jmp®
Statistical Discovery.™ From SAS.

Disproportionality Analysis and Its Application to Spontaneously Reported Adverse Events in Pharmacovigilance

Richard C. Zink, SAS, Cary, NC

Table of Contents

Introduction 1
Spontaneously Reported Adverse Events 2
The Disproportionality Analysis Dialog 3
Signal Summary and Tree Maps 5
Bubble Plots and Time Trends 6
Hierarchical and K-Means Clustering of Drugs or Events 7
Add and Maintain Comments for Drug-Event Pairs 8
Summary of Event and Record Characteristics 9
Conclusions 9
References 9
Technical Appendix 10

Introduction

While randomized clinical trials are the gold standard for evaluating the efficacy of a new intervention, the available sample size is often insufficient to fully understand its safety profile. The risk a new therapy may pose may not be well understood until it has been on the market for many years, taken by individuals who differ from those studied under the inclusion criteria of the clinical development program.

Spontaneously reported adverse events are collected by regulatory agencies, pharmaceutical companies and device manufacturers to monitor the safety of a product once it reaches the market. The FDA, for example, maintains databases for pharmaceuticals (AERS: Adverse Event Reporting System), vaccines (VAERS: Vaccine Adverse Event Reporting System) and medical devices (MAUDE: Manufacturer and User Facility Device Experience). These data are generally obtained from physicians, patients or the medical literature.

Spontaneously reported adverse events present a unique challenge in that there is no measure of total exposure. In other words, there is no clear denominator to define an adverse event incidence for a particular drug. In order to identify potential safety-signals, the rate at which a particular event of interest co-occurs with a given drug is compared to the rate this event occurs without the drug in the event database. This is referred to as disproportionality analysis.

This paper will introduce the **Disproportionality Analysis** analytical process (AP) available in JMP® Clinical.

Spontaneously Reported Adverse Events

Data are assumed to be of the format displayed in **Figure 1**. A case, which can refer to an individual patient or report, can have one or more adverse events associated with it. Each event, in turn, can have one or more drugs associated with it. Events that occur on different cases are considered distinct for the purposes of determining a total count of adverse events for the analysis. Within a case, JMP Clinical determines event uniqueness based on the variables supplied for analysis such as the event name and classification, which could refer to MedDRA preferred term and system organ class; the onset date of the event; or any variables used for stratification, such as event severity.

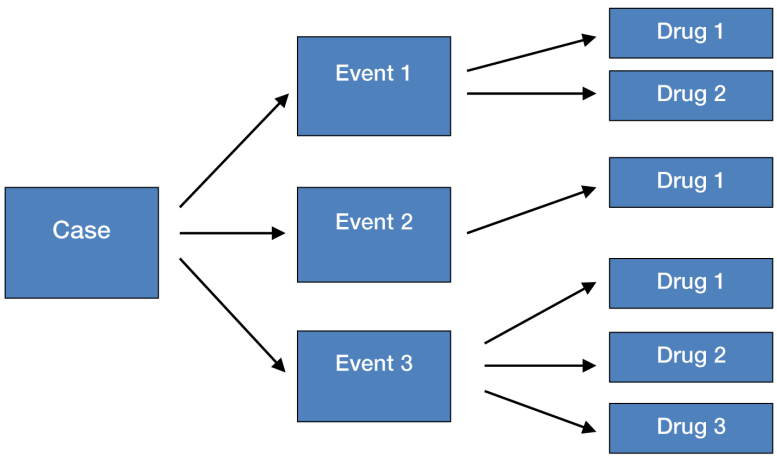


Figure 1. Data Structure

Once the total number of events is determined, contingency tables similar to **Table 1** are used to define measures of disproportionality for the *i*th drug and the *j*th adverse event. JMP Clinical calculates four measures of disproportionality – the reporting odds ratio (Meyboom et al., 1997), the proportional reporting ratio (Evans, Waller & Davis, 2001), the multi-item gamma poisson shrinker (DuMouchel, 1999) and the Bayesian confidence propagation neural network (Bate et al., 1998; Gould, 2003). Technical details are left to the Appendix.

		Event of Interest		
Drug of interest		**Mentioned**	**Not Mentioned**	
Mentioned		n_{hij}	$n_{hi.} - n_{hij}$	$n_{hi.}$
Not mentioned		$n_{h.j} - n_{hij}$	$N_h - n_{hi.} - n_{h.j} + n_{hij}$	$N_h - n_{hi.}$
		$n_{h.j}$	$N_h - n_{h.j}$	N_h

Table 1. Contingency Table for Stratum h, Drug i and Event j

The Disproportionality Analysis Dialog

The **Disproportionality Analysis** AP (**Figure 2**) has minimal data requirements to perform an analysis – an event field and a drug field. If you supply a data set of drug and event pairs, JMP Clinical assumes each record belongs to a unique case and conducts the analysis accordingly. JMP Clinical ships with a simulated data set of 10,000 cases. Generally, each case represents a single event for which a single drug is listed. However, there are two single-event cases with multiple drugs listed – one with four drugs and one with three drugs for a total of 10,005 records.

Though data requirements are minimal, additional specified variables provide a more informative analysis. For example, specifying an event classification (e.g., MedDRA system organ class) or drug classification (e.g., WHODrug ATC classification) can allow for straightforward subsetting of appropriate records of interest to perform cluster analyses. Since an event or drug may be associated with multiple classes, JMP Clinical associates the most commonly occurring classes with the computed disproportionality statistics. Specifying a location provides you with a geographical summary of where events are reported, which may provide insight into geographical trends of event reporting. Finally, providing an onset date generates analyses of disproportionality statistics over time to illustrate how robust findings are to accumulating data.

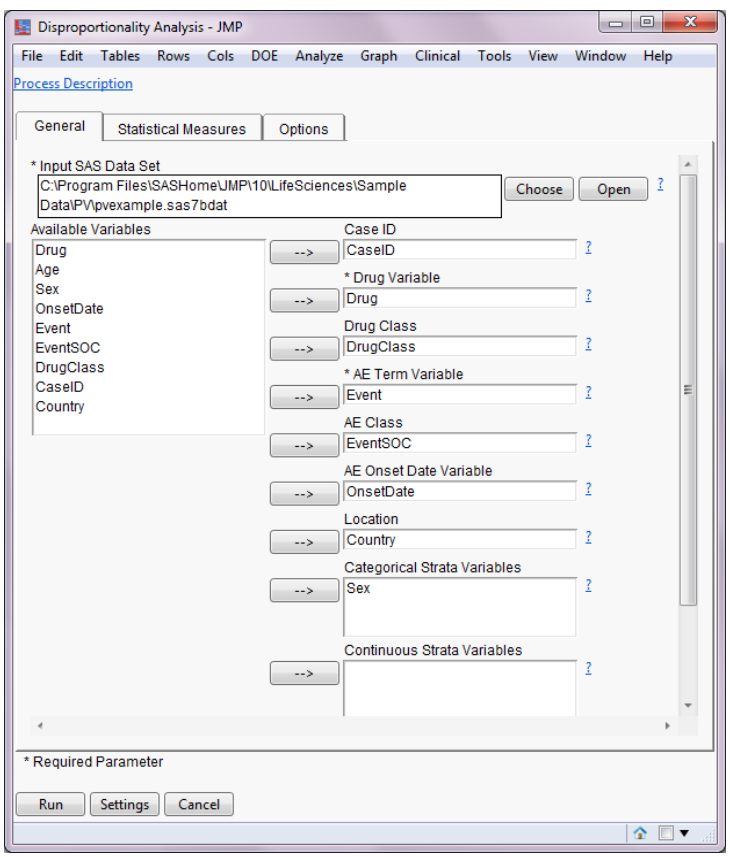


Figure 2. The Disproportionality Analysis Dialog

Stratification is an important consideration for disproportionality analysis. How drugs are prescribed, disease severity and how individuals may respond to treatment can all be influenced by demographic and other background characteristics. Therefore, disproportionality statistics should be calculated among homogeneous groups to avoid inappropriate conclusions (Woo et al., 2008). One or more variables can be specified to perform a stratified analysis. JMP Clinical will calculate strata based on the cross-classification of any supplied variables, categorizing any continuous variables into a user-defined number of subgroups.

As described above, JMP Clinical calculates four measures of disproportionality – the reporting odds ratio (ROR), the proportional reporting ratio (PRR), the multi-item gamma poisson shrinker (MGPS) and the Bayesian confidence propagation neural network (BCPNN). Further, the specific signal criteria can be tailored to the particular analysis requirements (**Figure 3**). In most cases, the criteria for signal generation (whether a particular drug-event pair occurs disproportionally) is a combination of event frequency and the magnitude of the lower limit of the confidence or credible interval. However, a common alternate specification for PRR based on the magnitudes of the chi-square statistic, the PRR estimate and event frequency is also available.

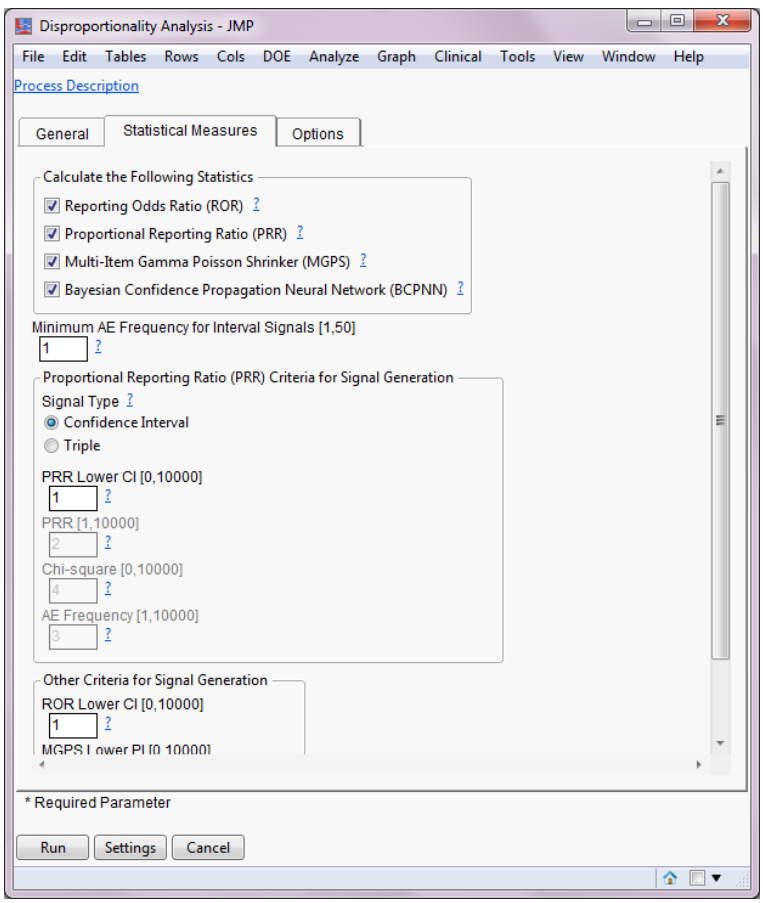
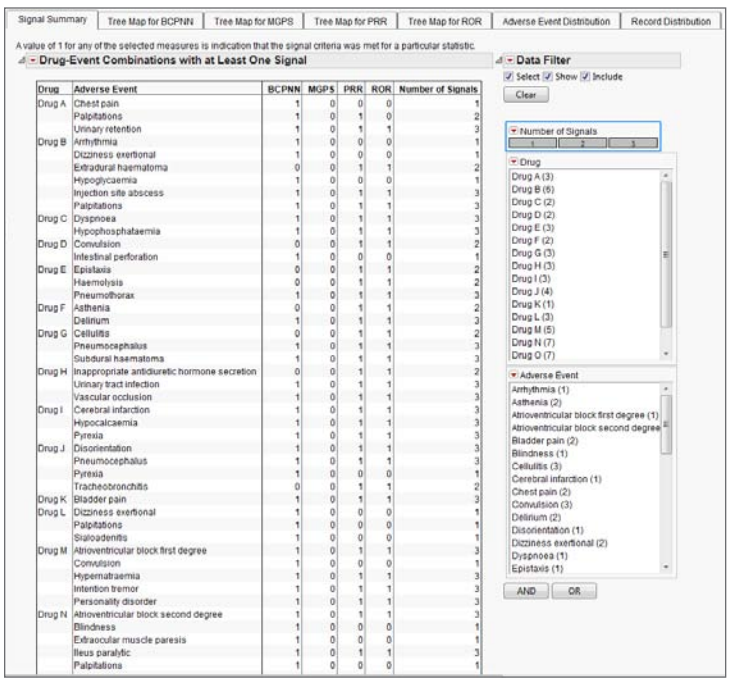


Figure 3. Disproportionality Measures and Signal Criteria

Signal Summary and Tree Maps

The results dashboard is shown below. The **Signal Summary** tab (**Figure 4**) summarizes any drug-event pairs that had a signal from at least one of the selected disproportionality statistics, and highlights any differences among the selected statistics for generating a signal. For each requested statistic, a **Tree Map** uses color and size to visually highlight the magnitude of the lower limits of the confidence or credible intervals. From **Figure 5**, it is straightforward to see that Drug P had a very large estimate of disproportionality for Vascular Occlusion. The available data filter can be used to subset **Tree Maps** based on drugs, events, classifications or whether drug-event pairs meet signal criteria or not.



Drug	Adverse Event	BCPNN	MGPS	PRR	ROR	Number of Signals
Drug A	Chest pain	1	0	0	0	1
	Palpitations	1	0	1	0	2
	Urinary retention	1	0	1	1	3
Drug B	Arrhythmia	1	0	0	0	1
	Dizziness exertional	1	0	0	0	1
	Extradural haematoma	0	0	1	1	2
	Hypoglycaemia	1	0	0	0	1
	Injection site abscess	1	0	1	1	3
	Palpitations	1	0	1	1	3
Drug C	Dyspnoea	1	0	1	1	3
	Hypophosphataemia	1	0	1	1	3
Drug D	Convulsion	0	0	1	1	2
	Intestinal perforation	1	0	0	0	1
Drug E	Epistaxis	0	0	1	1	2
	Haemolysis	0	0	1	1	2
	Pneumothorax	1	0	1	1	3
Drug F	Asthenia	0	0	1	1	2
	Delirium	1	0	1	1	3
Drug G	Cellulitis	0	0	1	1	2
	Pneumocephalus	1	0	1	1	3
	Subdural haematoma	1	0	1	1	3
Drug H	Inappropriate antidiuretic hormone secretion	0	0	1	1	2
	Urinary tract infection	1	0	1	1	3
	Vascular occlusion	1	0	1	1	3
Drug I	Cerebral infarction	1	0	1	1	3
	Hypocalcaemia	1	0	1	1	3
	Pyrexia	1	0	1	1	3
Drug J	Disorientation	1	0	1	1	3
	Pneumocephalus	1	0	1	1	3
	Pyrexia	1	0	0	0	1
	Tracheobronchitis	0	0	1	1	2
Drug K	Bladder pain	1	0	1	1	3
Drug L	Dizziness exertional	1	0	0	0	1
	Palpitations	1	0	0	0	1
	Sialoadenitis	1	0	0	0	1
Drug M	Atrioventricular block first degree	1	0	1	1	3
	Convulsion	1	0	0	0	1
	Hypernatraemia	1	0	1	1	3
	Intention tremor	1	0	1	1	3
	Personality disorder	1	0	1	1	3
Drug N	Atrioventricular block second degree	1	0	1	1	3
	Blindness	1	0	0	0	1
	Extraocular muscle paresis	1	0	0	0	1
	Ileus paralytic	1	0	1	1	3
	Palpitations	1	0	0	0	1

Figure 4. Summary of Signals Generated by Selected Methods

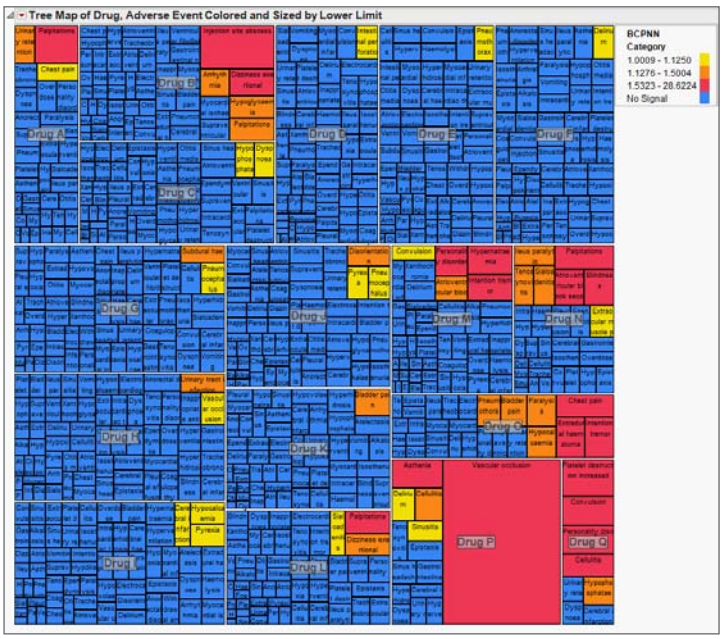


Figure 5. Tree Map of Bayesian Confidence Propagation Neural Network

Bubble Plots and Time Trends

From any **Tree Map**, you can highlight any number of cells to perform additional analyses. For example, bubble plots from the **Overall Bubble** (**Figure 6**, left) drill-down button display the magnitudes of disproportionality across all drugs reported for selected events. Bubble size indicates the frequency of the event providing insight into whether a strong signal occurred due to a very rare event. For a stratified analysis, within-strata estimates of disproportionality are available from **Strata Bubble** (**Figure 6**, right) to highlight differences among strata. Finally, should you supply a variable for the event onset date, **Overall Bubble Across Time** displays animation of the change in event frequency and disproportionality over time, changing color as the signal criteria is met or lost.

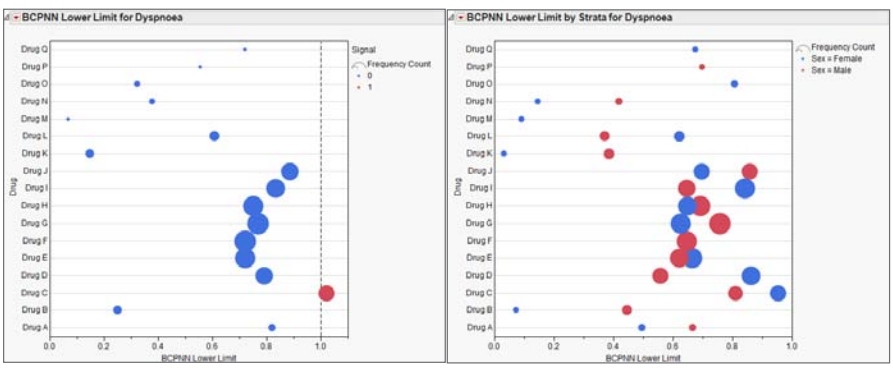


Figure 6. Overall and Strata Bubble Plots for Dyspnoea

More traditional across-time displays are available using the **Time Trend** drill-down button. **Figure 7** shows the lower limit of the 95 percent credible limit for dyspnoea. It is straightforward to see that Drug A was initially disproportional, although this signal disappeared over time. However, in 2004 Drug C began showing signs of disproportionality, which has continued to the present (though the lower limit is just beyond the specified signal threshold). Based on dialog options, you are able to specify whether data is summarized annually, biannually or quarterly and the number of most recent time points to include for figures. Limiting the analysis to a reasonable number of recent time points can have implications on signals (since rare events may have occurred in a limited number early on) and the speed at which an analysis is generated. By default, JMP Clinical limits analyses across time to the most recent 12 time points. An additional "overall" analysis is performed that includes all events, including those with missing onset dates. This overall analysis is presented as the primary results in all summary output that is not presented across time.

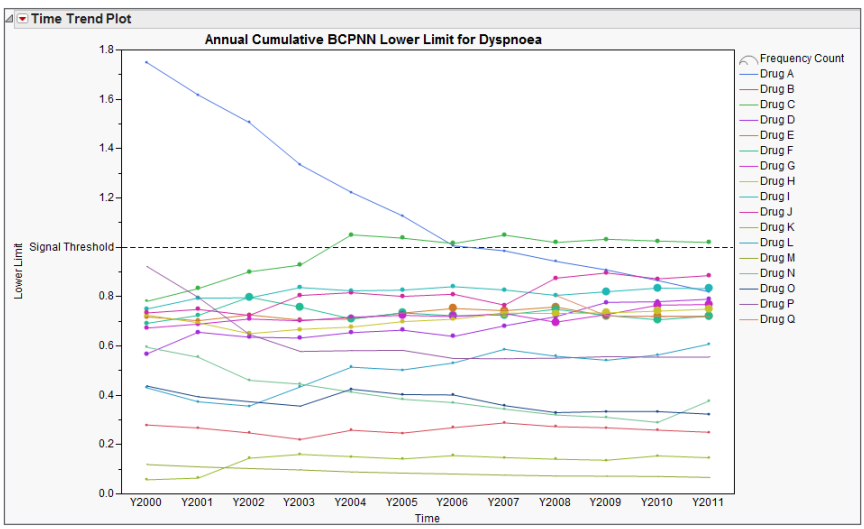


Figure 7. Time Trend Plot of Dyspnoea for Most Recent 12 Years

Hierarchical and K-Means Clustering of Drugs or Events

Some drugs may have similar mechanisms of action or risks associated with them. To identify drugs with similar adverse event patterns, selected drugs from a **Tree Map** can be clustered using the **Hierarchical for Drug** or **K-Means for Drug** drill-down buttons. Sample hierarchical clustering output is displayed in **Figure 8**. Similarly, adverse events can be clustered across drugs using the **Hierarchical for Event** or **K-Means for Event** buttons.

Add and Maintain Comments for Drug-Event Pairs

In the course of an analysis, certain facts may surface for a particular drug-event pair, or certain findings may warrant comment. For example, based on the drug label, a treatment may be known to cause a particular adverse event. To remind yourself of this fact during future analyses, JMP Clinical allows you to highlight **Tree Map** cells and save comments for selected drug-event pairs. These comments will be available for viewing in subsequent analyses, such as when the database is refreshed with additional records.

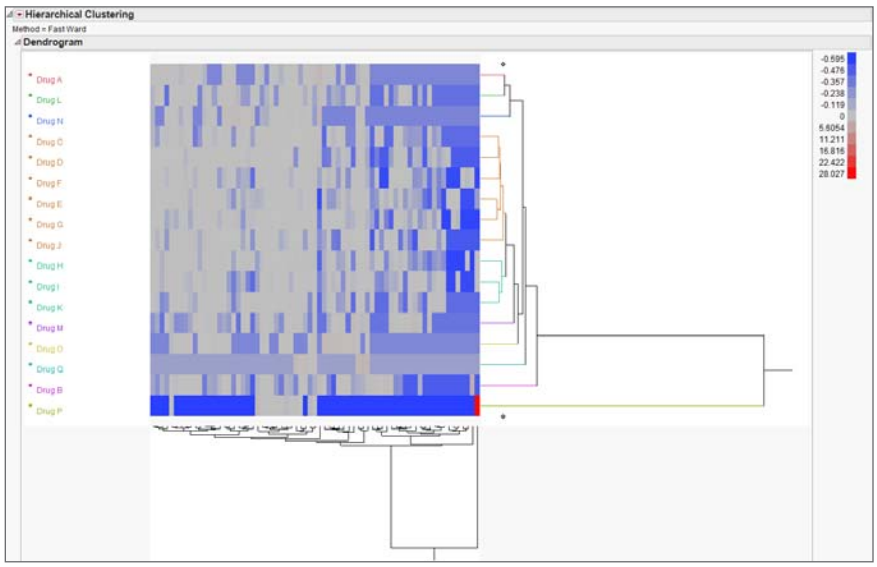


Figure 8. Hierarchical Clustering of Drugs Using BCPNN Estimate of Disproportionality

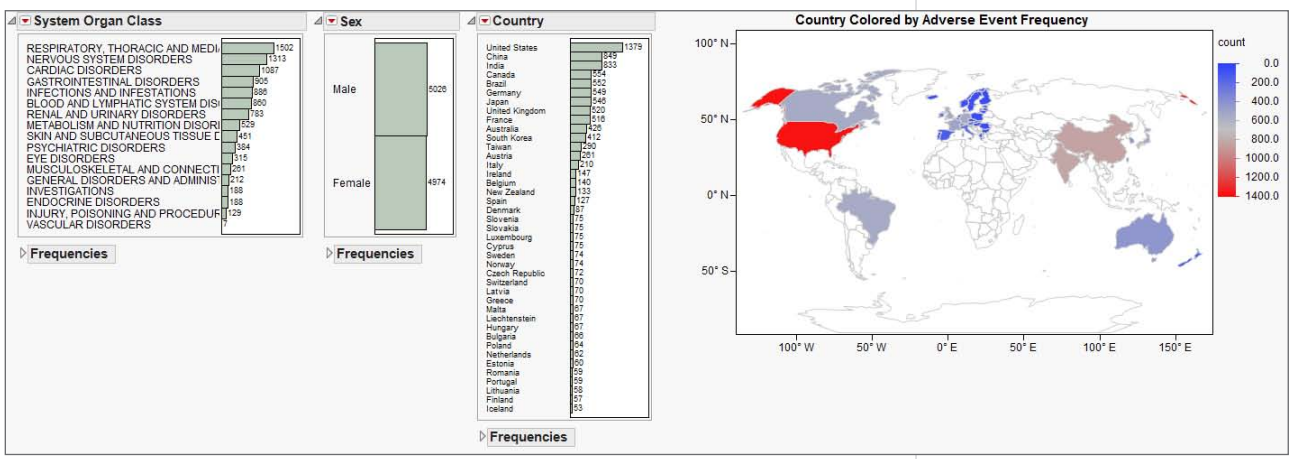


Figure 9. Distributions of Adverse Event Characteristics and Geographical Display

Summary of Event and Record Characteristics

JMP Clinical provides distributional summaries of all user-specified analysis variables at the event- and record-levels. Further, if event location such as country is supplied, full-color maps are generated to identify any geographical trends.

Conclusions

Using a compelling integration of dynamic graphics and statistics, JMP Clinical offers the most popular Frequentist and Bayesian pharmacovigilance measures that enable you to explore spontaneously reported adverse events and identify potential safety signals.

References

Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A & De Freitas RM. (1998). A Bayesian neural network method for adverse drug reaction signal generation. *European Journal of Clinical Pharmacology* 54: 315-21.

DuMouchel W. (1999). Bayesian data mining in large frequency tables with an application to the FDA spontaneous reporting system. *The American Statistician* 53: 177-90.

Evans SJW, Waller PC & Davis S. (2001). Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and Drug Safety* 10: 483-86.

Gould, AL. (2003). Practical pharmacovigilance analysis strategies. *Pharmacoepidemiology and Drug Safety* 12: 559-74.

Meyboom RHB, Egberts ACG, Edwards IR, Hekster YA, de Koning FHP & Gribnau FWJ. (1997). Principles of signal detection in pharmacovigilance. *Drug Safety* 16: 355-65.

Woo EJ, Ball R, Burwen DR & Braun MM. (2008). Effects of stratification on data mining in the US vaccine adverse event reporting system (VAERS). *Drug Safety* 31: 667-74.

Technical Appendix

For the hth strata, the ith drug and the jth event, the following summary table can be written:

	Event of Interest		
Drug of interest	**Mentioned**	**Not Mentioned**	
Mentioned	n_{hij}	$n_{hi.} - n_{hij}$	$n_{hi.}$
Not mentioned	$n_{h.j} - n_{hij}$	$N_h - n_{hi.} - n_{h.j} + n_{hij}$	$N_h - n_{hi.}$
	$n_{h.j}$	$N_h - n_{h.j}$	N_h

with the total number of cases $N = \sum_h N_h$ by default. Alternatively, **JMP Clinical** has options for performing the analysis using counts based on unique events.

Proportional Risk Ratio (PRR)

The Mantel-Haenszel Estimate of the PRR is

$$\text{PRR}_{MH} = \frac{\sum_h n_{hij}(N_h - n_{hi.})/N_h}{\sum_h n_{hi.}(n_{h.j} - n_{hij})/N_h}.$$

The 95% Confidence Interval for PRR_{MH} is

$$\text{PRR}_{MH} \times \exp(-1.96\hat{\sigma}), \text{PRR}_{MH} \times \exp(1.96\hat{\sigma})$$

where

$$\hat{\sigma}^2 = \frac{\sum_h (n_{hi.}(N_h - n_{hi.})n_{h.j} - n_{hij}(n_{h.j} - n_{hij})N_h)/N_h^2}{\sum_h n_{hij}(N_h - n_{hi.})/N_h \sum_h n_{hi.}(n_{h.j} - n_{hij})/N_h}.$$

Reporting Odds Ratio (ROR)

The Mantel-Haenszel estimate of the ROR is

$$\text{ROR}_{MH} = \frac{\sum_h n_{hij}(N_h - n_{hi.} - n_{h.j} + n_{hij})/N_h}{\sum_h (n_{hi.} - n_{hij})(n_{h.j} - n_{hij})/N_h}.$$

The 95% Confidence Interval for ROR_{MH} is

$$\text{ROR}_{MH} \times \exp(-1.96\hat{\sigma}), \text{PRR}_{MH} \times \exp(1.96\hat{\sigma})$$

where

$$\hat{\sigma}^2 = \frac{\sum_h \left(N_h - n_{hi.} - n_{h.j} + 2n_{hij}\right) n_{hij} \left(N_h - n_{hi.} - n_{h.j} + n_{hij}\right)/N_h^2}{2\left(\sum_h n_{hij}\left(N_h - n_{hi.} - n_{h.j} + n_{hij}\right)/N_h\right)^2}$$
$$+ \frac{\sum_h \left(N_h - n_{hi.} - n_{h.j} + 2n_{hij}\right)\left(n_{hi.} - n_{hij}\right)\left(n_{h.j} - n_{hij}\right)/N_h^2}{2\left(\sum_h n_{hij}\left(N_h - n_{hi.} - n_{h.j} + n_{hij}\right)/N_h\right)\left(\sum_h \left(n_{hi.} - n_{hij}\right)\left(n_{h.j} - n_{hij}\right)/N_h\right)}$$
$$+ \frac{\sum_h \left(n_{hi.} + n_{h.j} - 2n_{hij}\right) n_{hij} \left(N_h - n_{hi.} - n_{h.j} + n_{hij}\right)/N_h^2}{2\left(\sum_h n_{hij}\left(N_h - n_{hi.} - n_{h.j} + n_{hij}\right)/N_h\right)\left(\sum_h \left(n_{hi.} - n_{hij}\right)\left(n_{h.j} - n_{hij}\right)/N_h\right)}$$
$$+ \frac{\sum_h \left(n_{hi.} + n_{h.j} - 2n_{hij}\right)\left(n_{hi.} - n_{hij}\right)\left(n_{h.j} - n_{hij}\right)/N_h^2}{2\left(\sum_h \left(n_{hi.} - n_{hij}\right)\left(n_{h.j} - n_{hij}\right)/N_h\right)^2}.$$

Bayesian Confidence Propagation Neural Network (BCPNN)

Assume $n_{hi.} \sim \text{Bin}(N_h, p_{hi})$, $n_{h.j} \sim \text{Bin}\left(N_h, q_{hj}\right)$, $n_{hij} \sim \text{Bin}\left(N_h, r_{hij}\right)$ where p_{hi}, q_{hj} and r_{hij} have prior distributions Beta(1,1), Beta(1,1) and Beta(1,3), respectively. The estimate for the information component of the BCPNN is

$$IC_{ij} = \sum_h \omega_h \ln(2)^{-1}\left\{\Psi\left(1 + n_{hij}\right) - \Psi(4 + N_h)\right.$$
$$\left. - \left[\Psi(1 + n_{hi.}) - \Psi(2 + N_h) + \Psi\left(1 + n_{h.j}\right) - \Psi(2 + N_h)\right]\right\}$$

with variance

$$\hat{\sigma}^2 = \sum_h \omega_h^2 \ln(2)^{-2}\left\{\Psi'\left(1 + n_{hij}\right) - \Psi'(4 + N_h)\right.$$
$$\left. + \left[\Psi'(1 + n_{hi.}) - \Psi'(2 + N_h) + \Psi'\left(1 + n_{h.j}\right) - \Psi'(2 + N_h)\right]\right\}$$

where Ψ and Ψ' are the digamma and trigamma functions, respectively and $\omega_h = \frac{N_h}{N}$. An approximate 95% credible interval is IC±2$\hat{\sigma}$. **JMP Clinical** presents BCPNN=2^{IC} with approximate 95% credible interval $2^{IC \pm 2\hat{\sigma}}$.

Multi-Item Gamma Poisson Shrinker (MGPS)

Let $E_{.ij}$ be the expected number of events for the *i*th drug and *j*th event under independence, $E_{.ij} = \sum_h \frac{n_{hi.} n_{h.j}}{N_h}$ and let $N_{.ij} = \sum_h N_{hij}$ be the observed count of events for the *i*th drug and *j*th event. Define $L\left(\alpha_1, \beta_1, \alpha_2, \beta_2, \rho\right) = \prod_{i,j} \rho f\left(N_{.ij}; \alpha_1, \beta_1, E_{.ij}\right) + (1 - \rho) f\left(N_{.ij}; \alpha_2, \beta_2, E_{.ij}\right)$,

where $f(n;\alpha,\beta,E) = \left(1+\frac{\beta}{E}\right)^{-n}\left(1+\frac{E}{\beta}\right)^{-\alpha} \times \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)n!}$

and obtain maximum likelihood estimators for parameters to calculate

$$\begin{aligned}\mathrm{E}[\log(\lambda)|N_{\cdot ij}] = Q_{N_{ij}}&\left[\Psi(\alpha_1+N_{\cdot ij}) - \ln(\beta_1+E_{\cdot ij})\right] \\ &+ (1-Q_{N_{\cdot ij}})\left[\Psi(\alpha_2+N_{\cdot ij}) - \ln(\beta_2+E_{\cdot ij})\right]\end{aligned}$$

where $Q_{N_{\cdot ij}} = \frac{\rho f(N_{\cdot ij};\alpha_1,\beta_1,E_{\cdot ij})}{\rho f(N_{\cdot ij};\alpha_1,\beta_1,E_{\cdot ij})+(1-\rho)f(N_{\cdot ij};\alpha_2,\beta_2,E_{\cdot ij})}$ and Ψ is the digamma function. Finally, define $\mathrm{MGPS}_{ij} = e^{\mathrm{E}[\log(\lambda)|N_{\cdot ij}]}$. A 95% credible interval is $(\lambda_{ij,0.025}, \lambda_{ij,0.975})$ where the limits are solutions to

$$0.025 = \int_0^{\lambda_{ij,0.025}} \pi\left(\lambda;\alpha_1+N_{\cdot ij},\beta_1+E_{\cdot ij},\alpha_2+N_{\cdot ij},\beta_2+E_{\cdot ij},Q_{N_{\cdot ij}}\right)\partial\lambda$$

$$0.975 = \int_0^{\lambda_{ij,0.975}} \pi\left(\lambda;\alpha_1+N_{\cdot ij},\beta_1+E_{\cdot ij},\alpha_2+N_{\cdot ij},\beta_2+E_{\cdot ij},Q_{N_{\cdot ij}}\right)\partial\lambda$$

where $\pi(\lambda;\alpha_1,\beta_1,\alpha_2,\beta_2,\rho) = \rho g(\lambda;\alpha_1,\beta_1) + (1-\rho)g(\lambda;\alpha_2,\beta_2)$ and $g(\lambda;\alpha,\beta) = \frac{\beta^\alpha \lambda^{\alpha-1}e^{-\beta\lambda}}{\Gamma(\alpha)}$.