Article

Verification of the "Upward Variation in the Reporting Odds Ratio Scores" to Detect the Signals of Drug–Drug Interactions

Yoshihiro Noguchi *[ID]**, Shunsuke Yoshizawa, Keisuke Aoyama, Satoaki Kubo, Tomoya Tachi and Hitomi Teramachi ***

Laboratory of Clinical Pharmacy, Gifu Pharmaceutical University, 1-25-4, Daigakunishi,
Gifu-shi 501-1196, Gifu, Japan; 185127@gifu-pu.ac.jp (S.Y.); 175003@gifu-pu.ac.jp (K.A.);
175050@gifu-pu.ac.jp (S.K.); tachi@gifu-pu.ac.jp (T.T.)
* Correspondence: noguchiy@gifu-pu.ac.jp or noguchi-yshr@umin.ac.jp (Y.N.); teramachih@gifu-pu.ac.jp (H.T.);
Tel.: +81-58-230-8100 (Y.N. & H.T.)

Abstract: The reporting odds ratio (ROR) is easy to calculate, and there have been several examples of its use because of its potential to speed up the detection of drug–drug interaction signals by using the "upward variation of ROR score". However, since the validity of the detection method is unknown, this study followed previous studies to investigate the detection trend. The statistics models (the Ω shrinkage measure and the "upward variation of ROR score") were compared using the verification dataset created from the Japanese Adverse Drug Event Report database (JADER). The drugs registered as "suspect drugs" in the verification dataset were considered as the drugs to be investigated, and the target adverse event in this study was Stevens–Johnson syndrome (SJS), as in previous studies. Of 3924 pairs that reported SJS, the number of positive signals detected by the Ω shrinkage measure and the "upward variation of ROR score" (Model 1, the Susuta Model, and Model 2) was 712, 2112, 1758, and 637, respectively. Furthermore, 1239 positive signals were detected when the Haldane–Anscombe 1/2 correction was applied to Model 2, the statistical model that showed the most conservative detection trend. This result indicated the instability of the positive signal detected in Model 2. The ROR scores based on the frequency-based statistics are easily inflated; thus, the use of the "upward variation of ROR scores" to search for drug–drug interaction signals increases the likelihood of false-positive signal detection. Consequently, the active use of the "upward variation of ROR scores" is not recommended, despite the existence of the Ω shrinkage measure, which shows a conservative detection trend.

Keywords: spontaneous reporting systems; drug–drug interaction; reporting odds ratio (ROR); Ω shrinkage measure

1. Introduction

To ensure the proper use of drugs, it is important to understand the related adverse events. However, pre-marketing randomized clinical trials focus on establishing the safety and efficacy of a single drug, rather than investigating drug–drug interactions. Therefore, patients using other drugs along with the drug being studied in a clinical trial are excluded from an investigation. However, unlike pre-marketing studies, in actual clinical practice, multiple drugs are generally used for treatment.

Recent reports have estimated that the proportion of adverse events caused by drug–drug interactions is approximately 30% of unexpected adverse events [1]. Considering the numerous reports on polypharmacy in treatment in recent years [2–6], early identification of adverse events that may be caused by drug–drug interactions is an important issue that should be addressed.

Spontaneous reporting systems, which play an important role in pharmacovigilance, are a source of information for the detection of previously unknown adverse events not identified in clinical trials, including adverse events caused by drugs in post-marketing.

Using spontaneous reporting systems, there are several reports of safety assessments that reflect real-world use in specific populations and clinical practice. However, the databases used in spontaneous reporting systems contain only cases of adverse events caused by the use of drugs and do not include the number of users of the drugs; therefore, the incidence of adverse events cannot be calculated. Thus, instead of incidence, the disproportionality analysis signals have been used to search for unknown adverse events [7–9]. The disproportionality analysis focuses on differences in the proportion of adverse event reports. If the reporting rate of the medicinal product of interest is high compared to the average reporting rate for all other medicinal products, this indicates that "the medicinal product and the adverse event have an association" [10]. For this analysis, a number of algorithms have also been reported to search for signals of drug–drug interactions [11,12]. Among these, the Ω shrinkage measure [13–15], proposed by Noren et al., is used by the World Health Organization Uppsala Monitoring Center (WHO-UMC), and previous studies have shown that it has the most conservative signal detection trend among signal detection methods based on frequentist statistics [16]. This study defined a conservative signal detection algorithm as one that detects few signals that are specific to that detection algorithm, and many signals that are common to other algorithms.

Surprisingly, however, there have only been few papers using the Ω shrinkage measure [17]; rather, several reports have used a detection method that extends the reporting odds ratio (ROR) score [18–20]. In addition, there is also a report that evaluated the ROR score of the concomitant use [21]. However, the validity of the analysis method using the ROR score has been questioned by Kuss et al., and it was discussed together with the results using another detection algorithm [22].

The ROR is an algorithm used by the Pharmaceuticals and Medical Devices Agency (PMDA) in Japan and the Pharmacovigilance Center (Lareb) in the Netherlands, which generally searches for adverse events caused by a single drug rather than drug–drug interactions [23].

A detection method that expands the ROR score evaluates a combination as having an increased risk of adverse events if the ROR score for the combination of drugs is higher than the ROR score for a single drug (i.e., if there is an upward variation in the ROR score). Although Susuta et al. were the first to propose a detection method that uses the "upward variation in the ROR score" as the signals for drug–drug interactions in Japan [18], the proposed cutoff values of the scores and the comparison with other methods for detecting signals of drug–drug interactions have not been sufficiently verified, and the validity of the signals detected by this analysis method is unknown. Therefore, in this study, we aimed to verify an analysis method that uses the "upward variation in the ROR scores" as the signals for drug–drug interactions, referring to similar previous studies that evaluated the detection tendency of search algorithms [16,24–26].

2. Materials and Methods

2.1. Data Sources

The dataset used for validation was the same dataset used in our previous study [16]. The dataset was created from the Japanese Adverse Drug Event Report database (JADER) data from the first quarter of 2004 to the fourth quarter of 2015. The JADER consists of four comma-separated values (csv) files as data tables: DEMO.csv (patient information), DRUG.csv (medicine information), HIST.csv (patient past history), and REAC.csv (AE event information). These files were linked by identification numbers, and the four csv files were combined using the identification numbers to create a database for analysis. In this study, we did not use patient information and past history, because we were only investigating the differences in detection trends between algorithms. Therefore, we removed this information when creating the verification dataset.

The Japanese authority, the PMDA, which owns these data, does not permit sharing the data directly. The latest version of the file can be accessed directly here: (http://

2.2. Targeted Drugs and Adverse Events

The drugs registered as "suspected drugs" in the verification dataset were considered as the drugs to be investigated. As in previous studies [16,18,24–26], the only adverse event targeted for this signal search was set to be the preferred term; that is, "Stevens-Johnson syndrome" (SJS) in the Medical Dictionary for Regulatory Activities Japanese version (MedDRA/J), which is registered in the verification dataset.

2.3. Statistical Models and Criteria

2.3.1. Ω Shrinkage Measure

The Ω shrinkage measure that detected the most conservative signal [16] is used as the control model in this study. The score of the Ω shrinkage measure was calculated from the number of reports (n_{111}) and its expected value (E_{111}), as shown in Table 1, using Equations (1)–(6). The lower limit of the 95% confidence interval (CI) for the Ω shrinkage measure was Ω_{025}, and a positive signal was considered to exist when $\Omega_{025} > 0$, as previously reported [13].

$$\Omega = log_2 \frac{n_{111} + 0.5}{E_{111} + 0.5} \tag{1}$$

where n_{111} is the number of reports and E_{111} is the expected value.

$$f_{00} = \frac{n_{001}}{n_{00+}}, \quad f_{10} = \frac{n_{101}}{n_{10+}}, \quad f_{01} = \frac{n_{011}}{n_{01+}}, \quad f_{11} = \frac{n_{111}}{n_{11+}} \tag{2}$$

where n is the number of reports shown in Table 1. For example, n_{111} is the number of reported target adverse event caused by *drug D_1* and *drug D_2*.

$$g_{11} = 1 - \frac{1}{\max\left(\frac{f_{00}}{1-f_{00}}, \frac{f_{10}}{1-f_{10}}\right) + \max\left(\frac{f_{00}}{1-f_{00}}, \frac{f_{01}}{1-f_{01}}\right) - \frac{f_{00}}{1-f_{00}} + 1} \tag{3}$$

when $f_{10} < f_{00}$, which denotes no risk of an adverse event caused by *drug D_1*, the most sensible estimator $g_{11} = \max (f_{00}, f_{01})$ is yielded and vice versa when $f_{01} < f_{00}$.

$$E_{111} = g_{11} \times n_{11+} \tag{4}$$

$$\text{Var}(\Omega_0) = \text{Var}\left(log_2 \frac{n_{111}}{E_{111}}\right) \approx \frac{1}{n_{111} \log(2)^2} \tag{5}$$

$$\Omega_{025} = \Omega - \frac{\phi(0.975)}{log(2)\sqrt{n_{111}}} \tag{6}$$

where ϕ (0.975) is 97.5% of the standard normal distribution.

Table 1. The 4×2 contingency table for signal detection.

	Target AE	Other AEs	Total
drug D_1 **and** *drug D_2*	n_{111}	n_{110}	n_{11+}
drug D_1 **without** *drug D_2*	n_{101}	n_{100}	n_{10+}
drug D_2 **without** *drug D_1*	n_{011}	n_{010}	n_{01+}
Neither *drug D_1* **or** *drug D_2*	n_{001}	n_{000}	n_{00+}
Total	n_{++1}	n_{++0}	n_{+++}

AE: adverse event, n: the number of reports (e.g., n_{111}: the number of target drugs (*drug D_1* and *drug D_2*) induced AE, n_{+++}: the number of all reports).

2.3.2. Upward Variation in Reporting Odds Ratio Scores

The ROR scores were calculated using the data from the number of reports shown in Table 2 and Equations (7) and (8). The lower limit of the 95% CI for ROR was ROR_{025}, and the upper limit was ROR_{975} [23].

$$ROR = \frac{N_{11} / N_{10}}{N_{01} / N_{00}} \tag{7}$$

$$ROR \text{ (95\% CI)} = e^{ln \text{ (ROR)} \pm 1.96 \sqrt{\frac{1}{N_{11}} + \frac{1}{N_{10}} + \frac{1}{N_{01}} + \frac{1}{N_{00}}}} \tag{8}$$

where N is the number of reports shown in Table 2. For example, N_{11} is the number of reported target adverse events caused by *drug D_1* and *drug D_2*.

Table 2. The 2×2 contingency table for signal detection.

	Target AE	Other AEs	Total
Target drug (s)	N_{11}	N_{12}	N_{1+}
Other drugs	N_{21}	N_{22}	N_{2+}
Total	N_{+1}	N_{+2}	N_{++}

AE: adverse event, N: the number of reports (e.g., N_{11}: the number of target drug-induced AE, N_{++}: the number of all reports).

Additionally, to calculate the ROR of *drug $D_1 \cap$ drug D_2*, *drug D_1* and *drug D_2*, replace it as follows: *drug $D_1 \cap$ drug D_2*: $N_{11} = n_{111}$, $N_{00} = n_{000} + n_{010} + n_{100}$, $N_{10} = n_{110}$, $N_{01} = n_{001} + n_{011} + n_{101}$, $N_{1+} = n_{11+}$, $N_{+1} = n_{++1}$, $N_{0+} = n_{00+} + n_{01+} + n_{10+}$, $N_{+0} = n_{++0}$; *drug D_1*: $N_{11} = n_{111} + n_{101}$, $N_{00} = n_{000} + n_{010}$, $N_{10} = n_{110} + n_{100}$, $N_{01} = n_{001} + n_{011}$, $N_{1+} = n_{11+} + n_{10+}$, $N_{+1} = n_{++1}$, $N_{0+} = n_{00+} + n_{01+}$, $N_{+0} = n_{++0}$; *drug D_2*: $N_{11} = n_{111} + n_{011}$, $N_{00} = n_{000} + n_{100}$, $N_{10} = n_{110} + n_{010}$, $N_{01} = n_{001} + n_{101}$, $N_{1+} = n_{11+} + n_{01+}$, $N_{+1} = n_{++1}$, $N_{0+} = n_{00+} + n_{10+}$, $N_{+0} = n_{++0}$.

The signal of drug–drug interactions detected using ROR has been recently reported in Japan as "the lower limit of the 95% CI of the ROR score when *drug D_1* and *drug D_2* are used together ($ROR_{025 \text{ drug } D1 \cap \text{ drug } D2}$) > 1" and "when *drug D_1* and *drug D_2* are used together ($ROR_{\text{drug } D1 \cap \text{ drug } D2}$) is greater than either the ROR score of *drug D_1* ($ROR_{\text{drug } D1}$) or the ROR score of *drug D_2* ($ROR_{\text{drug } D2}$), whichever is greater (Model 1)." The ROR score ratio (Model 1) is calculated using Equation (9).

$$\text{ROR score ratio (Model 1)} = \frac{ROR_{\text{drug } D_1 \cap \text{ drug } D_2}}{\max\left(ROR_{\text{drug } D_1}, \ ROR_{\text{drug } D_2}\right)} \tag{9}$$

Susuta et al. proposed "($ROR_{025 \text{ drug } D_1 \cap \text{ drug } D_2}$) > 1" and "ROR score ratio (Model 1) > 2" as signal detection criteria for drug–drug interactions [18]. This was evaluated as the "Susuta model."

However, Model 1 and the Susuta model failed to account for the overlap between the lower 95% CI of the ROR score for *drugs D_1* and *drug D_2* together and the upper 95% CI of the ROR score for either *drug D_1* or *drug D_2*, whichever is greater (Figure 1).
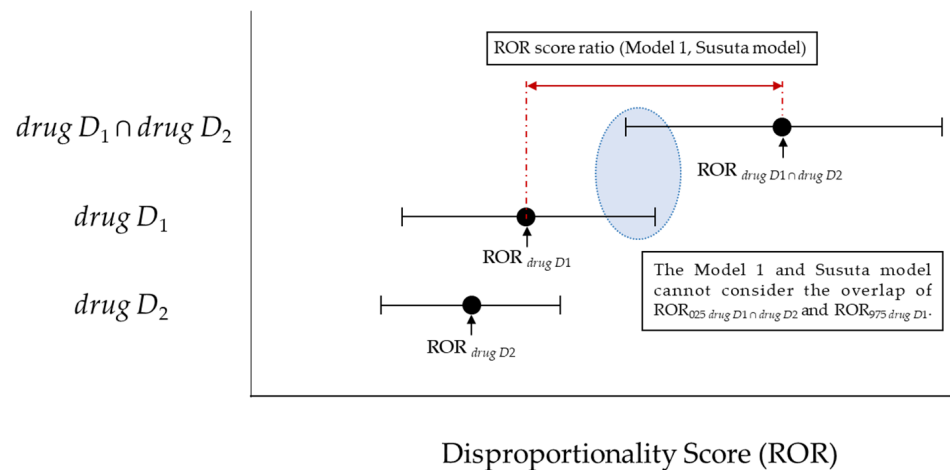
Disproportionality Score (ROR)

Figure 1. The association between the ROR score ratio (Model 1, Susuta model) and disproportionality score.

If such an overlap occurs, a risk signal may not be detected for concomitant use. Therefore, with reference to the interaction signal score (INTSS) [27] and concomitant signal score (CSS) [26], we used the criteria to detect if "($ROR_{025\ drug\ D1\ \cap\ drug\ D2}$) > 1" and "the lower limit of the 95% CI of the ROR of the combination of *drug D_1* and *drug D_2* ($ROR_{025\ drug\ D1\ \cap\ drug\ D2}$) is greater than either the upper limit of the 95% CI of *drug D_1* ($ROR_{975\ drug\ D1}$) or the upper limit of the 95% CI of the ROR of *drug D_2* ($ROR_{975\ drug\ D2}$), whichever is greater (Model 2)" (Figure 2, Equation (10)).

$$\text{ROR score ratio (model 2)} = \frac{ROR_{025\ drug\ D_1 \cap\ drug\ D_2}}{\max\left(ROR_{975\ drug\ D_1},\ ROR_{975\ drug\ D_2}\right)} \tag{10}$$

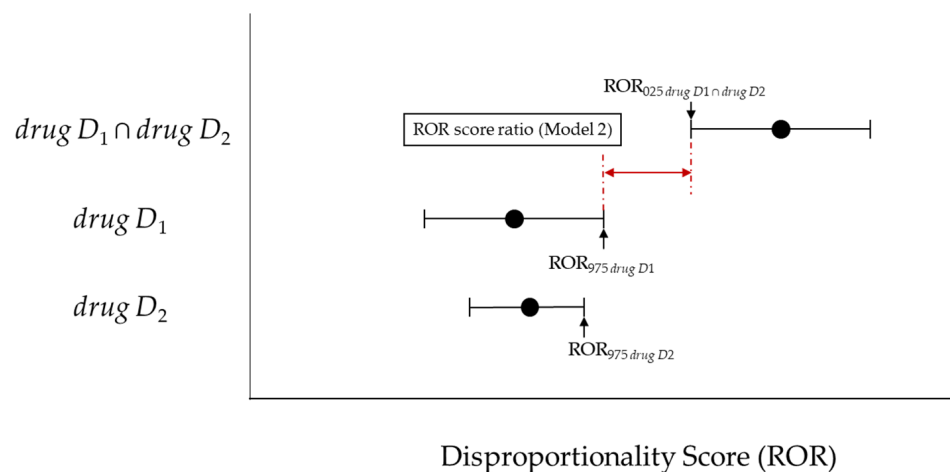
Disproportionality Score (ROR)

Figure 2. The association between the ROR score ratio (Model 2) and disproportionality score.

However, in the estimation of the ROR, if any one of the four cells in Table 2 is 0, the ROR will be 0 or ∞. In addition, if there is zero in the perimeter sum, the ROR cannot be defined. Furthermore, when estimating ROR when the sample size in a cell is small, as seen in a few reports and in this study, the effect of a single case in a small sample cell is very large, making the estimation unstable [28]. The Haldane–Anscombe 1/2 correction, which adds 1/2 to each cell, is known to solve this problem [29]. In this study, among the detection methods (Model 1, Susuta Model, and Model 2) that utilize the "upward variation in the ROR scores", the statistical model that showed the most conservative detection tendency was also analyzed with the Haldane–Anscombe 1/2 correction.

2.4. Targeted Drugs and Adverse Events

The signal similarity of each statistical detection method was evaluated using Cohen's kappa coefficient (κ), proportionate agreement for the positive rating (P_{pos}), and proportionate agreement for the negative rating (P_{neg}), as reported in previous studies [16,24]. Cohen's kappa coefficient and its 95% CI were obtained from Table 3 and Equations (11)–(14) [30].

Table 3. Agreement between the Ω shrinkage measure and target model.

		Ω Shrinkage Measure		Total
		Yes	**No**	
Target model	**Yes**	N_{yy}	N_{yn}	$N_{y.}$
	No	N_{ny}	N_{nn}	$N_{n.}$
Total		$N_{.y}$	$N_{.n}$	$N_{..}$

Target model: model 1; model 2; Susuta model; Yes: signal detection; No: not signal detection.

$$P_o = \frac{N_{yy} + N_{nn}}{N_{..}} \tag{11}$$

$$P_e = \frac{N_{y.}}{N_{..}} \times \frac{N_{.y}}{N_{..}} + \frac{N_{n.}}{N_{..}} \times \frac{N_{.n}}{N_{..}} \tag{12}$$

The number of N_{yy}, $N_{y.}$, $N_{.y}$, N_{nn}, $N_{n.}$, $N_{.n}$ and $N_{..}$ can be obtained from Table 3.

$$Cohen's\ kappa\ coefficient\ (\kappa) = \frac{P_o - P_e}{1 - P_e} \tag{13}$$

$$95\%\ CI\ of\ kappa\ coefficient = \kappa \pm 1.96\sqrt{\frac{P_o \times (1 - P_o)}{N_{..} \times (1 - P_e)^2}} \tag{14}$$

The P_{pos} and P_{neg} can be obtained from Equations (15) and (16) and Table 3 [27].

$$P_{pos} = \frac{2N_{yy}}{N_{y.} + N_{.y}} \tag{15}$$

$$P_{neg} = \frac{2N_{nn}}{N_{n.} + N_{.n}} \tag{16}$$

However, the number of N_{yy}, $N_{y.}$, $N_{.y}$, N_{nn}, $N_{n.}$ and $N_{.n}$ can be obtained from Table 3.

3. Results

Table 4 shows the number of reports (n_{111}) and the number of signals detected for each of the Ω shrinkage measure, Model 1, the Susuta model, Model 2, and Model 2 (Haldane–Anscombe 1/2 correction).

Of the 374,327 cases used in this study, there were 3924 *drug D_1–drug D_2*-SJS, wherein the number of positive signals was 712 for the Ω shrinkage measure [16]; 2,112 for Model 1; 1758 for the Susuta model; and 637 for Model 2. In Model 1, the Susuta Model, and Model 2, 1239 positive signals were detected when Haldane–Anscombe 1/2 correction was applied to Model 2, which, consequently, was the statistical model that showed the most conservative detection trend.

Table 5 shows the κ, P_{pos}, and P_{neg} between the Ω shrinkage measure and the detection methods of Model 1, the Susuta model, Model 2, and Model 2 (Haldane–Anscombe 1/2 correction).

Table 4. The *drug D₁–drug D₂*-SJS combinations detected as a potential signal in five frequency statistical models.

Statistical Models	Signal (Y/N)	Number (%) of Combinations					
		$n_{111} < 3$	$n_{111} = 3$	$n_{111} = 4$	$n_{111} = 5$	$n_{111} > 5$	Total
Model 1	Y	1363 (47.8)	243 (65.3)	159 (66.8)	110 (80.3)	237 (72.3)	2112 (53.8)
	N	525 (18.4)	48 (12.9)	23 (9.7)	14 (10.2)	42 (12.8)	652 (16.6)
	N (no criterion)	961 (33.7)	81 (21.8)	56 (23.5)	13 (9.5)	49 (14.9)	1160 (29.6)
Susuta model	Y	1142 (40.1)	207 (55.6)	136 (57.1)	97 (70.8)	176 (53.7)	1758 (44.8)
	N	746 (26.2)	84 (22.6)	46 (19.3)	27 (19.7)	103 (31.4)	1006 (25.6)
	N (no criterion)	961 (33.7)	81 (21.8)	56 (23.5)	13 (9.5)	49 (14.9)	1160 (29.6)
Model 2	Y	239 (8.4)	106 (28.5)	84 (35.3)	74 (54.0)	134 (40.9)	637 (16.2)
	N	1649 (57.9)	185 (49.7)	98 (41.2)	50 (36.5)	145 (44.2)	2127 (54.2)
	N (no criterion)	961 (33.7)	81 (21.8)	56 (23.5)	13 (9.5)	49 (14.9)	1160 (29.6)
Corrected Model 2	Y	678 (23.8)	178 (47.8)	118 (49.6)	92 (67.2)	173 (52.7)	1239 (31.6)
	N	2171 (76.2)	194 (52.2)	120 (50.4)	45 (32.8)	155 (47.3)	2685 (68.4)
	N (no criterion)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Total		2849	372	238	137	328	3924

SJS: Stevens–Johnson syndrome; n_{111}: the number of reported target adverse event caused by *drug D₁* and *drug D₂* (see Table 1).

Table 5. The Cohen's kappa coefficient and proportionate agreement for positive rating (P_{pos}) and that for negative rating (P_{neg}) between Ω Shrinkage measure and four frequency statistical models.

	Model 1	Susuta Model	Model 2	Corrected Model 2 *
Ω Shrinkage measure	κ: 0.074 95% CI: 0.058–0.089 P_{pos}: 0.325 P_{neg}: 0.621	κ: 0.152 95% CI: 0.135–0.168 P_{pos}: 0.371 P_{neg}: 0.711	κ: 0.495 95% CI: 0.475–0.514 P_{pos}: 0.581 P_{neg}: 0.913	κ: 0.479 95% CI: 0.459–0.493 P_{pos}: 0.597 P_{neg}: 0.867

κ: Cohen's kappa coefficient; P_{pos}: proportionate agreement for positive rating; P_{neg}: proportionate agreement for negative rating,
*: Model 2 with the Haldane–Anscombe 1/2 correction.

The detection method with the most similar detection tendency as the Ω shrinkage measure was Model 2 (κ: 0.495; 95% CI: 0.475–0.514; P_{pos}: 0.581; P_{neg}: 0.913), followed by the Susuta model (κ: 0.152; 95% CI: 0.135–0.168; P_{pos}: 0.371; P_{neg}: 0.711) and Model 1 (κ: 0.074; 95% CI: 0.058–0.089; P_{pos}: 0.325; P_{neg}: 0.621).

However, the similarity of the detection trend of Model 2 decreased when the Haldane–Anscombe 1/2 correction was applied (κ: 0.476; 95% CI: 0.459–0.492; P_{pos}: 0.597; P_{neg}: 0.867).

Figure 3 shows the relationship between the number of reports (n_{111}), the ROR score ($ROR_{drug\ D1 \cap drug\ D2}$), and the ROR score ratio when two *drug D₁* and *drug D₂* were used in combination.

In cases where the number of reports ($N_{11} = [n_{111}]$) was less than 10, there were many cases where the $ROR_{drug\ D1 \cap drug\ D2}$ exceeded 50, and the maximum score was over 700. The ROR score ratio (Model 1) was also inflated by the inflation of the $ROR_{drug\ D1 \cap drug\ D2}$, the maximum score of ROR score ratio (Model 1) was over 450 (Figure 3).

However, some of the combinations with the very high $ROR_{drug\ D1 \cap drug\ D2}$ or the ROR score ratios do not meet the detection criteria for the Ω shrinkage measure, e.g., N_{11}: 11, $ROR_{drug\ D1 \cap drug\ D2}$: 57.45, ROR score ratio (Model 1): 3.43, Ω_{025}: −0.18.
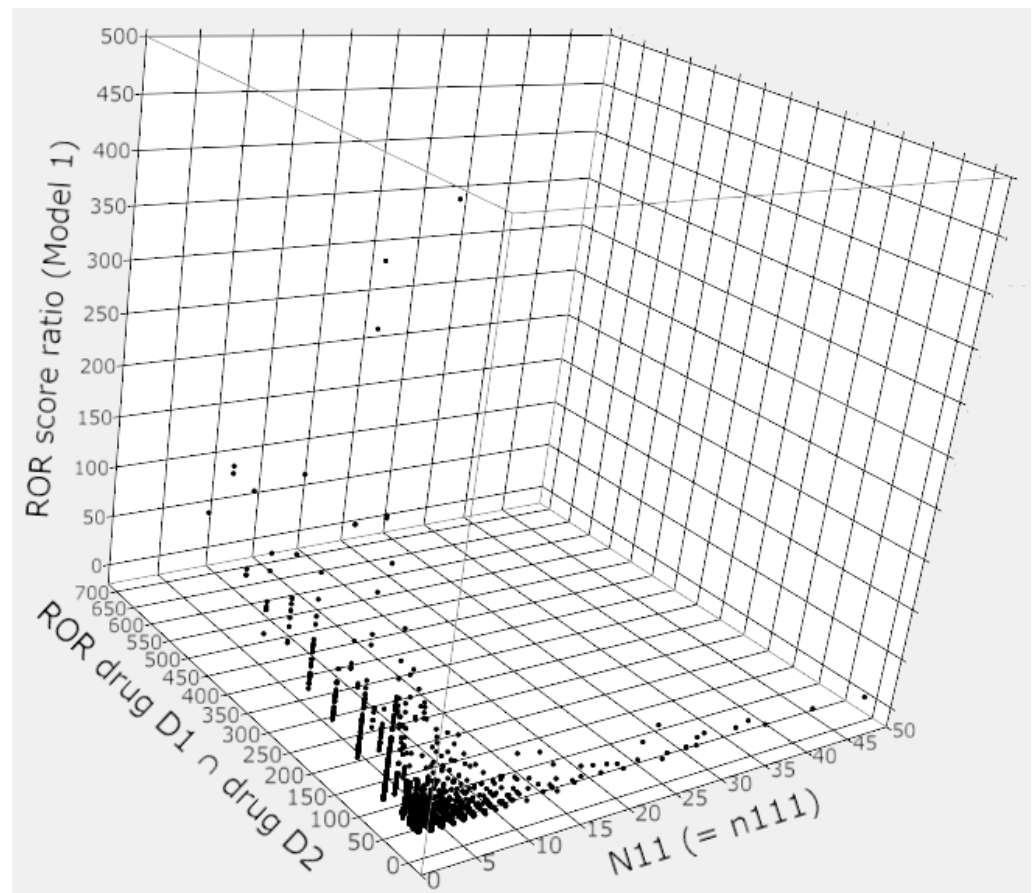
Figure 3. Relationship between the number of reports (N_{11} [= n_{111}]), $ROR_{drug\ D1 \cap drug\ D2}$, and ROR score ratio (Model 1).

4. Discussion

In this study, we examined a method of using the "upward variation in the ROR score" as a signal of drug–drug interactions. Strictly, detection trends should be compared for all adverse events registered in the database. Unfortunately, even with fast and powerful computers, calculating signal scores for all combinations of multiple drugs and adverse events can be expected to take an enormous amount of time; thus, targeting all combinations was not a realistic research method. Therefore, in this study, as in a previous study [16,18,24], the target adverse event was SJS.

Of the 3924 combinations of *drug D_1*–*drug D_2*-SJS, 2112 positive signals were detected in Model 1, 1758 in the Susuta model, and 637 in Model 2. All of the combinations detected in Model 2 also had a positive signal detected in Model 1. Additionally, of the 637 combinations detected in Model 2, 636 pairs also had a positive signal detected in the Susuta model (Table 4).

The Ω shrinkage measure detected 712 positive signals [16]. The detection method with the most similar detection tendency as the Ω shrinkage measure was Model 2 (Table 5), with 392 positive signals in common. However, among the detection methods that utilize the "upward variation in ROR score", even in Model 2, which showed the most conservative detection tendency, even one of the four cells in Table 2 became a 0 cell, and there were 1160 pairs in which the ROR score ratio could not be calculated (Table 4). The problem of negative signals due to the inability to calculate this score was solved by Haldane–Anscombe 1/2 correction; however, the number of positive signals increased by 602 from that before the correction, for a total of 1239 (Table 4). This result indicated the instability of the positive signal detected in Model 2.

Furthermore, even Model 2, which had the highest similarity to the Ω shrinkage measure in this study, had a similarity lower than that between the Ω shrinkage measure

and the Chi-square model [31] in a previous study using the same verification database and the same targeted adverse event [16].

The signals obtained from JADER, the database used in this study, require verification to confirm that they are true adverse events, and true data are needed to evaluate the validity of the detection results. However, it is not possible to prepare true data, including the data of "unknown" adverse events.

In the verification of a signal detection method for a single drug, Szarfman et al. reported [32] that the information in the medical package inserts was set as the "true" data and evaluated using receiver operating characteristic (ROC) curves with different cutoff values for signal scores [32]. However, as Watanabe et al. pointed out, it is unclear whether the information in the package insert is the only "true" data, and this verification method has its limitations [33].

In fact, considering that signal detection methods are designed to search for "unknown" adverse events, it is not appropriate to evaluate the performance of detection methods using only "known" information. Thus, following the combinations of drug–drug interactions described in the medical package insert, as in Kubota et al. and our previous studies [16,24], we only compared the detection trends of each statistical model in this study. This study is not affected by patient background, as it only shows the difference between the calculated results of each statistical model and the interpretation of the signal scores. This is the same as previous studies [16,18,24].

However, this limitation of not being able to provide "true" data makes it difficult to determine whether the signal detection results from the detection method using the "upward variation in ROR score" are overestimated or whether the signal detection results from the Ω shrinkage measure are underestimated.

This study was conducted under limited conditions and may require further investigation using simulation data. However, unlike the Bayesian confidence neural network (BCPNN) [34] and the empirical Bayes geometric mean (EBGM) [35] based on Bayesian statistical methods, the ROR based on frequency-based statistical methods is prone to signal score inflation when the number of reports is small [36], leading to unstable detection results [37].

In general, the number of reports for the combination of two drugs ($drug\ D_1 \cap drug\ D_2$) will be less than the number of reports for single drug use ($drug\ D_1$ or $drug\ D_2$); therefore, the ROR score ($ROR_{drug\ D1 \cap drug\ D2}$) is likely to be inflated. In fact, in this study, there were many cases of inflation of $ROR_{drug\ D1 \cap drug\ D2}$, as shown in Figure 3. The $ROR_{drug\ D1 \cap drug\ D2}$ is a numerator in the equation of Model 1 and the Susuta model, and such signal score inflation may make it easier to detect false-positive signals of drug–drug interactions.

Further, it is known that the 95% CIs are wider when the number of reports is small, and in this study, the method that did not take into account the overlap between the 95% CI of the signal score for the combination of two drugs ($drug\ D_1 \cap drug\ D_2$) and the 95% CI of the signal score for single drug ($drug\ D_1$ or $drug\ D_2$) use (Model 1, Susuta model) resulted in a higher likelihood of detecting false-positive signals than Model 2, the method that took into account the overlap.

The results indicated that when the number of reports was small, as in the case of drug–drug interactions, it was important to consider the overlap between the 95% CI of the signal score for the combination of two drugs ($drug\ D_1 \cap drug\ D_2$) and the 95% CI of the signal score for single drug use ($drug\ D_1$ or $drug\ D_2$) when calculating the ROR score ratio.

The spontaneous reporting systems, which are used for disproportionality analysis, consist only of spontaneously "reported" cases and, naturally, do not include those that occur but are not reported. Additionally, reports are known to contain a variety of biases [38–40]. Therefore, the calculated signal score is also affected by the biases. Spontaneous reports often lack information on concomitant medications, leading to underestimation of the drug–drug interaction signal score in some cases. Thus, the use of any statistical analysis method cannot overcome the inherent qualitative and quantitative limitations of spontaneous reporting systems [37].

The ROR is simple to calculate, and utilizing the "upward variation in the ROR score" may speed up the detection of drug–drug interaction signals. However, even Model 2, which showed the most conservative detection tendency among the detection methods utilizing the "upward variation in the ROR score", did not show stable signal detection results due to the small number of reports, making it difficult to deny the overestimation of positive signals of drug–drug interactions, as in Model 1 and the Susuta model.

It is known that there are many attentive points when analyzing a spontaneous reporting database, and various analysis algorithms have been proposed [10]. Considering the history of the development of the BCPNN and the EBGM based on Bayesian statistical models [37] to avoid signal score inflation that detects a large number of false-positive signals in the detection of single drug signals, even though there is not only the Ω shrinkage measure used in WHO-UMC, but also several alternative detection methods for signal detection of drug–drug interactions [10,11], there is no reason to actively recommend the use of "upward variation in the ROR score", which is more likely to detect false-positive signals.

5. Conclusions

Recently, many patients have been concomitantly using drugs, and in order to use drugs appropriately, it is necessary to screen not only for single drugs but also for safety signals such as drug–drug interactions. In this study, Model 2, which corrects the problems contained in Model 1 by referring to INTSS and CSS, was also examined. However, the ROR scores based on the frequency-based statistics are easily inflated; thus, the use of the "upward variation of ROR scores" in either statistic model to search for drug–drug interaction signals increases the likelihood of false-positive signal detection. Although, some researchers have used "upward variation of ROR scores" (the active use of this algorithm is not recommended), because of the existence of the Ω shrinkage measure, which shows a conservative detection trend. In order to reduce false-positive signals, the selection of appropriate detection algorithms is desired.

Author Contributions: Conceptualization, Y.N.; funding acquisition, Y.N.; investigation, Y.N., S.Y., K.A. and S.K.; methodology, Y.N. and T.T.; project administration, Y.N. and H.T.; supervision, Y.N.; validation, Y.N. and T.T.; visualization, Y.N.; writing—original draft, Y.N.; writing—review and editing, Y.N., T.T. and H.T. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This study was used the dataset from the first quarter of 2004 to the fourth quarter of 2015 from the Japanese Adverse Drug Event Report database (JADER). However, the Japanese authority, the Pharmaceuticals and Medical Devices Agency (PMDA), which owns this data, does not permit sharing the data directly. Therefore, it can be accessed directly here: (http://www.info.pmda.go.jp/fukusayoudb/CsvDownload.jsp (accessed date: 10 August 2021)) (in Japanese only).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Iyer, S.V.; Harpaz, R.; LePendu, P.; Bauer-Mehren, A.; Shah, N.H. Mining clinical text for signals of adverse drug-drug interactions. *J. Am. Med. Inf. Assoc.* **2013**, *21*, 353–362. [CrossRef] [PubMed]
2. Gnjidic, D.; Hilmer, S.N.; Blyth, F.M.; Naganathan, V.; Waite, L.; Seibel, M.J.; McLachlan, A.J.; Cumming, R.G.; Handelsman, D.J.; Le Couteur, D.G. Polypharmacy cutoff and outcomes: Five or more medicines were used to identify community-dwelling older men at risk of different adverse outcomes. *J. Clin. Epidemiol.* **2012**, *65*, 989–995. [CrossRef] [PubMed]
3. Davies, L.E.; Spiers, G.; Kingston, A.; Todd, A.; Adamson, J.; Hanratty, B. Adverse Outcomes of Polypharmacy in Older People: Systematic Review of Reviews. *J. Am. Med. Dir. Assoc.* **2020**, *21*, 181–187. [CrossRef] [PubMed]

4. Sugioka, M.; Tachi, T.; Mizui, T.; Koyama, A.; Murayama, A.; Katsuno, H.; Matsuyama, T.; Aoyama, S.; Osawa, T.; Noguchi, Y.; et al. Effects of the number of drugs used on the prevalence of adverse drug reactions in children. *Sci. Rep.* **2020**, *10*, 21341. [CrossRef] [PubMed]

5. Liu, Y.; Wang, R.; Huang, R.; Cao, Y.; Wang, J.; Feng, D. Influencing factors and their relationships of risk perception and decision-making behaviour of polypharmacy in patients with chronic diseases: A qualitative descriptive study. *BMJ Open* **2021**, *11*, e043557. [CrossRef] [PubMed]

6. Uragami, Y.; Takikawa, K.; Kareki, H.; Kimura, K.; Yamamoto, K.; Iihara, N. Effect of number of medications and use of potentially inappropriate medications on frailty among early-stage older outpatients. *J. Pharm. Health Care Sci.* **2021**, *7*, 15. [CrossRef]

7. Noguchi, Y.; Esaki, H.; Murayama, A.; Sugioka, M.; Koyama, A.; Tachi, T.; Teramachi, H. Association between dipeptidyl peptidase-4 inhibitor and aspiration pneumonia: Disproportionality analysis using the spontaneous reporting system in Japan. *Eur. J. Clin. Pharmacol.* **2020**, *76*, 299–304. [CrossRef]

8. Yokoyama, S.; Wakamoto, S.; Tanaka, Y.; Nakagawa, C.; Hosomi, K.; Takada, M. Association between Antipsychotics and Osteoporosis Based on Real-World Data. *Ann. Pharmacother.* **2020**, *54*, 988–995. [CrossRef]

9. Toyoshima, M.; Noguchi, Y.; Otsubo, M.; Tachi, T.; Teramachi, H. Differences in detected safety signals between benzodiazepines and non-benzodiazepine hypnotics: Pharmacovigilance study using a spontaneous reporting system. *Int. J. Med. Sci.* **2021**, *18*, 1130–1136. [CrossRef]

10. Noguchi, Y.; Tachi, T.; Teramachi, H. Detection algorithms and attentive points of safety signal using the spontaneous reporting systems as a clinical data source. *Brief. Bioinform.* **2021**. [CrossRef]

11. Vilar, S.; Friedman, C.; Hripcsak, G. Detection of drug-drug interactions through data mining studies using clinical sources, scientific literature and social media. *Brief. Bioinform.* **2018**, *19*, 863–877. [CrossRef]

12. Noguchi, Y.; Tachi, T.; Teramachi, H. Review of Statistical Methodologies for Detecting Drug-Drug Interactions Using Spontaneous Reporting Systems. *Front. Pharmacol.* **2019**, *10*, 1319. [CrossRef] [PubMed]

13. Noren, G.N.; Sundberg, R.; Bate, A.; Edwards, I.R. A statistical methodology for drug-drug interaction surveillance. *Stat. Med.* **2008**, *27*, 3057–3070. [CrossRef] [PubMed]

14. Uppsala Monitoring Centre (UMC). The UMC Measures of Disproportionate Reporting. Available online: https://www.who-umc.org/media/164041/measures-of-disproportionate-reporting_2016.pdf (accessed on 6 July 2021).

15. Hult, S.; Sartori, D.; Bergvall, T.; Hedfors Vidlin, S.; Grundmark, B.; Ellenius, J.; Norén, G.N. A Feasibility Study of Drug-Drug Interaction Signal Detection in Regular Pharmacovigilance. *Drug Saf.* **2020**, *43*, 775–785. [CrossRef] [PubMed]

16. Noguchi, Y.; Tachi, T.; Teramachi, H. Comparison of signal detection algorithms based on frequency statistical model for drug-drug interaction using spontaneous reporting systems. *Pharm. Res.* **2020**, *37*, 86. [CrossRef] [PubMed]

17. Noguchi, Y.; Takaoka, M.; Hayashi, T.; Tachi, T.; Teramachi, H. Antiepileptic combination therapy with Stevens-Johnson syndrome and toxic epidermal necrolysis: Analysis of a Japanese pharmacovigilance database. *Epilepsia* **2020**, *61*, 1979–1989. [CrossRef]

18. Susuta, Y.; Takahashi, Y. Safety risk evaluation methodology in detecting the medicine concomitant use risk which might cause critical drug rash. *Jpn. J. Pharmacoepidemiol.* **2014**, *19*, 39–49. [CrossRef]

19. Asakura, M.; Yamaori, S. Analyses of Risk for Liver Injury Associated with Dipeptidyl Peptidase-4 Inhibitors and of Influence of Co-administered Drugs on Vildagliptin-induced Liver Injury Using the Japanese Adverse Drug Event Report (JADER) Database. *Jpn. J. Pharm. Diabet.* **2020**, *9*, 123–134. [CrossRef]

20. Okumura, S.; Ishihara, Y.; Echizen, H. Detection of Signals of Convulsion-inducing Effect for Fluoroquinolone Antibiotics Alone and with Co-administration of Nonsteroidal Anti-inflammatory Drugs in Association with Aging Using Japanese Adverse Drug Event Reporting Database. *Jpn. J. Clin. Pharmacol. Ther.* **2019**, *50*, 203–209. [CrossRef]

21. Antonazzo, I.C.; Poluzzi, E.; Forcesi, E.; Salvo, F.; Pariente, A.; Marchesini, G.; De Ponti, F.; Raschi, E. Myopathy with DPP-4 inhibitors and statins in the real world: Investigating the likelihood of drug-drug interactions through the FDA adverse event reporting system. *Acta Diabetol.* **2020**, *57*, 71–80. [CrossRef]

22. Kuss, O.; Rathmann, W. A missed protective drug-drug interaction of DPP-4 inhibitors and statins on myopathy risk. *Acta Diabetol.* **2020**, *57*, 113–114. [CrossRef] [PubMed]

23. van Puijenbroek, E.P.; Bate, A.; Leufkens, H.G.M.; Lindquist, M.; Orre, R.; Egberts, A.C.G. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol. Drug Saf.* **2002**, *11*, 3–10. [CrossRef] [PubMed]

24. Kubota, K.; Koide, D.; Hirai, T. Comparison of data mining methodologies using Japanese spontaneous reports. *Pharmacoepidemiol. Drug Saf.* **2004**, *13*, 387–394. [CrossRef]

25. Noguchi, Y.; Tachi, T.; Teramachi, H. Subset Analysis for Screening Drug-Drug Interaction Signal Using Pharmacovigilance Database. *Pharmaceutics* **2020**, *12*, 762. [CrossRef]

26. Noguchi, Y.; Aoyama, K.; Kubo, S.; Tachi, T.; Teramachi, H. Improved Detection Criteria for Detecting Drug-Drug Interaction Signals Using the Proportional Reporting Ratio. *Pharmaceuticals* **2021**, *14*, 4. [CrossRef] [PubMed]

27. Suzuki, A.; Yuen, N.A.; Ilic, K.; Miller, R.T.; Reese, M.J.; Brown, H.R.; Ambroso, J.I.; Falls, J.G.; Hunt, C.M. Comedications alter drug-induced liver injury reporting frequency: Data mining in the WHO VigiBase™. *Regul. Toxicol. Pharmacol.* **2015**, *72*, 481–490. [CrossRef]

28. Breslow, N.E. Odds ratio estimators when the data are sparse. *Biometrika* **1981**, *68*, 73–84. [CrossRef]

29. Lawson, R. Small sample confidence intervals for the odds ratio. *Commun. Stat. Simul. Comput.* **2004**, *33*, 1095–1113. [CrossRef]

30.　Feinstein, A.R. Evaluating concordances. In *Principles of Medical Statistics*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2002; pp. 407–436.

31.　Gosho, M.; Maruo, K.; Tada, K.; Hirakawa, A. Utilization of chi-square statistics for screening adverse drug-drug interactions in spontaneous reporting systems. *Eur. J. Clin. Pharmacol.* **2017**, *73*, 779–786. [CrossRef] [PubMed]

32.　Szarfman, A.; Machado, S.G.; O'Neill, R.T. 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*, 381–392. [CrossRef]

33.　Watanabe, H.; Matsushita, Y.; Watanabe, A.; Maeda, T.; Nukui, K.; Ogawa, Y.; Sawa, J.; Maeda, H. Early Detection of Important Safety Information -Recent Methods for Signal Detection. *Jpn. J. Biomet.* **2004**, *25*, 37–60. [CrossRef]

34.　Bate, A.; Lindquist, M.; Edwards, I.R.; Olsson, S.; Orre, R.; Lansner, A.; De Freitas, R.M. A Bayesian neural network method for adverse drug reaction signal generation. *Eur. J. Clin. Pharmacol.* **1998**, *54*, 315–321. [CrossRef]

35.　DuMouchel, W. Bayesian Data Mining in Large Frequency Tables, with an Application to the FDA Spontaneous Reporting System. *Am. Stat.* **1999**, *53*, 177–190.

36.　Fujita, T. Signal Detection of Adverse Drug Reactions. *Jpn. J. Pharmacoepidemiol.* **2009**, *14*, 27–36. [CrossRef]

37.　Urushihara, H. Basic Dos and Don'ts in Applying Signal Detection Methods to Spontaneous Reporting Systems Databases. *Jpn. J. Drug Inform.* **2019**, *21*, 135–141. [CrossRef]

38.　Weber, J. Epidemiology of adverse reactions to nonsteroidal anti-inflammatory drugs. *Adv. Inflamm. Res.* **1984**, *6*, 1–7.

39.　Pariente, A.; Gregoire, F.; Fourrier-Reglat, A.; Haramburu, F.; Moore, N. Impact of safety alerts on measures of disproportionality in spontaneous reporting databases: The notoriety bias. *Drug Saf.* **2007**, *30*, 891–898. [CrossRef] [PubMed]

40.　Wang, H.W.; Hochberg, A.M.; Pearson, R.K.; Hauben, M. An experimental investigation of masking in the US FDA adverse event reporting system database. *Drug Saf.* **2010**, *33*, 1117–1133. [CrossRef] [PubMed]