Primer on Disproportionality Analysis

Version: 2018-10-16
Author: Ruwen Böhm <ruwen.boehm@pharmakologie.uni-kiel.de>

Contingency Tables 2
Number of cases ("DE") 2
χ^2 (chi squared) with Yates' correction 3
Relativ Reporting Ratio (RRR) 3
Confidence intervals for RRR 3
Proportional Reporting Ratio (PRR) 4
Confidence intervals for PRR 4
Reporting Odds Ratio (ROR) 4
Confidence Intervals for ROR 5
Non-frequentist, Bayesian methods 5
Multi-item associations 6
Example 1 6
Multi-item-association analysis methods 7
References 8

Abstract

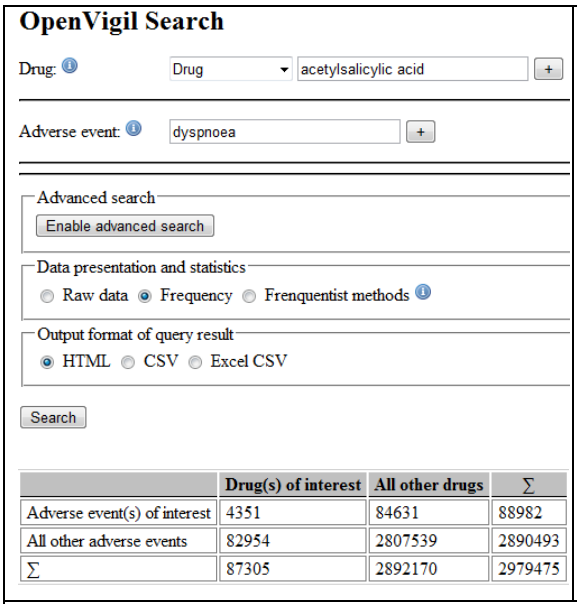
This primer explains meaning, calculation and uses of various methods for assessing disproportionality in pharmacovigilance data by observed-expected ratios. Disproportionality can stimulate further research whether an adverse event (AE) should be considered an adverse drug reaction (ADR). Disproportionality analysis is thus **only suited for hypothesis generation**, not for hypothesis testing. Note the cave-at documents from OpenVigil, FDA and WHO before drawing any conclusions from the ratios presented and explained below [1].

Contingency Tables

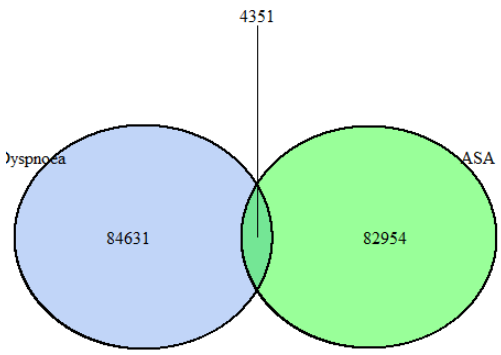
Each analysis of an association of drug exposure and adverse event („signal“) is based on a 2x2 contingency table which can be easily created in OpenVigil:

	Drug Exposure	No Drug Exposure	Sums
Adverse Event occurred	DE	dE	E
No Adverse Event occurred	De	de	e
Sums	D	d	N
1) We use more explicit terms instead of a,b,c and d. Capital letters denote occurrence of drug exposure (D) or adverse event (E). Lowercase letters denote no drug exposure (d) or no occurrence of the adverse event (e). The PRR is slightly vulnerable to transformation (De vs dE) of the table. DE ist also called “support” in some contexts [2, 3].			

The primary choice of analysis of such a contingency table are frequentist methods of disproportionality analysis (DPA) which are all based on observed and expected (OE) numbers.



	Drug(s) of interest	All other drugs	∑
Adverse event(s) of interest	4351	84631	88982
All other adverse events	82954	2807539	2890493
∑	87305	2892170	2979475




```
draw.pairwise.venn(area1 = 88982, area2 = 87305, cross.area=4351, fill = c("cornflowerblue", "green"), alpha=0.4, euler.d=T, scaled=T, category=c("Dyspnoea","ASA"))
```

Exampel #1: drug “acetylsalicylic acid” and “dyspnoea”
Data extracted with OpenVigil 2.0-experimental. Database contains all reports from 2003-10-29 to 2012-06-30. Venn diagram created with R 3.1.1 and library VennDiagram.

Number of cases (“DE”)

Cut off: Any meaningful statistics starts with DE>3, better more...

χ^2 (chi squared) with Yates' correction

Predication/Interpretation: Testing for independence in a contingency table as precautionary measure before applying the OE ratios below. The higher the χ^2 value, the more the observed numbers deviate from expected numbers.

Calculation (for χ^2 with Yates' correction for continuity) according to [4]: $\chi^2_{Yates} = N * (| DE*de – dE*De | - N/2)^2 / (D * d * E * e)$

Cut off: χ^2 values greater than 3.841 indicate statistical significance with $p \leq 0.05$, i.e., more than 95% chance that the observed numbers are really different from the expected numbers. Routinely, a χ^2 value of 4 (p=0.045) is used as cut-off [5].

Comparison to other tests: Fisher's exact test should be used for small values, a condition which is usually not given for pharmacovigilance data. Yates' correction is necessary for non χ^2 distributions and for 2x2 tables with only one degree of freedom [4].

Relativ Reporting Ratio (RRR)

Definition:

1. risk of an event = probability that the event occurs
2. relative reporting ratio (RRR) = probability for event in a group / probability for event in the whole population

Predication/Interpretation: Ratio of observed frequency (risk) to expected frequency in the complete population. It is a measure of association.

Statistical notation: Pr (ae | drug) / Pr (ae) = Pr (ae, drug) / Pr (ae) * Pr (drug)

	+ Drug	- drug	Sums
+ Adverse Event	DE	dE	E
- adverse event	De	de	e
Sums	D	d	N

Calculation [6-8]: RRR = DE * (DE + De + dE + de) / ((DE + De) * (DE + dE)) = DE * N / (D * E)

Cut-off: A cut-off value of 2 could be used [3].

Comparison to other uses: RRR is sometimes just called Reporting Ratio (RR, not to be confused with Relative Risk, see below).

Confidence intervals for RRR

Calculation of standard deviation: s = sqr(De/(DE*D) + e/(E*N))

The sampling distribution of RRR is positively skewed but approximately a normal distribution after z-transformation (log, ln). Converting RRR to natural log (ln), applying 1,96 times the standard deviation and converting this back to original scale, yields the confidence interval, i.e., the interval which contains the true value with $p \leq 0.05$.

CI = e ^ (ln RRR ± 1,96s)

This is method can be used instead of combining RRR with χ^2_{Yates}.

Proportional Reporting Ratio (PRR)

Definition:

1. risk of an event = probability that the event occurs
2. relative risk ("RR", ≈ PRR) = probability for event in group 1 / probability for event in group 2

Predication/Interpretation: Ratio of observed frequency in exposed population to non-exposed population. It is a measure of association.

Statistical notation: Pr (ae | drug) / Pr (ae | -drug)

Calculation [5-9]: PRR = (DE / D) / (dE / d)

	+ Drug	- drug	Sums
+ Adverse Event	DE	dE	E
- adverse event	De	de	e
Sums	D	d	N

Cut-off: Routinely, a cut off value of 2 is used to identify signals [5].

Comparison to other uses: The PRR is the pharmacovigilancy equivalent of the Relative Risk (RR, sometimes also called prevalence ratio) which is used for cohort studies [10].

Confidence intervals for PRR

Calculation of standard deviation [9]: s = sqr(De/(DE*D) + de/(dE*d))

Applying the standard deviation to the non-normal distributed PRR is the same like for RRR (see above):

CI = e ^ (ln PRR ± 1,96s)

This is method can be used instead of combining PRR with χ^2_{Yates}.

Reporting Odds Ratio (ROR)

Definition:

1. odds of an event = probability that the event occurs / probability that the event does not occur
2. odds ratio = odds for event in group 1 / odds for event in group 2

Predication/Interpretation: It is a measure of association.

Statistical notation: (Pr (ae | drug) / Pr (-ae | drug)) / (Pr (ae | -drug) / Pr (-ae | drug))

Calculation [6-8, 11]: ROR = (DE / De) / (dE / de) = DE*de / De*dE

	+ Drug	- drug	Sums
+ Adverse Event	DE	dE	E
- adverse event	De	de	e
Sums	D	d	N

Cut-off: As with the PRR and RRR, a ROR value of 2 can be used as threshold. However, usually the lower and upper bounds of the confidence interval of ROR are instead used (see below). The confidence interval must not cross the value 1 for statistical significance.

Comparison to other uses: The ROR is the pharmacovigilance equivalent of the Odds Ratio (OR) which is used for case-control-studies.

Comparison to PRR: The ROR will always be similar and a bit greater than PRR [10]. This was also shown for OR (≈ROR) and RR (≈PRR).

Confidence Intervals for ROR

Calculation of standard deviation [11]: s = sqr(1/DE + 1/De + 1/dE + 1/de)

Applying the standard deviation to the non-normal distributed ROR is the same like for RRR (see above):

CI = e ^ (ln ROR ± 1,96s)

This is method can be used instead of combining ROR with χ^2_{Yates}.

Non-frequentist, Bayesian methods

The Information Component (IC) is calculated as the $\log_2$ of the observed-expected ratios presented above:

Calculation [8]: IC = $\log_2$(RRR)

It is the basis for the Information Component (IC) for Multiple Gamma Poisson Shrinker (MGPS) and the Bayesian Confidence Propagation Neural Network (BCPNN) [6, 8].

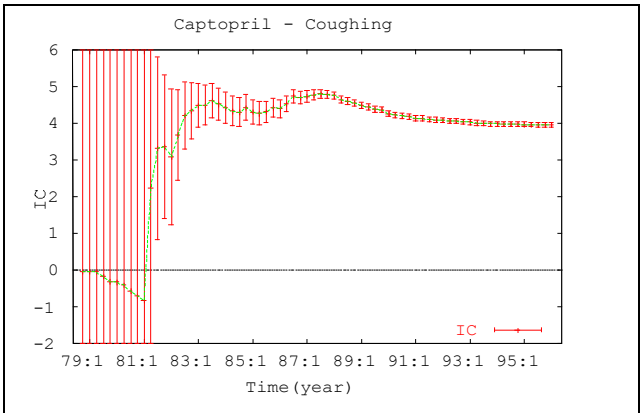


Fig. 1: Shrinkage of CI of IC over time
With increasing data, the confidence interval gets smaller. Once the value 0 is not included in the CI, a signal is flagged.
(Sten Olsson, the Uppsala Monitoring Centre, presentation, 2004)

Multi-item associations

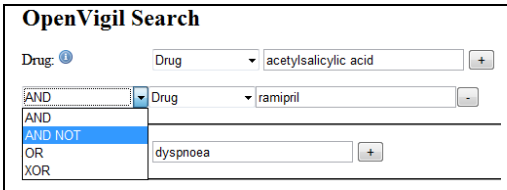
Instead of focusing on the association of only one drug to only one adverse event, associations of multiple items can be analysed, e.g.:

- two drugs and one adverse event: detection of a **drug-drug interaction (DDI)**, like superadditive effects on cardiac toxicity
- one drug and two adverse events: detection of a **syndrome with multiple symptoms**, like propofol infusion syndrome

The contingency table has to be expanded, e.g.,

	Drug Exposure to both drugs	Exposure to Drug 1 but not Drug 2	Exposure to Drug 2 but not Drug 1	No Exposure to either drug	Sums
Adverse Event occurred	D_1D_2E	D_1d_2E	d_1D_2E	d_1d_2E	E
No Adverse Event occurred	D_1D_2e	D_1d_2e	d_1D_2e	d_1d_2e	e
Sums	D_1D_2	D_1	D_2	d_1d_2	N

Example 1



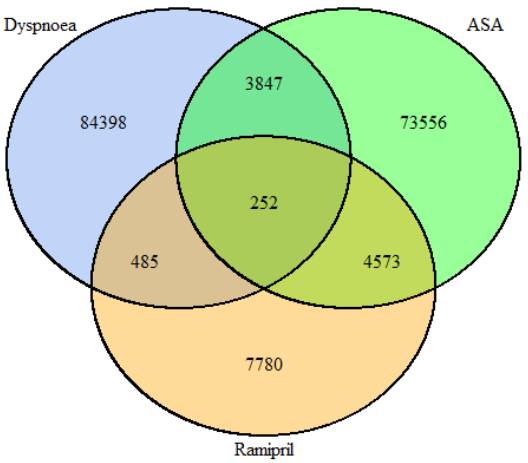
3 queries:

"acetylsalicylic acid" AND "rampril"

"acetylsalicylic acid" AND NOT "ramipril"

"ramipril"AND NOT "acetylsalicylic acid"

Calculation of d_1d_2E, d_1d_2e and d_1d_2.




```
draw.triple.venn(fill=c("cornflowerblue","green","orange"), alpha=0.4, euler.d=T, scaled=T, category=c("Dyspnoea","ASA","Ramipril"), area1=88982, area2=82228, area3=13090, n123=252,n12=4099,n13=737,n23=4825)
```

Example #1 (continued)
OpenVigil allows for combination of several drugs or events in the interface and via direct SQL statements.

Tab. for example #1 (continued) Drug #1: acetylsalicylic acid (ASA), Drug #2: ramipril, Event: Dyspnoea					
	Drug Exposure to both drugs	Exposure to Drug 1 but not Drug 2	Exposure to Drug 2 but not Drug 1	No Exposure to either drug	Sums
Adverse Event occurred	252	4099	737	83894	88982
No Adverse Event occurred	4825	78129	12353	2795186	2890493
Sums	5077	82228	13090	2879080	2979475

Multi-item-association analysis methods

Several approaches to combine multiple items for association analysis exist [2, 3, 12, 13].

Based on an early approach by IBM in 1993, originally used to mine associations found in shopping baskets and thus being able to propose other products to the buyer, Harpaz has used this concept to mine for associations in pharmacovigilance data. This approach aims to reduce the computing power needed. Therefore, only possible associations showing a so-called support (= DE, s. above) are included in the dataset for analysis. DE > 50, RRR > 2 [3]

The non-frequentist, Bayesian methods can be enhanced to include more than one item, the so-called **multi-item** gamma poisson shrinker (MGPS) [13].

Another approach is to compare the RRR for two drugs ($f_{11} = D_1D_2E / d_1d_2E$) with the expected RRR ($E[f_{11}]$). The latter can be estimated from other ratios (named f_{10}, f_{01} and f_{00}), called g_{11} [12]. A measure for interaction is $\Omega = \log_2 (f_{11}/g_{11})$ which can be expanded to $\Omega(\alpha) = \log_2 ((f_{11}+ \alpha) / (g_{11} + \alpha))$. α is a tuning parameter for shrinkage strength and is equivalent to α additional expected reports. Finally, a confidence interval can be constructed.

References

1. Böhm, R., *OpenVigil cave-at document*. 2011.
2. Agrawal, R., T. Imieliński, and A. Swami. *Mining association rules between sets of items in large databases*. in *ACM SIGMOD Record*. 1993: ACM.
3. Harpaz, R., H.S. Chase, and C. Friedman, *Mining multi-item drug adverse effect associations in spontaneous reporting systems.* BMC Bioinformatics, 2010. **11 Suppl 9**: p. S7.
4. Wikipedia, *Yates's correction for continuity*. 2014.
5. Evans, S.J., P.C. Waller, and S. Davis, *Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports.* Pharmacoepidemiol Drug Saf, 2001. **10**(6): p. 483-6.
6. Agah, A., *Medical Applications of Artificial Intelligence*. 2013: Taylor & Francis.
7. Holle, L.V. and V. Bauchau, *The upper bound to the Relative Reporting Ratio—a measure of the impact of the violation of hidden assumptions underlying some disproportionality methods used in signal detection.* Pharmacoepidemiology and drug safety, 2014.
8. Madigan, D., et al., *Bayesian methods in pharmacovigilance.* Bayesian Statistics, 2011. **9**: p. 421-438.
9. Wikipedia, *Relative risk*. 2014.
10. Waller, P., et al., *The reporting odds ratio versus the proportional reporting ratio: 'deuce'.* Pharmacoepidemiology and drug safety, 2004. **13**(8): p. 525-526.
11. Wikipedia, *Odds ratio*. 2014.
12. Norén, G.N., et al., *A statistical methodology for drug–drug interaction surveillance.* Statistics in medicine, 2008. **27**(16): p. 3057-3070.
13. Szarfman, A., S.G. Machado, and R.T. O'Neill, *Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database.* Drug Saf, 2002. **25**(6): p. 381-92.