ORACLE
HEALTH SCIENCES

An Oracle White Paper
November 2012

Regression-Adjusted GPS Algorithm (RGPS)

William DuMouchel and Rave Harpaz

ORACLE

Disclaimer

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Executive Overview

This document describes a new algorithm that is a hybrid of Extended Logistic Regression (ELR) and the Multi-item Gamma Poisson Shrinker (MGPS). It is similar to MGPS in that relative reporting rates (RR) are input into a Bayesian gamma-Poisson shrinking algorithm to get more reliable estimated rates and confidence intervals. The main difference is that instead of using stratification and a Mantel-Haentzel approach to adjust for patient covariates and compute an adjusted expected value (E), which is used as a denominator for RR, the values of E are computed using the results of an ELR analysis. This new Regression-adjusted GPS algorithm will be abbreviated as RGPS. This paper describes the algorithm in some (but not complete) detail, including its method of screening for drug-drug interactions. It also uses compares the results of RGPS to those of MGPS and ELR on a recent AERS database, and uses the OMOP gold standard set of nearly 400 drug-event combinations (DECs) to show that RGPS has greater discriminatory power than the earlier methods.

Introduction

Databases of reports of adverse drug reactions are primary sources of information about possible harms from prescription drugs. One of the largest of such databases is the Adverse Event Reporting System (AERS) maintained by the US FDA[1]. AERS now has about 5 million reports involving over 5 thousand drugs and over 10 thousand coded Preferred Terms (PTs) within the Medical Dictionary for Regulatory Activities (MedDRA)[2] structured vocabulary. Analysis of AERS for discovering potential associations of drugs with adverse reactions is challenging because of the severe limitations of the data reliability within spontaneous reports[3, 4]. The voluntary reports don't follow a research protocol, and adverse report rates seem to vary from year to year, as well as by drug and by adverse event type. There is no certainty that a reported reaction was caused by the drug(s) in the report. Since there are no accompanying data on the numbers exposed to each drug, incidence rates cannot be determined. Instead, it is common to compute *disproportionalities*, which compare the number of reports of each adverse event (AE) within reports mentioning the drug of interest to a comparison or expected count based on a null hypothesis of no association of the drug and the event within the database[5-7].

Several disproportionality measures are commonly computed. The proportional reporting ratio (PRR)[8] and the reporting odds ratio (ROR)[9] are based on simple ratios of counts from a two-by-two table of presence/absence of a drug of interest and an event of interest. Bayesian measures of disproportionality also attempt to use a prior distribution to smooth or adjust observed-to-expected ratios for their expected statistical variability in order to obtain more reliable estimates. The multi-item gamma Poisson shrinker (MGPS)[10, 11] estimates an empirical Bayes prior distribution for the observed-to-expected ratio based on examining such ratios for all the drug-event combinations (DEC) in the

database. MGPS also uses a Mantel-Haentzel adjustment when computing expected values to correct for potential confounding caused by patient covariates such as age or sex, or by secular trends when reports many years apart have been pooled. MGPS has the advantages that it can be efficiently and automatically run to get disproportionality estimates for the potentially millions of DECs in a database, and that its adjustment for all covariate combinations and its Bayesian shrinkage algorithm enhance accuracy and interpretability compared to the simpler methods such as PRR or ROR. However, none of these methods adjust for the effects of polypharmacy—the fact that many patients are taking multiple drugs at the same time, and that drugs for the same indication will naturally often show up in the same patient reports, creating confounding issues and other potential statistical biases[5]. A masking bias occurs when the expected counts for a DEC of interest are inflated because other drugs in the database cause the event of interest, whereas, ideally, reports involving those other causal drugs should be excluded from the calculations of the expected or null hypothesis count. This masking bias can reduce the disproportionality estimate for the drug of interest, causing a missed signal. A confounding or "innocent bystander" bias occurs when a drug with no causal connection to the AE of interest is often co-prescribed with a drug that does sometimes cause the AE. In that case the co-prescribing causes both drugs to appear associated with the AE.

The main methodology for coping with the fact that spontaneous reports mention multiple drugs per report is logistic regression or its variants[7, 12-15]. A regression model attempts to compute the probability that a given AE will be mentioned in a report based on the set of drugs mentioned in the report as well as the before-mentioned report covariates such as age, sex or report year. Logistic regression assumes that a linear combination of predictors can combine with the logistic probability function to fit the probability of any particular AE being present. This can have the unexpected disadvantage of implicitly assuming that multiple drugs affect risk multiplicatively, whereas an additive effects model or some other model may fit the observed event frequencies much better. This assumption can be relaxed by extending the logistic probability function to a wider family of probability distributions. We denote one such extension, implemented within Empirica Signal, as extended logistic regression (ELR)[16]. Standard LR or ELR regression algorithms are challenged by the fact that with over 5,000 drugs in the database, there could be at least that many predictors in the regression model. If the response AE is rare, there may be fewer than 1,000 reports of the event, even if the database has millions of reports, which can cause standard logistic regression algorithms to fail. The usual approach would be to use medical knowledge or statistical expertise to greatly reduce the number of drugs used in the model, perhaps to a few or at most a few hundred drugs. But this then may require user interaction or statistical expertise which would have to be applied to every separate modeling example, one for each of the thousands of AEs in the database. Often the spontaneous reports analysts have medical knowledge but not statistical expertise, and they may not be comfortable with the whole idea of selecting predictors to use in a regression model, preferring a more automatic approach such as MGPS provides.

By combining ideas from ELR and MGPS, RGPS aims to achieve the best of both methods. Like ELR, RGPS is designed to account for confounding by concomitant drugs and for masking effects. Like ELR, RGPS allows for a better fit when risk accumulation is not multiplicative. Like MGPS, the shrinkage model protects from false alarms due to multiple comparisons. Like MGPS, RGPS is user friendly and does not require complicated model choice decisions as with ELR. In RGPS all computation is done automatically. Similar to MGPS, RGPS computes pairs of observed and expected counts, which can be used subsequently to compute more complex associations such as drug-drug interactions and more elaborate shrinkage models based on drug classes and event hierarchies. The RGPS method for computing drug-drug interaction signals is less computing intensive than the 3-D run required to get interaction scores in MGPS. In addition, unlike MGPS RGPS is able to compute shrinkage estimates for cases where N=0.

RGPS processes one response (AE) at a time (similar to ELR). The computation of each response is independent of other responses so that several responses can be processed in parallel. Similar to MGPS a gamma-Poisson model is used to shrink N towards E, but the shrinkage model is estimated separately for each response and is simpler in the sense that it uses a single gamma prior model rather than a mixture of two gammas (although a different gamma prior is estimated for each response).

In what follows it is assumed that the spontaneous report data is efficiently available in a database, and that a specific set of drugs, events and categorical report covariates have been specified to be in the analysis. The analysis repeats for each selected AE, producing values of N_j, E_j, $RR_j = N_j/E_j$, $EBRRR_j$, $RRR05_j$, $RRR95_j$ for every AE that is analyzed, where j varies across all the drugs, and where EBRRR, RRR05, RRR95 are the analogous RGPS versions of what MGPS calls EBGM, EB05, and EB95.

The RGPS Methodology

The RGPS methodology consists of three main steps. First, a set of predictors (drugs and grouped-strata covariates) are automatically selected to fit a Bayesian ELR model. For clarity, we distinguish between the current ELR analysis that is available in Empirica Signal v7, to be denoted ES-ELR, and the way in which ELR is implemented in RGPS, which we denote RGPS-ELR. The selection of which drugs to use in ES-ELR is semi-automated, in that there are default rules based solely on N_j, the number of events reported with Drug j, but the user is encouraged both to modify those rules as well as to add any other drugs to the model if they are of particular interest. Disproportionality results are only reported for those drugs included in the model, and the disproportionalities are solely a function of estimated coefficients from the model fit, with no Bayesian shrinkage aspect to the algorithm. The lack of Bayesian shrinkage means that, especially for rare events, the number of drugs for which estimates can be obtained from

ES-ELR is limited by concerns for computational estimability, statistical collinearity and reliability, and concern for issues of multiple comparisons. The RGPS-ELR estimation involves an empirical Bayes shrinkage prior for the coefficients, which allows many more degrees of freedom to be included in the model than are used in ES-ELR. The handling of report covariates also differs in RGPS from that in both ES-ELR and MGPS. ES-ELR includes covariate values as ordinary predictors in the regression model, included additively on the logistic or extended logistic scale. This means that interactions among covariates like age and sex are not modeled, whereas in MGPS all combinations of covariate values play a separate role in the Mantel-Haenzel stratification method. The RGPS approach can be viewed as a hybrid of these approaches, in which a Bayesian clustering of the strata based on all combinations is performed in order to group strata that seem to have similar response frequencies, and then the resulting groups determine separately estimated intercepts for the regression model.

In the second step, expected counts for each drug in the database are estimated based on the RGPS-ELR model computed in the first step. In the third step, a two parameter gamma Poisson shrinkage algorithm is used to compute adjusted relative reporting rates and their associated confidence intervals based on the observed counts (N_j) and the expected counts (E_j) computed in the second step. This three step process is repeated for each event (response) being computed.

Step One: Selecting Predictors and Fitting the ELR

The RGPS-ELR analysis does not use all the drugs that are available and for which later steps eventually produce values of E_j. An automatic variable selection process selects drugs to be included in the RGPS-ELR analysis based on their event rates. The RGPS-ELR analysis also includes extra covariates such as age, sex and year. However, these covariates are not modeled in the usual way, e.g., as 0-1 indicator variables representing different covariate categories. Rather, in the RGPS-ELR analysis each covariate category combination is considered a stratum, and these strata are then aggregated into several groups which we call grouped-strata. Assuming there are C covariates each consisting of g_c categories, c = 1, …, C, then the total number of strata (covariate category combinations) considered is $K = \prod_c g_c$. These K strata are then partitioned into G groups each containing strata with similar event rates. The event rates are determined by a local gamma-Poisson shrinkage model. Let *s* be a stratum defined by a specific covariate category combination, N_s be the number of events reported for stratum s, and n_s be the number of reports included in stratum s. Define $p_0=N_+/n_+$ to be the prior probability for the event, and $e_s=p_0n_s$ the expected number of events for stratum s. Then, a gamma-Poisson shrinkage algorithm with a single parameter gamma prior Gamma(γ_0, γ_0) (mean 1) is used to estimate the posterior mean of the event rate (reporting ratio) for stratum s as

$$\lambda_s = (N_s + \gamma_0) / (e_s + \gamma_0)$$

where the hyperparameter γ_0 is estimated empirically using all the K pairs (N_s,e_s). The K strata are then ordered according to $\lambda_1>\lambda_2>\lambda_3> \ldots>\lambda_K$ and partitioned into groups of successive strata according to this ordering such that each group has approximately the same number of observed events. Specifically, let

$$N_{min} = \max(10, \sqrt{N_+} / 2)$$

Then, starting with the first (highest frequency) stratum, the strata are pooled until at least N_{min} observed events are included, after which a new pooling begins for the second group until at least N_{min} events have been included in that group, and continuing until all K strata have been included in G groups. With this algorithm, all but the final group will have N_{min} or more observed events. The rationale is to have fewer groups for rarer events and to have groups that are about equally informative as to their average event frequency. This method also groups many low-frequency strata together while doing less grouping for the high-frequency strata, which are of greater interest for determining safety signals. Each of these G grouped-strata are included in the ELR model as separate indicator variables, or, equivalently, the regression model has a separate intercept for each of the G groups.

The criterion for including a drug predictor in the ELR analysis is based on its group-stratified event rate, similar to the stratified RR computed by MGPS.

Let g(i) be the strata-group associated with report *i* (*i* ranges over all reports). We define the prior grouped-strata adjusted event probability for report *i* as

$$p_i = \left(\left(\sum_{s\in g(i)} N_s\right) + p_0\right) / \left(\left(\sum_{s\in g(i)} n_s\right) + 1\right)$$

At this stage, p_i is the same for all reports within a group of strata. The estimate for each group is shrunk slightly toward p_0 by adding p_0 and 1 to the numerator and denominator above. This mainly serves to ensure that every p_i is greater than 0 and less than 1. Let $X_{ij}=1$ if drug j is included in report i, $X_{ij} = 0$ otherwise, and let N_j be the number of events reported with drug j. Define $f_j=\Sigma_i X_{ij}\cdot p_i$ to be the grouped-strata adjusted expected number of events for drug j, where j ranges over all drugs. The (N_j, f_j) pairs are input into a gamma-Poisson shrinkage algorithm with a Gamma(δ_0, δ_0) prior to estimate the hyperparamter δ_0. The posterior mean of the grouped-strata adjusted event rate is thus

$$\mu_j=(N_j + \delta_0)/(f_j + \delta_0)$$

and the lower and upper 99% posterior limits $\mu_{j.01}$ and $\mu_{j.99}$ are computed using the gamma distribution

$$\text{Gamma}(N_j + \delta_0, f_j + \delta_0)$$

Define $\mu_0 = \text{median}(\mu_j)$ and $\mu_{j.q}$ to be the q^{th} percentile of the above gamma distribution. A drug j is then included in the model if:

$$\text{abs}(N_j - f_j) > 2.5 \text{ and } (\mu_{j.01} > \mu_0 \text{ or } \mu_{j.99} < \mu_0).$$

Prior distributions for the ELR coefficients

For the model including J drugs, the assumed probability that report i will have an event is

$$p_i = P_\alpha(\beta_{0g(i)} + \Sigma_j X_{ij}\beta_j) \qquad 0 < \alpha < 1$$

where P_α is the extended logistic probability function, β_{0g} is the intercept term for reports in the gth grouping of strata, and β_j is the coefficient for drug j. The formula for P_α is

$$P_\alpha(z) = 2\alpha / [1 + \exp(-2(1 - \alpha)z)] \qquad [z \leq 0]$$
$$P_\alpha(z) = 2\alpha - 1 + 2(1- \alpha)/[1 + \exp(-2\alpha z)] \qquad [z \geq 0]$$

which reduces to the standard logistic distribution function when $\alpha = ½$. The estimation of the ELR parameter α is achieved by first fitting the model for a few prespecified values of α and then optimizing the product of prior*likelihood with respect to α. The likelihood, L, is the standard product of binomials:

$$\log(L) = \Sigma_i[Y_i \log(p_i) + (1 - Y_i) \log(1 - p_i)]$$

where Y_i is 1 if the ith report contains the response event, 0 otherwise, and p_i depends on the parameters α, β_{0g}, and β_j. The prior distributions for β_{0g} and β_j are independent and centered at their null hypothesis values, assuming no effects of strata or drugs. That is, the estimates of β_{0g} are shrunk toward z_0, where

$$P_\alpha(z_0) = p_0 \qquad \text{(where } z_0 \text{ depends on both } \alpha \text{ and } p_0\text{)}$$

and the strength of the shrinkage is determined by the parameter γ_0, estimated above. The estimates of β_j are shrunk toward 0, and the strength of the shrinkage is determined by the parameter δ_0, estimated above. Since the relationship between these parameters and the p_i depends on the ELR parameter α, we define the prior distributions as beta distributions depending on transformations of the βs to probabilities depending on P_α. Let $\gamma^* = \gamma_0(1 - p_0)/p_0$ and let $\delta^* = \delta_0(1 - p_0)/p_0$. The resulting prior density L_0 is a product of beta distributions defined as

$$\log(L_0) = \Sigma_g[\gamma_0 \log(P_\alpha(\beta_{0g})) + \gamma^* \log(1 - P_\alpha(\beta_{0g}))] + \Sigma_j[\delta_0 \log(P_\alpha(z_0 + \beta_j)) + \delta^* \log(1 - P_\alpha(z_0 + \beta_j))]$$

Since both (γ_0, γ^*) and (δ_0, δ^*) are proportional to $(p_0, 1 - p_0)$, this implies that L_0 will be maximum when all the values of $P_\alpha(\,)$ in the above expression are equal to p_0, which occurs when every $\beta_{0g} = z_0$ and every $\beta_j = 0$.

Note that the values (γ_0, δ_0) used in the definition of the prior L_0 above were derived from gamma-Poisson fits that involved modeling the by-strata counts N_s and the DEC counts N_j (with an abuse of notation, since these are actually two sets of counts). We are applying these hyperparameters from much simpler models to define the prior distributions for the RGPS-ELR model. The justification is that the parameters of the conjugate gamma distributions in the gamma-Poisson model have interpretations very similar to those of the parameters of the conjugate beta distributions in the beta-binomial model defined by L_0. The value γ_0 was derived above from a gamma-Poisson model involving the effect of stratification, with one interpretation as the number of prior pseudo-observations of the Poisson response event that just fit the null hypothesis of probability p_0 per report within each stratum, and we see from the form of $\log(L_0)$ above that now γ_0 could be interpreted as that number of pseudo-observations of a binomial response event concordant to the same null hypothesis, thus justifying its value here. Similarly, the value of δ_0 was estimated from a simple gamma-Poisson model involving the effect of drugs on the response counts, and one interpretation of the resulting gamma(δ_0, δ_0) prior is that δ_0 is the number of pseudo-observations of the DEC consistent with the null hypothesis of no drug effect, with the same interpretation for the analogous binomial model in the definition of L_0. The difference is that the ELR model involves modeling the concomitancy of the drugs in each report, unlike the two gamma-Poisson models. Thus we are assuming that the empirical Bayes priors for the simpler models can be useful when applied to the ELR model, and also that the prior L_0, which depends on the ELR parameter α, can be used as is for each value of α.

The resulting ELR estimation provides a model for how the probability of a response event being in a report depends on the patient covariates and the set of drugs in that report. However, the coefficients from this model are not used directly to estimate drug-event disproportionalities. Rather the model is merely used to adjust for concomitant drugs as described in Step Two.

Step Two: Computing Expected Counts E for Every Drug

The next step is to compute a set of baseline or expected counts E_j for every drug available, including those that were used as drug predictors in the RGPS-ELR model $j = 1, \ldots, J_0$, as well as those drugs that were not included as predictors in the RGPS-ELR

model. So in this case j runs over all drugs. This computation uses the results of the RGPS-ELR analysis, namely a set of J_0 drug coefficients β, a set of G group-strata coefficients β_0, and the ELR parameter α. We will use the subscript i to index the reports that are included in the analysis. As above, let $X_{ij} = 1$ if potential drug predictor j is included in report i, $X_{ij} = 0$ otherwise, and g(i) be the stratum-group associated with report i. Now define p_i as the predicted probability that report i will include the response, based on the formula fit by the RGPS-ELR. The prediction formula can be represented as

$$p_i = P_\alpha(\mu_i = \beta_{0_{g(i)}} + \Sigma_j X_{ij}\beta_j)$$

where P_α is the function that links the linear predictor μ_i to the probability scale and β_j and $\beta_{0_{g(i)}}$ are the estimated coefficients for the drugs and intercepts, where the intercept depends on which grouped-stratum g(i) report i belongs to. For standard logistic regression ($\alpha = ½$)

$$P_{1/2}(\mu) = 1/(1 + \exp(-\mu))$$

while for extended logistic regression the formula is more complicated and depends on α as defined above. In the formula for p_i, the summation over j can be thought of as summing only over the predictors actually included in the RGPS-ELR model, or, equivalently, assume the summation runs over all potential predictors but that the β_j for drugs not included in the model are set to 0. This latter interpretation is assumed now, so that j runs over all the potential drug predictors in the summation but that $\beta_j = 0$ for drugs not included in the RGPS-ELR model.

The values of E_j, j = 1, …, J are computed by summing over i (i.e. across all reports) so that we can represent

$$E_j = \Sigma_i e_{ij}, \qquad \text{where} \quad e_{ij} = X_{ij} P_\alpha(\mu_i - \beta_j) \qquad \text{for } j = 1, \ldots, J$$

The factor X_{ij} makes $e_{ij} = 0$ if report i does not include predictor j. Also note that $e_{ij} = p_i$ if $X_{ij} = 1$ but predictor j is a drug not included in the RGPS-ELR model, since then $\beta_j = 0$.

The sum of the e_{ij}, E_j, is interpreted as the expected number of reports having the response when drug j is present under the null hypothesis that $\beta_j = 0$, but adjusted for all the (other) predictors used in the regression model. The non Bayesian relative reporting ratio is just $RR_j = N_j/E_j$.

In both the current MGPS and in ES-ELR, drug-event pairs where the count $N_j = 0$ are excluded. This exclusion was done for computing and storage space efficiency (especially for 3D and higher runs) in MGPS, and for estimability in ES-ELR. Our recommendation

is to include drugs having 0 counts in RGPS as potential predictors for which the values E_j are computed. Although such estimated relative reporting ratios won't have a large value, they can be useful for other purposes, for example to compare drug values across adverse events.

Step Three: Computing the Bayesian Shrinkage Estimates and Confidence Intervals

For each response, the (N_j, E_j) pairs from the previous step are input into a gamma-Poisson shrinkage algorithm. The prior distributions are assumed to be simple gamma distributions rather than a mixture of two gamma distributions as is done in MGPS. Specifically, a two-parameter gamma Poisson model is used to produce shrinkage estimates, where the prior distribution of the relative reporting ratios is assumed to be

$$\text{Gamma}(\gamma, \delta)$$

and where the (N_j, E_j) pairs are used to estimate the hyperparameters γ and δ. The posterior mean of a drug relative reporting ratio is then $EBRRR_j = (N_j + \gamma)/(E_j + \delta)$, and RRR05 and RRR95 are computed using the appropriate gamma distribution $\text{Gamma}(N_j + \gamma, E_j + \delta)$. The four hyperparameters $(\gamma_0, \delta_0, \gamma, \delta)$ will differ for each response variable, just as the RGPS-ELR parameter α does. They should be stored with the coefficients and made available to the user as in the current ES-ELR implementation.

Note the recommendation to replace the use of the posterior geometric mean (EBGM) by the posterior mean (EBRRR) in RGPS. Historically, we recommended posterior geometric mean rather than posterior mean in MGPS at a time when EB05 was not being computed and we felt that the EBGM would be preferred because it would be smaller and thus more conservative, less subject to false alarms. Now EB05 fulfills this role. In more general epidemiology applications, statisticians don't estimate the geometric mean of a relative risk, they estimate the mean of a relative risk. The introduction of RGPS might serve as an opportunity to change back to the more intuitive measure. Maybe the difference would also help remind the user that they are using the new program. [But possibly a switch would only serve to confuse the previous users of MGPS.]

Empirical Assessment of RGPS

Signal Detection Statistics- A Comparison with MGPS

We compare MGPS and RGPS to identify differences between the two methods. As part of this evaluation we provide a high-level summary of the differences between the two methods with respect to the disproportionality estimates (relative reporting ratios) they produce, and aim to identify cases (DECs) for which the two methods produce contradictory results. Contradictory results are identified by examining whether or not the two methods produce disproportionality confidence intervals which overlap one another. Interesting cases are those in which the methods produce non-overlapping confidence intervals. These are also cases which are potential indicators of confounding and masking biases that are detectable by RGPS, but not by MGPS.

The comparison is based on the public release version of AERS (and SRS) covering the period from 1968 through 2011Q3 (4,784,337 reports). We focus only on those events (preferred terms) having at least 100 reports, and on DECs having at least 5 reports. This limits our comparison to DECs with tighter nominal confidence intervals, and thus allows us to focus on differences that are likely attributed to the methods. The stratification variables used for the computation of MGPS and RGPS include, age (0-1, 2-4, 5-12, 13-16, 17-45, 46-75, 76-85, >85), gender, and year of report. The age and gender strata include an extra category "unknown".

Table 1 provides high-level summary statistics of the differences between MGPS and RGPS with respect to the disproportionality estimates, EBGM, EB05, EB95, EBRRR, RRR05, and RRR95. Differences are measured on the log scale. The table suggests that on average RGPS produces more conservative (smaller) disproportionality estimates than MGPS, and that on average RGPS produces tighter confidence intervals than MGPS. The average amount of reduction in the disproportionality estimates and the confidence intervals is roughly 35% and 50% respectively (on the linear scale).

Table 1. high-level summary statistics of the differences between MGPS and RGPS

	log (EBGM/EBRRR)	log(EB05/RRR05)	log [(EB95-EB05)/(RRR95-RRR05)]
Min.	-3.91	-3.92	-3.82
1st Qu.	-0.01	0.00	0.06
Median	0.24	0.25	0.33
Mean	0.30	0.30	0.40
3rd Qu.	0.54	0.54	0.65
Max.	6.74	6.74	6.79
Sd	0.54	0.52	0.59

Table 2 provides summary statistics (counts) related to cases (DECs) where the two methods produce non-overlapping confidence intervals or suggest contradictory signaling conclusions based on the commonly used default thresholds values of 1 and 2. The table shows that for roughly 20% of the DECs examined the two methods produce non-overlapping confidence intervals, of which the majority (18%) represent cases where MGPS produces larger disproportionality estimates (RRR95< EB05). That is, cases where the lower bound EB05 of MGPS is larger than the upper bound RRR95 of RGPS. For roughly 5% of the DECs the methods suggest contradictory signaling conclusions. The majority of these DECs (4%) represent cases were RGPS suggests that there is no signal (RGPS' confidence interval covers the null hypothesis, i.e., RRR05<1<RRR95), whereas the lower bound of MGPS suggests a potential signal (EB05>2). These are also the cases that potentially identify confounding biases undetectable by MGPS, but accounted for by RGPS. By the same token the table shows that there are cases (very few) that represent potential masking biases, which are accounted for by RGPS. Overall, it appears that RGPS reports many fewer signals than MGPS using the default threshold value of 2. Whether this represents better specificity or worse sensitivity merits further examination, but the next experiment provides some insights into this question.

Table 2. Difference between MGPS and RGPS based on confidence intervals.

Condition	Number of DECs with condition		Type of potential bias
Baseline			
Events with at least 100 reports and DECs with at least 5 reports	1,041,404	100%	
CI overlap			
RRR95 < EB05	192,486	18%	
EB95 < RRR05	18,847	2%	
RRR05 < 1 < RRR95 < EB05	103,991	10%	confounding
EB05 < 1 < EB95 < RRR05	894	0%	masking
Signals based on default thresholds			
EB05 ≥ 2	127,852	12%	
RRR05 ≥ 2	28,311	3%	
RRR05 < 1 < RRR95 < EB05 and EB05 ≥ 2	38,956	4%	confounding
EB05 < 1 < EB95 < RRR05 and RRR05 ≥ 2	82	0%	masking

Performance Evaluation Based on the OMOP Gold Standard

We evaluate the diagnostic performance of RGPS, and compare it with performance of other algorithms based on the OMOP gold standard[17, 18]. The OMOP gold standard consists of a total 398 test cases, of which 164 (41%) are positive test cases and 234 (59%) are negative controls. Each test case corresponds to a drug-event pair. The entire gold standard spans 181 unique drugs and 4 unique events, acute myocardial infarction, acute renal failure, acute liver injury, and gastrointestinal bleeding. The majority of drugs in the gold standard belong to the drug classes of NSAIDs, antibiotics, antidepressants, ACE inhibitors, beta blockers, antiepileptics, and glucose lowering drugs. Each test case (a drug-event pair) is classified as a positive test case or negative test case (control) based on the following criteria:

Positive test cases

- Event listed in Boxed Warning or Warnings/Precautions section of active FDA structured product label.
- Drug listed as 'causative agent' in Tisdale et al, 2010: "Drug-Induced Diseases"[19]
- Literature review identified no powered studies with refuting evidence of effect

Negative control

- Event not listed anywhere in any section of active FDA structured product label
- Drug not listed as 'causative agent' in Tisdale et al, 2010: "Drug-Induced Diseases"
- Literature review identified no powered studies with evidence of potential positive association

The AERS data used in this evaluation corresponds to the public release version of AERS (and SRS) covering the period from 1968 through 2011Q3. The drugs making up the gold standard are specified at the ingredient level. Therefore, drug names in AERS were normalized at the single ingredient level, so as to be consistent with the drug specification of the gold standard. In particular, an AERS report mentioning a combination drug with two or more ingredients was treated as if each ingredient had been mentioned separately in the report. Each event in the gold standard is defined by a group of MedDRA preferred terms (PTs). OMOP provides alternative definitions for each event ranging from broad to narrow (more specific) definitions. We used the broadest definition for each event. To match these definitions we created user-defined (custom) event terms within Empirica Signal. Of the 181 drugs appearing in the gold standard, 163 matched exactly with single ingredients appearing in AERS, and 17 drugs were manually matched. One drug (endopeptidases) did not appear in the AERS drugs dictionary, resulting in 4 negative test cases being removed from the gold standard. The remaining 394 test cases were used in this analysis.

The methods evaluated include: RGPS, ES-ELR, ES-LR (logistic regression), MGPS, PRR, and ROR as implemented in Empirica Signal v7.3. The stratification variables used for the computation of RGPS, ES-ELR, ES-LR, and MGPS include age (0-1, 2-4, 5-12, 13-16, 17-45, 46-75, 76-85, >85), gender, and year of report. The age and gender strata included an extra category- "unknown". The computation of PRR and ROR did not use stratification. ES-LR and ES-ELR were configured to include a set of 300 drugs as predictors, of which 181 are the drugs defining the gold standard and the remaining are automatically selected by Empirica Signal (the drugs most frequently reported with the response AE and having a minimum of 5 such reports). For each of these methods we examine two disproportionality values, the point estimate centrality measure (EBRRR, ELROR, LROR, EBGM, PRR, ROR), and the lower bound of its associated 90% confidence interval (RRR05, ELR05, LR05, EB05, PRR05, ROR05).

Of the 394 tests cases used in this analysis, 9 drug-event pairs (including 3 positive test cases) were not reported in AERS. Each method evaluated was assigned a disproportionality value equal to 0 for each of these 9 test cases.

To separate the intrinsic properties of the methods from their threshold implementation we use the threshold agnostic performance metric—the area under the receiver operating characteristic curve (AUC) metric, as the main performance metric with which we compare diagnostic performance.

Figure 1 displays a performance (diagnostic accuracy) comparison based on the AUC metric. The performance of each method is evaluated using the two disproportionality measures described above. The evaluation suggests that RGPS outperforms the competing methods in terms of diagnostic accuracy, including the main comparators MGPS and ES-ELR, and often by a significant margin (e.g., 10% for RGPS vs. MGPS). Based on the RRR05 measure RGPS results in an AUC=0.86, whereas MGPS (EB05) and ES-ELR (ELR05) result in AUC=0.79 and AUC=0.83 respectively. While the AUC confidence intervals display a pattern of overlap between methods, p-values for the hypothesis of no difference between AUCs, based on a test for correlated ROCs, suggest that the differences between RGPS, MGPS, and ES-ELR are significant at the standard 5% level. The figure also suggests that the lower bound measures are better signaling proxies.

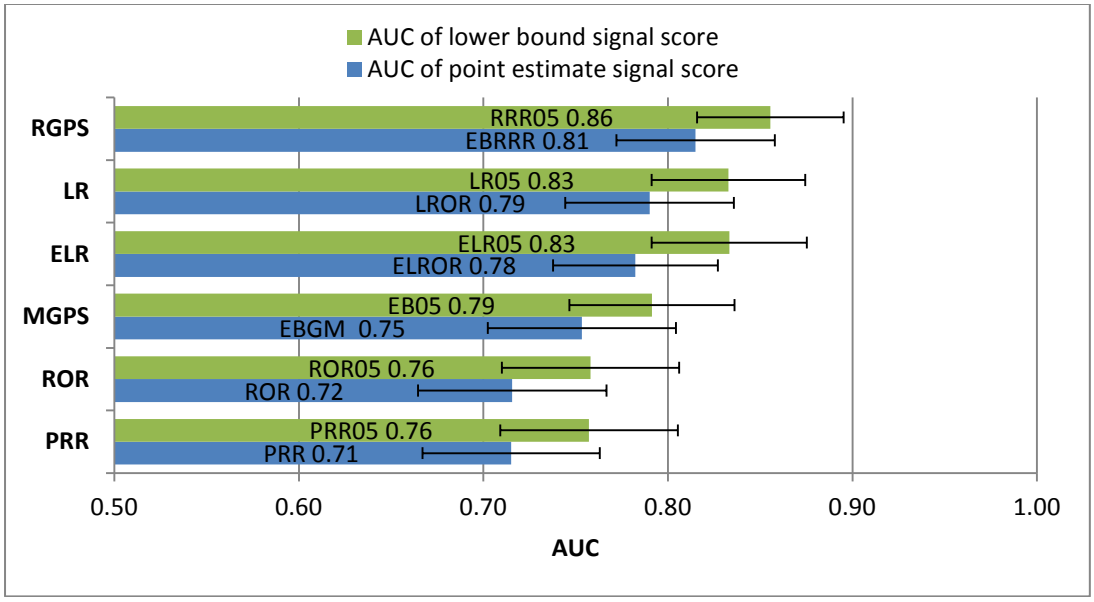

Figure 1. Performance comparison based on the AUC metric and the OMOP gold standard

Figure 2 displays the receiver operating characteristic curves produced by each method. The figure demonstrates that RGPS (RRR05) generally provides greater specificity at a given level of sensitivity than any of the other methods for almost the complete range of operating scenarios.

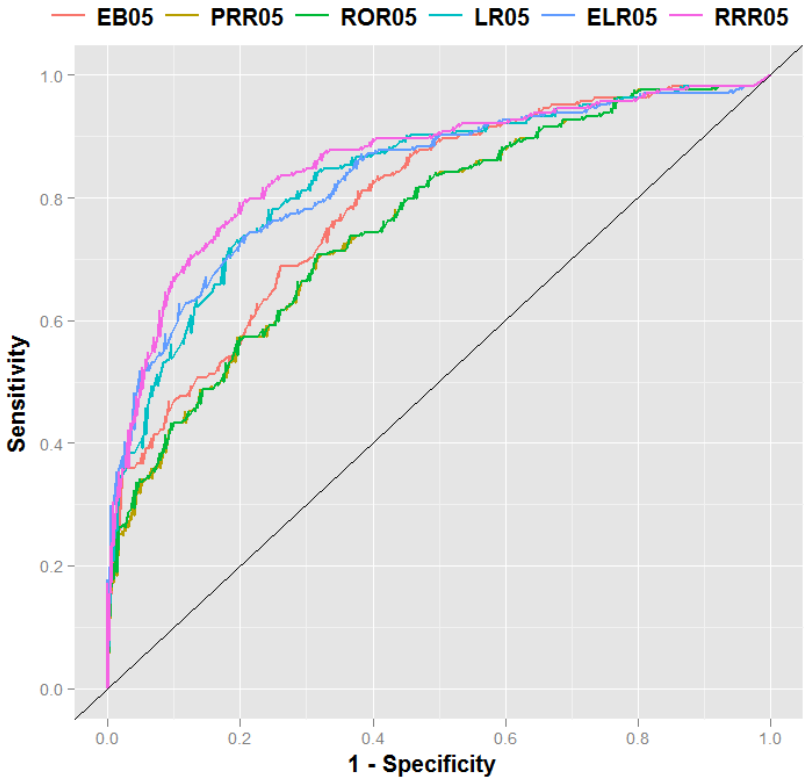

Figure 2. Receiver Operating Characteristic curves for the EB05, PRR05, ROR05, LR05, ELR05, RRR05 measures, based on the OMOP gold standard.

Screening for Drug-Drug Interactions in RGPS

The calculations for estimating the effect of drug-drug interactions on the response are based on reusing the ELR estimates from the main effects estimation, so there is no additional model fitting, but there are steps for computing expected counts and for Bayesian shrinkage of ratios of counts that represent the potential effects of drug interactions. Our measure of drug-drug interaction is to compare the relative reporting rate when both drugs are present to the higher of the two drugs' relative reporting rates. Calculating, storing and presenting interaction estimates for all drug1-drug2-AE triples would be burdensome and potentially wasteful, and so we also suggest strategies for filtering the set of all triples. A primary strategy is to only look at pairs of drugs (j, k) among drugs where both N_j and N_k (the cross-counts of each of the two drugs with the response AE) are at least equal to some value N_Int_{min}. (For example, $N\_Int_{min} = 25$.) Thus, assume that J_{int} such drugs have been identified, so that interaction calculations are performed for all $J_{int}(J_{int} - 1)/2$ drug pairs (j, k).

Expected counts

Let n_{jk} be the number of reports containing both drug j and drug k, and let N_{jk} be the number of those reports that also mention the response AE of interest. Let $EBRRR_j$ and $EBRRR_k$ be the corresponding disproportionality estimates for the two drugs based on the RGPS analysis discussed in Section 2. As in that Section, define, for each report i,

$$p_i = P_\alpha(\mu_i = \beta 0_{g(i)} + \Sigma_j X_{ij}\beta_j) \qquad \text{[in the summation, j varies over } \textit{all} \text{ drugs]}$$

Now define E_{jk} as the expected value of N_{jk} under the null hypothesis that *both* drug j and drug k have no relative reporting ratio effect, namely

$$E_{jk} = \Sigma_i X_{ij} X_{ik} P_\alpha(\mu_i - \beta j - \beta_k) \qquad 1 \le j < k \le J_{int}$$

where, as before, we take β_j or β_k as 0 if the corresponding drug was not in the ELR model. As mentioned above, we interpret "no interaction" to mean that the disproportionality ratio for both drugs (that is, the ratio N_{jk}/E_{jk}) is expected to be the higher of $EBRRR_j$ and $EBRRR_k$. Therefore, we define the no-interaction expected count as

$$E^*_{jk} = E_{jk} * \max(EBRRR_j, EBRRR_k)$$

Bayesian interaction estimates

There will be $J_{int}(J_{int} - 1)/2$ raw interaction ratios of the form

$$INTRR_{jk} = N_{jk}/E^{*}_{jk}$$

These ratios will have large relative sampling variances because many of the N_{jk} and/or E^{*}_{jk} will be small. Bayesian interaction ratios are estimated using yet another application of the gamma-Poisson model. We use a one-parameter prior distribution Gamma(γ_1, γ_1), having mean 1, as a model for the means of $INTRR_{jk}$, and estimate γ_1 by inputting the set of (N_{jk}, E^{*}_{jk}) into the empirical Bayes estimation. Then we have the posterior mean of the interaction ratio as

$$INTEB_{jk} = (N_{jk} + \gamma_1)/(E^{*}_{jk} + \gamma_1)$$

The posterior 5% and 95% limits, $INT05_{jk}$ and $INT95_{jk}$, are the corresponding quantiles of the distribution Gamma($N_{jk} + \gamma_1$, $E^{*}_{jk} + \gamma_1$).

Further filtering of interaction estimates

After computing the raw and Bayesian interaction ratios for all $J_{int}(J_{int} - 1)/2$ pairs of drugs for a given response AE, we may wish to suppress the "uninteresting" ones. This is justified by the fact that the empirical Bayes estimation largely corrects for reversion to the mean caused by excess variance. Since a complete run of RGPS may involve up to 15,000 MedDRA preferred terms (PTs) as response AEs, even the reduction to $J_{int}(J_{int} - 1)/2$ pairs for each AE may involve ungainly amounts of disk storage and an awkward user interface. We suggest only presenting interaction estimates if

$$INT05 > int05_{min} \quad \underline{\text{or}} \quad INT95 < int95_{max}$$

with default values $int05_{min} = 1$ and $int95_{max} = {}^{1}/_{3}$. (The value of searching for very low interaction ratios has not yet been established, but they often do occur and could merit further investigation.)

References

(1) Adverse Event Reporting System. http://www.fda.gov/cder/aers/default.htm (accessed Nov. 2012)

(2) Medical Dictionary for Regulatory Activities (MedDRA). http://www.meddramsso.com/ (accessed Nov. 2012)

(3) Bate A, Evans SJ. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol Drug Saf* 2009;18(6):427-436.

(4) Stephenson W, Hauben M. Data mining for signals in spontaneous reporting databases: proceed with caution. *Pharmacoepidemiol Drug Saf* 2007;16(4):359-365.

(5) Hauben M, Madigan D, Gerrits CM, Walsh L, van Puijenbroek EP. The role of data mining in pharmacovigilance. *Expert Opin Drug Saf* 2005;4(5):929-948.

(6) Almenoff JS, Pattishall EN, Gibbs TG, DuMouchel W, Evans SJ, Yuen N. Novel statistical tools for monitoring the safety of marketed drugs. *Clin Pharmacol Ther* 2007;82(2):157-166.

(7) Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. *Clin Pharmacol Ther* 2012;91(6):1010-1021.

(8) Evans SJW, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and Drug Safety* 2001;10(6):483-486.

(9) van Puijenbroek EP, Bate A, Leufkens HG, Lindquist M, Orre R, Egberts AC. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol Drug Saf* 2002;11(1):3-10.

(10) DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA Spontaneous Reporting System. *Am Stat* 1999;53(3):177-190.

(11) DuMouchel W, Pregibon D. Empirical bayes screening for multi-item associations. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining* 2001;67-76.

(12) DuMouchel W, Fram D, Yang X et al. Antipsychotics, glycemic disorders, and life-threatening diabetic events: a Bayesian data-mining analysis of the FDA adverse event reporting system (1968-2004). *Ann Clin Psychiatry* 2008;20(1):21-31.

(13) Berlin C, Blanch C, Lewis DJ et al. Are all quantitative postmarketing signal detection methods equal? Performance characteristics of logistic regression and Multi-item Gamma Poisson Shrinker. *Pharmacoepidemiol Drug Saf* 2012;21(6):622-630.

(14) Solomon R, DuMouchel W. Contrast media and nephropathy: findings from systematic analysis and Food and Drug Administration reports of adverse effects. *Invest Radiol* 2006;41(8):651-660.

(15) Caster O, Noren GN, Madigan D, Bate A. Large-scale regression-based pattern discovery: the example of screening the WHO global drug safety database. *Stat Analy Data Mining* 2010;3:197-208.

(16) http://www.oracle.com/us/industries/life-sciences/health-sciences-empirica-signal-364243.html (accessed Nov. 2012)

(17) Observational Medical Outcomes Partnership (OMOP). http://omop.fnih.org/ (accessed Nov. 2012)

(18) Stang PE, Ryan PB, Racoosin JA et al. Advancing the Science for Active Surveillance: Rationale and Design for the Observational Medical Outcomes Partnership. *Annals of Internal Medicine* 2010;153(9):600-W206.

(19) Tisdale J, Miller D. Drug-Induced Diseases: Prevention, Detection, and Management, 2nd edition. 2010. American Society of Health-System Pharmacists.

Oracle is committed to developing practices and products that help protect the environment

Regression-Adjusted GPS Algorithm (RGPS)
November 2012
Authors: William DuMouchel, Rave Harpaz

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200

oracle.com

Hardware and Software, Engineered to Work Together