RENAL FUNCTION TEST ABNORMAL	1	
19	STEVENS-JOHNSON SYNDROME	1	
20	PLACENTAL INFARCTION	1	
21	MALAISE	1	
22	NEONATAL DIABETES MELLITUS	1	
23	CHEST PAIN	1	
24	FOETAL DISORDER	1	
25	MUCOSAL DRYNESS	1	
26	ACUTE MYOCARDIAL INFARCTION	1	
27	RECTAL ADENOMA	1	
28	THYROID FUNCTION TEST ABNORMAL	1	
29	**HYPOGLYCAEMIA**	**1**	
30	VISUAL ACUITY REDUCED	1	
31	TOXIC EPIDERMAL NECROLYSIS	1	
32	SWELLING FACE	1	
33	MYOCARDIAL INFARCTION	1	
34	NEONATAL HYPOXIA	1	
35	HYPERHIDROSIS	1	
36	MATERNAL DRUGS AFFECTING FOETUS	1	
37	SOMNOLENCE	1	
38	CARDIOMEGALY	1	
39	HYPERVENTILATION	1	
40	BLOOD GLUCOSE DECREASED	1	
41			
42	**sum**	**60**	
43			

Extract certain data items, e.g., route of administration and dosage

Problem: How many QT prolongations are reported for haloperidol? How many were observed after intravenous, how many after different routes of administration? What was the average dosage for each route of administration?

Query construction: Wizard in professional mode; DRUG.DRUGNAME is “haloperidol”; REAC.PT contains both “QT” and “prolonged”; report statistics of DRUG.DOSE_VBM using factor DRUG.ROUTE



OpenVigil v1.2

This web application permits you to process a query on the Adverse Event Reporting System (AERS) data. Click **here for a tutorial** on how to use it. Further information/specifications can be found in the documentation.
Press show database info to display more information about the database structure and content.

Step 1: Chose how to construct your query. Create query...

- using a Wizard in basic mode or
- using a Wizard in professional mode or
- in self-made structured query language (SQL) or
- perform a Proportional Reporting Ratio (PRR) analysis (according to Evans 2001; this analysis might take some minutes) or
- show all records belonging to a specific ISR number (output only human-readable, no CSV!)

Step 2: Select the filtering conditions! Which records shall be extracted? E.g., focus on a specific drugname (DRUG.DRUGNAME is equal HALOPERIDOL) or adverse reaction (REAC.PT is equal DIZZINESS).

concat.	database field	operator	value
	DRUG.DRUGNAME	is equal	haloperidol
AND	DRUG.ROLE_COD	is equal	PS
AND	REAC.PT	contains	QT
AND	REAC.PT	contains	prolonged
AND	---	is equal	
AND	---	is equal	

Step 3: Select whether the above filtered records shall be extracted as raw data or whether this data shall be postprocessed. If you want to analyse the data (a horizontalized table) yourself, select raw data, otherwise request a statistical analysis:

- Report raw data or
- perform a statistical analysis.

This analysis module is currently very basic. Consider exporting your query results and using a real statistics program like Gnumeric/R.

Step 4a: You can select a data field from which categories/groups shall be made (a so called "factor"). Each group is counted and the number of records per group is presented. If you want this, select a data field here:

DRUG.ROUTE

Step 4b: The numerical data in a data field can be descriptive-statistically analysed (average, standard deviation, minimum, maximum). If you want this, select a data field here:
If a factor is selected, each group will be analyzed.

DRUG.DOSE_VBM

Note: You can also use both factor and descriptive statistics simultanously.

Export results as human readable HTML or standard CSV or Microsoft Excel CSV?
Please note that no pictures (e.g., PRR/ChiSquare charts or pie charts) can be exported via CSV!

Process query >>

Depending on query complexity and server load your query might take some seconds to several minutes to complete. There is no progress indicator, so please wait!

Results:

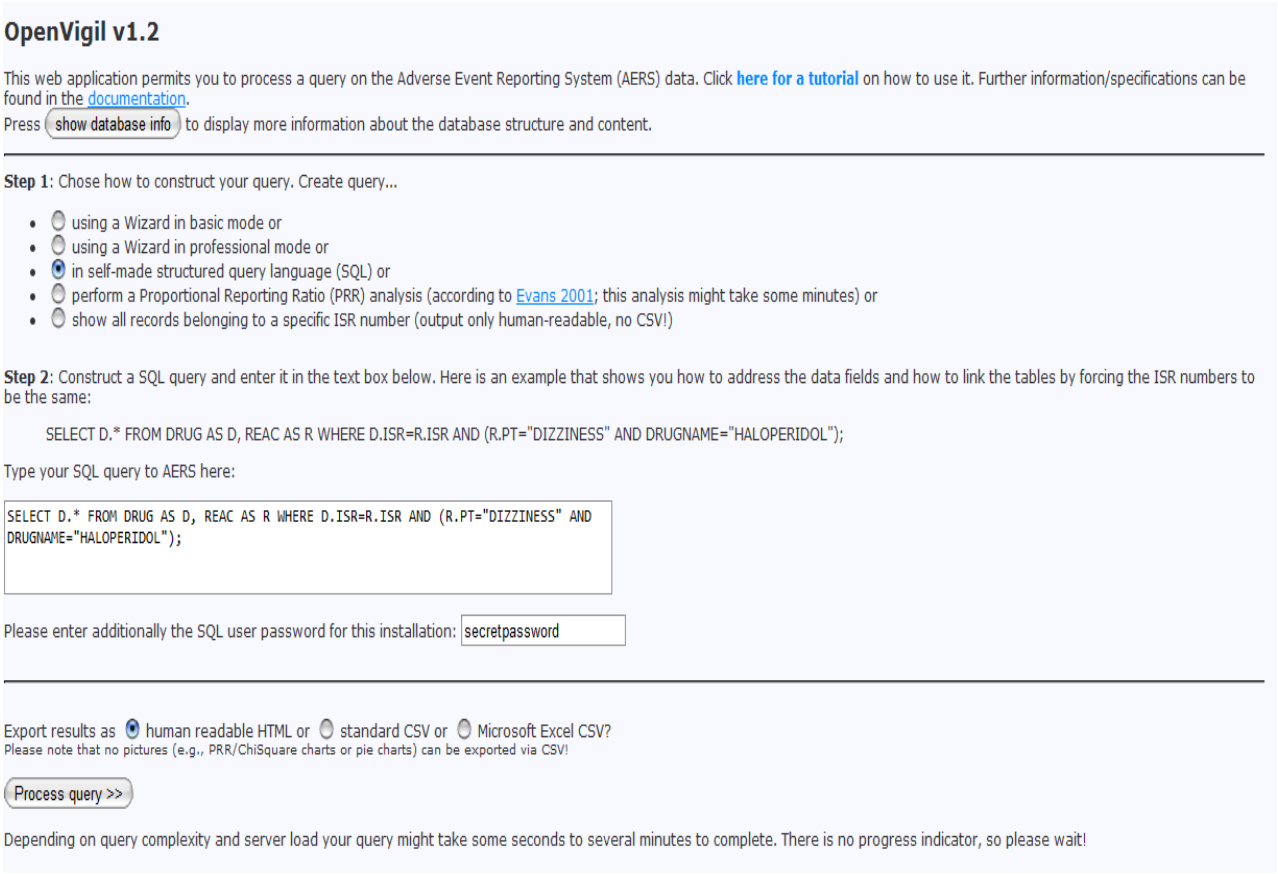
ROUTE	COUNT(DRUG.ROUTE)	AVG(DRUG.DOSE_VBM)	STD(DRUG.DOSE_VBM)	MIN(DRUG.DOSE_VBM)	MAX(DRUG.DOSE_VBI
UNKNOWN	19	0	0		
INTRAVENOUS	12	31.375	93.09563295343126		OVER 20 HOURS
	7	1.4285714285714286	2.2587697572631282		5 MG, QD
INTRAMUSCULAR	3	3.6666666666666665	4.4969125210773475		10 MG IM X 2
ORAL	3	0	0		PO
INTRAVENOUS BOLUS	1	2.5	0	2.5 MG IV BOLUS, 2 DOSES	2.5 MG IV BOLUS, 2 DOS

Note that DRUG.DOSE_VBM is a free-text field and thus calculating the average, min, max etc. is not possible. Data in DRUG.ROUTE is missing in some cases.

Customized filters and data item extraction via SQL

Problem: A very complex query was constructed that cannot be created with any of the available Wizards.

Query construction: If direct passing of structure query language (SQL) queries is enabled, the query written in SQL:



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

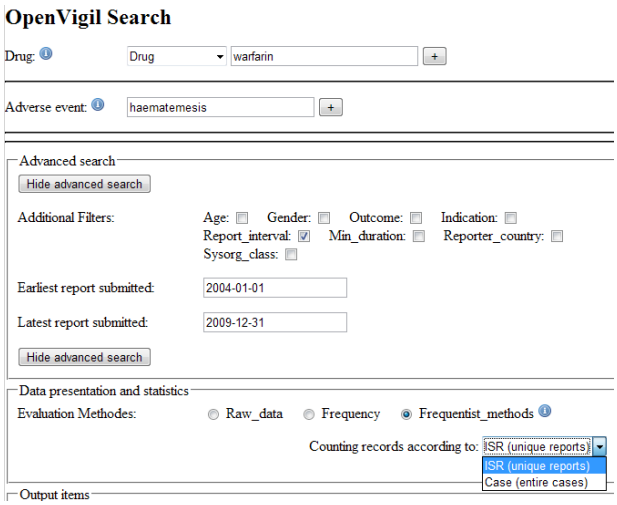
Introduction: This example stresses the importance of carefully checking any results obtained. Common pitfalls are

- 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.

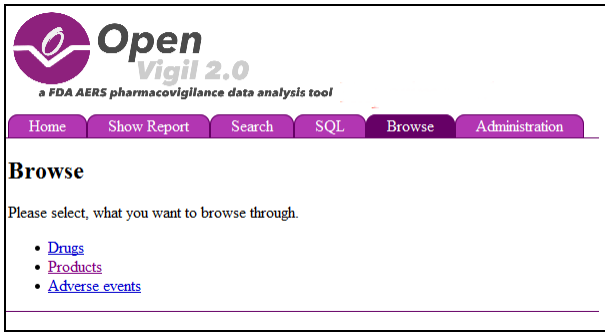


One first glance, both results appear way off: Too few reports and to few cases were found and the measurements of disproportionality indicate a rather strong association (i.e., haematemesis appears to be a real adverse reaction to warfarin). This in contrast to Sakaeda whose numbers do not fulfil Evans' 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

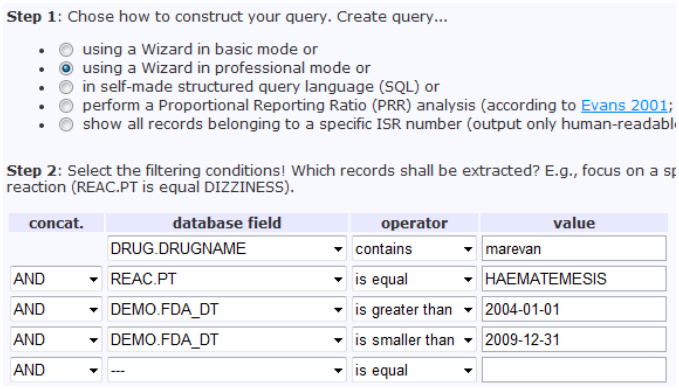
EVENT_DT	1655915	1653317	1184848
MFR_DT	2180288	2176768	1584290
FDA_DT minus data files DEMO04Q1 till DEMO05Q2	1805798	1803719	1331082
Sakaeda 2013	2231029	not provided	1644220
raw line count (minus headers)	2234931	n/a	n/a

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-2009BH004256
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.