Use of the Zipf-Mandelbrot Law in Modelling US FDA Adverse Reactions

Glen Atlas *, Sunil Dhar , George Tewfik , Dhvani Shihora

Posted Date: 8 April 2026

doi: 10.20944/preprints202604.0494.v1

Keywords: Zipf-Mandelbrot law; pareto distribution; adverse drug reactions

 doi:10.20944/preprints202604.0494.v1

Article

Use of the Zipf-Mandelbrot Law in Modelling US FDA Adverse Reactions

Glen Atlas [1,*], Sunil Dhar [2], George Tewfik [1] and Dhvani Shihora [1]

[1] Department of Anesthesiology, Rutgers New Jersey Medical School, Newark, NJ, USA

[2] Department of Mathematics, New Jersey Institute of Technology, Newark, NJ, USA

* Correspondence: atlasgm@njms.rutgers.edu

Abstract

Objective: The purpose of this preliminary study was to evaluate the use of the Zipf-Mandelbrot (ZM) law to mathematically model the percentage occurrence of adverse drug reactions (ADRs), as a function of rank, reported to the US FDA Adverse Event Monitoring System (AMES). **Methods**: Six commonly used hospital-based medications were examined. Nonlinear curve fitting of the two ZM coefficients was utilized to model the percentage occurrence of ADRs in a hierarchical or rank order for each drug examined. **Results**: The reported complications and their associated occurrence rates for all six medications were accurately modelled using the ZM law. Those medications which have a greater percentage of reported ADRs within their first ten ranks have a greater negative slope. Furthermore, a natural logarithmic transformation of both the reported FDA data and the predicted values utilizing the ZM law demonstrated a consistent statistically significant near-linear correlation. The ratio of the coefficients of the ZM law, $a \cdot b^{-1}$, was also found to be a potentially useful index which allows for describing and comparing the overall shape of the medication-specific distributions. **Conclusions**: Based upon this preliminary examination, the ZM law appears to be applicable to the mathematical modeling of US FDA reported ADRs. Additional research to assess and utilize this law for the analysis, economic management, and possible improvement in patient outcome may be warranted.

Keywords: Zipf-Mandelbrot law; pareto distribution; adverse drug reactions

Introduction

Adverse drug reactions (ADRs) represent a significant source of morbidity and mortality. Therefore, these medication-based complications negatively impact both the cost and quality of health care [1,2].

The Zipf-Mandelbrot (ZM) law has been applied to mathematically model a wide range of phenomena including linguistics [3], insurance risk [4], scientific citations [5], web hits [6], economics [7], and urban population [8]. Furthermore, it is frequently referred to as the Pareto-Zipf distribution [9].

This paper examines the applicability of the ZM law to US FDA reported ADRs for the following medications: fentanyl, propofol, albumin, succinylcholine, ketamine, and isoflurane. For the purposes of this research, the ZM law will be represented as:

$$f(r) = \frac{100}{(r+b)^{a}}, \qquad r = 1, 2 \ldots N. \tag{1}$$

Where $f(r)$ represents the percentage occurrence of each FDA adverse event associated with a sequential rank, r. Moreover, r is a dimensionless natural integer ranging from 1 to N. Where N represents the total number of unique medication-specific ADRs.

 doi:10.20944/preprints202604.0494.v1

Note that the maximum value for $f(r)$ always occurs at $r = 1$ which is the most frequently reported adverse reaction. Whereas $f(N)$ represents the percentage occurrence of the least reported adverse reaction which occurs at $r = N$.

Furthermore, $f(r)$, a, and b are all positive real numbers. Also, $a > 1$ and $b \geq 0$. It should be noted that small positive values for a which are less than one do not generate a sufficient skewness to adequately represent these clinical data.

In addition, coefficients a and b are both dimensionless and are determined using curve-fitting. Note that when $b = 0$ the ZM law reduces to a basic power law distribution.

The following limits are therefore straightforward and are fundamental in understanding the properties of the ZM law:

$$\lim_{a\to\infty}\frac{100}{(r+b)^{a}}=0, \tag{2}$$

$$\lim_{a\to 0}\frac{100}{(r+b)^{a}}=100, \tag{3}$$

$$\lim_{b\to\infty}\frac{100}{(r+b)^{a}}=0, \tag{4}$$

and:

$$\lim_{r\to\infty}\frac{100}{(r+b)^{a}}=0. \tag{5}$$

By definition, the sum of the total occurrences of ADRs expressed in percent form is:

$$\sum_{r=1}^{N} f(r)=100\%. \tag{6}$$

However, the reader should be cognizant that the original US FDA data is supplied or downloaded in decimal form:

$$\sum_{r=1}^{N}\frac{f(r)}{100}=1. \tag{7}$$

For this study, the non-linear ZM law is statistically evaluated by comparing it to the medication-specific reported data:

$$\underbrace{f(r)}_{\substack{\text{Reported}\\ \text{adverse}\\ \text{reactions}}} \approx \underbrace{\frac{100}{(r+b)^{a}}}_{\substack{\text{Predicted}\\ \text{adverse}\\ \text{reactions}}}, \qquad r=1,2...N. \tag{8}$$

Therefore, the reported adverse events for each medication's ADRs, $f(r)$, will be compared to the predicted adverse events by utilizing the ZM law and by curve fitting coefficients a and b.

Additionally, the ZM law can be differentiated with respect to rank:

$$\frac{df}{dr}=\frac{-a\cdot 100}{(r+b)^{(a+1)}}=\frac{-a\cdot f(r)}{(r+b)}. \qquad r=1,2...N. \tag{9}$$

The above interrelationship has dimensionless units of $\% \cdot r^{-1}$ and readily explains the slope of $f(r)$ at each rank. Note that $\frac{df}{dr}$ is consistently negative.

Inspection of the above derivative also demonstrates that for a given value of r increasing values of a and/or b yields an overall diminution in the magnitude of the slope of $f(r)$.

Therefore:

$$\lim_{a\to\infty}\frac{df}{dr}=0, \tag{10}$$

and:

$$\lim_{b\to\infty}\frac{df}{dr}=0. \tag{11}$$

 doi:10.20944/preprints202604.0494.v1

Furthermore, for given values of both a and b an increasing value of r will also generate an overall diminution in the magnitude of the slope of $f(r)$:

$$\lim_{r\to\infty}\frac{df}{dr}=0. \tag{12}$$

To compare different medications and their distributions of ADRs, it is also helpful to quantitate a medication-specific average slope within a specific range of ranks, $r = j, j+1, \ldots n.$ Note that: $j \geq 1$ and $n \leq N.$

$$\text{Average } \frac{df}{dr} \text{ within a range of ranks} = \frac{d\bar{f}}{dr} = \frac{\sum_{r=j}^{n}\left(\frac{-a}{(r+b)^{(a+1)}}\right)}{(n-j+1)}\cdot 100, \qquad 1 \leq j \leq n. \tag{13}$$

This is equivalent to:

$$\frac{d\bar{f}}{dr} = \frac{\sum_{r=j}^{n}\left(\frac{-a\cdot f(r)}{(r+b)}\right)}{(n-j+1)}, \qquad 1 \leq j \leq n. \tag{14}$$

Note that the area under a curve is approximately proportional to the summation of the points along a curve. Therefore, the definite integral of the ZM law, I, is an analogous function to the summation of the points along the $f(r)$ curve:

$$I = 100\int_{r_i}^{r_f}\frac{dr}{(r+b)^a} = \frac{100}{(1-a)}\left[(r_f+b)^{(1-a)} - (r_i+b)^{(1-a)}\right]. \tag{15}$$

Equivalently:

$$I = \frac{100}{(1-a)}\left[\frac{1}{(r_f+b)^{(a-1)}} - \frac{1}{(r_i+b)^{(a-1)}}\right]. \tag{16}$$

Thus, I illustrates how coefficients a and b interact with the summation process. The above equation also has the following clinical constraints: $r_f \geq r_i$, $b \geq 0$, and $a > 1$. Inspection of I demonstrates that, for a given value of a, decreasing values of b will lead to a greater value of I and therefore a greater summation over the specified range of r_i to r_f. The value of I and the range-based summation will also increase with decreasing values of a for a given value of b.

In addition to $\frac{d\bar{f}}{dr}$, the summation of sequential points along a specific ZM distribution, $f(r)$, provides useful information when comparing different medications with different distributions of ADRs. Note again that: $j \geq 1$ and $n \leq N.$

$$\text{Percentage occurence of ADRs within a range of ranks} = \sum_{r=j}^{n} f(r), \qquad 1 \leq j \leq n. \tag{17}$$

Equivalently:

$$\text{Percentage occurence of ADRs within a range of ranks} \approx \sum_{r=j}^{n}\frac{100}{(r+b)^a}, \qquad 1 \leq j \leq n. \tag{18}$$

Examination of Figure 1 demonstrates that the medication-specific values for $f(r)$ tend to coalesce at approximately $r = 10$. Therefore, the use of both the sum and average derivative, based upon the first ten ranks, is particularly beneficial when comparing the various medications' different ADR distributions (See: Results).

 doi:10.20944/preprints202604.0494.v1

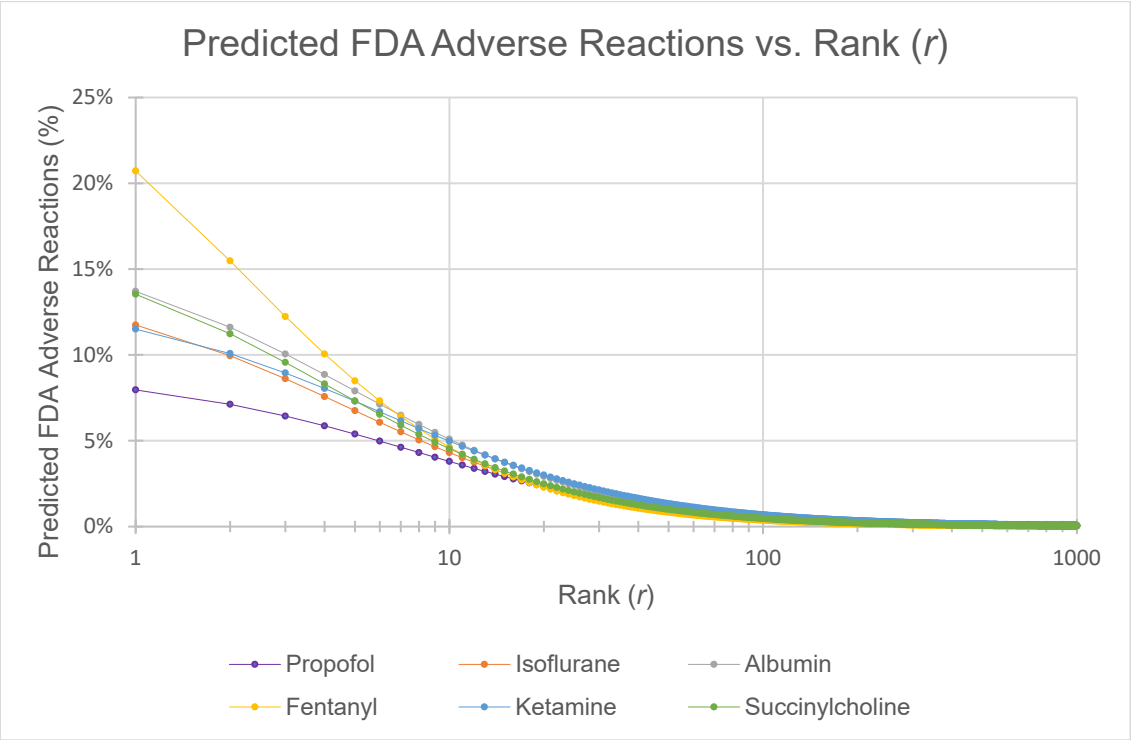


Figure 1. A comparison of the ZM law for each of the six medications examined. Note that these graphs were generated using their medication-specific predicted values based upon curve fitting coefficients a and b.

A natural logarithmic transformation has also been utilized for additional statistical analysis:

$$\ln\left(\frac{f(r)}{100}\right)=\ln\left(\frac{1}{(r+b)^{a}}\right). \tag{19}$$

Which is equal to:

$$\ln\left(\frac{f(r)}{100}\right)=-a\cdot\ln(r+b). \tag{20}$$

Inspection of the above equations demonstrates that the natural logarithm function operates on the unitless decimal form of both the reported data and the predicted values. This transformation also results in an approximate linearization of the predicted vs. the reported values.

Thus, with increasing rank, $r \gg b$ therefore: $\ln(r+b) \approx \ln(r)$. Consequently, the natural logarithm of a reported ADR and the natural logarithm of its rank will become approximately proportional as r increases:

$$\ln\left(\frac{f(r)}{100}\right)\approx -a\cdot\ln(r), \qquad r \gg b. \tag{21}$$

Moreover, the ZM law can also be represented using an exponential equation. This helps to explain its fundamental "exponential like" mathematical properties:

$$f(r)=100\cdot e^{-a\cdot\ln(r+b)}. \tag{22}$$

Materials and Methods

The raw data for this study were downloaded directly from the US FDA website [10]. Mathematical and statistical analyses were done using Microsoft Excel, XLSTAT, and PTC Mathcad Prime 10.0. All data were obtained and uploaded anonymously by the US FDA. Therefore, Institutional Review Board (IRB) approval was deemed unnecessary.

For each medication, all FDA adverse events and associated ranks were entered into a spreadsheet (MS Excel). Medication-specific values for coefficients a and b were determined using

 doi:10.20944/preprints202604.0494.v1

nonlinear curve fitting (XLStat). Note that the raw data were initially acquired in decimal form and subsequently multiplied by 100 to obtain an equivalent percentage.

Six frequently used generic medications and their reported ADRs were assessed: propofol, albumin, isoflurane, fentanyl, succinylcholine, and ketamine. It should be noted that these medications are commonly utilized in hospital-based anesthesia practice. However, ADRs resulting from the illicit use of propofol, fentanyl, and ketamine are also collectively documented within the AEMS. Thus, no explicit identification or differentiation of either legitimate or illicit use and associated ADRs can be determined from the analysis of the AEMS database. It should be noted that drug overdoses are also considered ADRs and are included in the AEMS database [11].

Appendix 1 documents the ten most prevalent adverse effects and their corresponding ranks associated with each of the six medications. Note that each ADR is expressed as a percentage for a given rank.

Results

Figure 1 demonstrates the predicted adverse reactions for all six medications utilizing the ZM law. Table 1 and Figure 2 show the medication-specific values for both a and b. Figure 2 also documents a slight negative correlation between coefficients a and b.

Furthermore, Table 1 and Appendix 2 illustrate the associated root mean square error as a function of percentage, RMSE(%), for each medication. Note that Appendix 3 shows the derivation of RMSE(%) based upon RSME(decimal format).

In addition, a natural logarithmic transformation was also calculated for each value of $f(r)$ and its associated predicted value. This generated medication-specific approximate linear models and associated coefficients of determination, R^2 (See: Table 1 and Appendix 2). This transformation process also yielded correlations which were statistically significant $(p < 0.0001)$ for each medication.

Table 1 and Figure 3 also document the dimensionless ratio: $a \cdot b^{-1}$. Note that this ratio was found to negatively correlate with the reported average derivative and positively correlate with the sum of the ADRs. Note that the average derivative and sum are calculated using only the ADRs associated with the first ten ranks (See: Figures 4 and 5).

Inspection of Table 1 and Figure 3 demonstrate that coefficient a tends to have a near-uniform approximate value of slightly greater than 1 whereas coefficient b has a range of approximately 2.67 to 8.59.

In addition, Figures 6 and 7 document the correlation between reported and predicted average derivatives and sums over the first ten ranks. Figure 8 shows that greater negative slopes correlate with greater sums.

Lastly, Figure 9 compares RMSE(%) to R^2. Note that RMSE(%) is determined using the non-linear ZM model (Eqns. 1 and 8) whereas R^2 is obtained using the natural logarithmic model (Eqns. 19 and 20).

 doi:10.20944/preprints202604.0494.v1

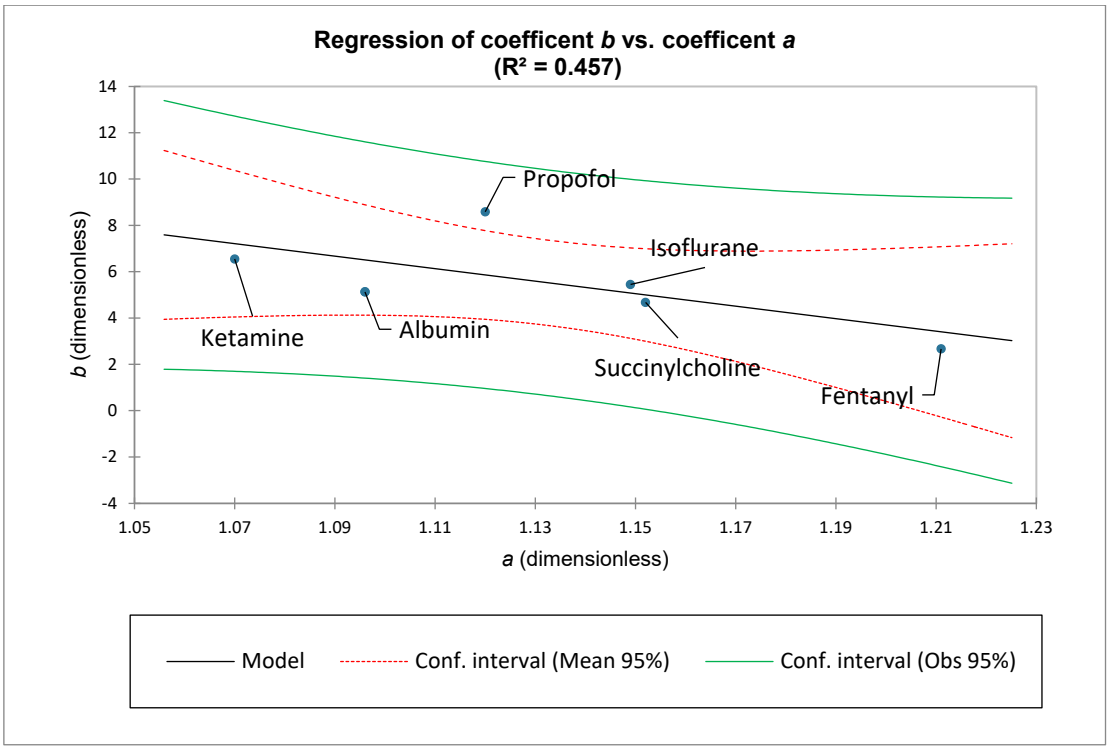


Figure 2. Coefficients *a* and *b* are illustrated for each medication. Note that some negative correlation is evident.

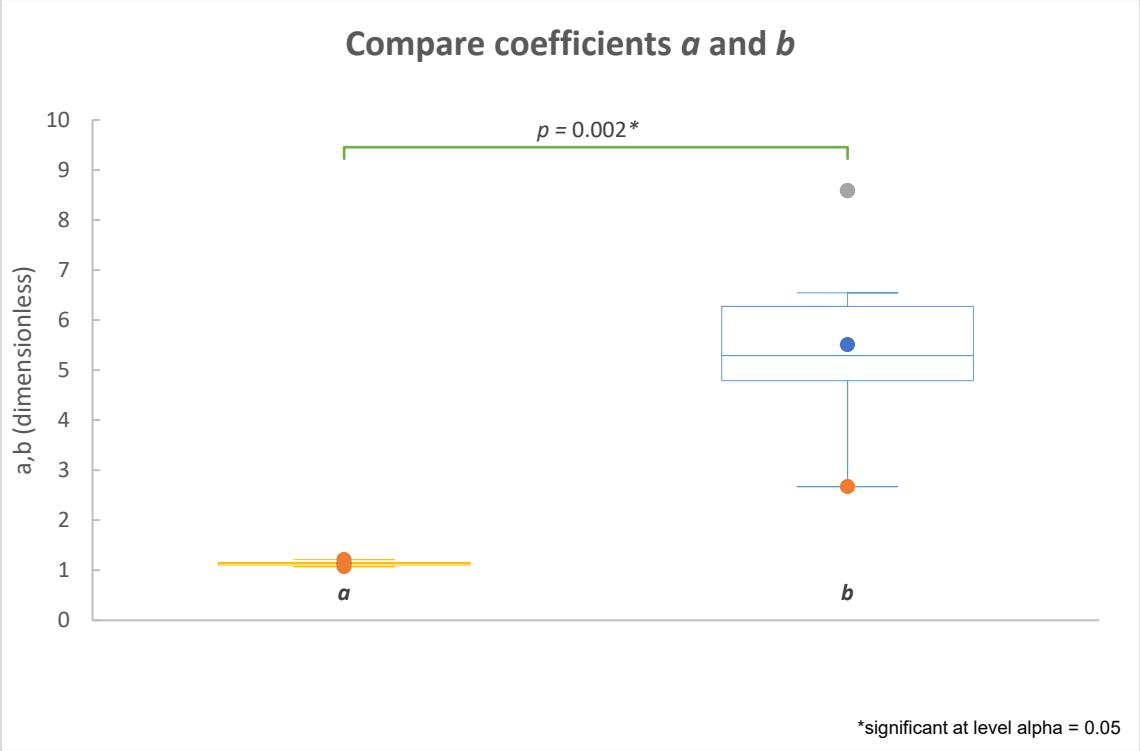


Figure 3. A statistical comparison of coefficients *a* and *b*.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 April 2026 doi:10.20944/preprints202604.0494.v1

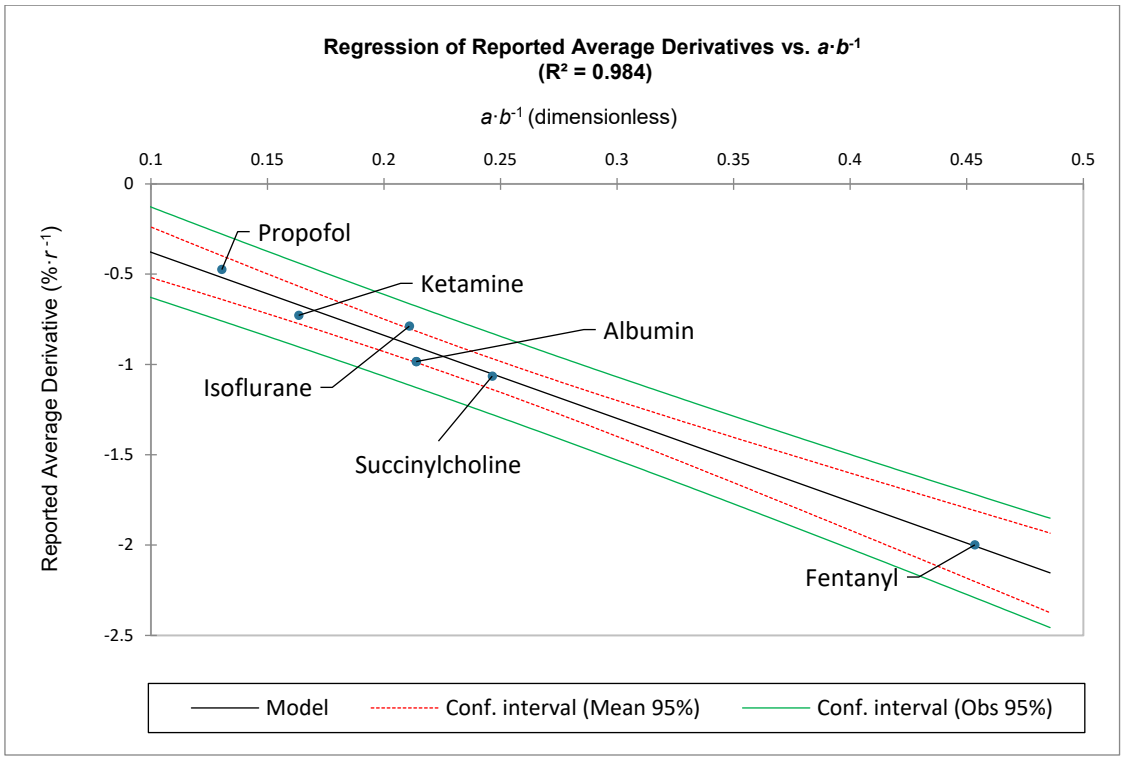


Figure 4. The ratio of $a \cdot b^{-1}$ negatively correlates with the the reported average derivative for each medication.

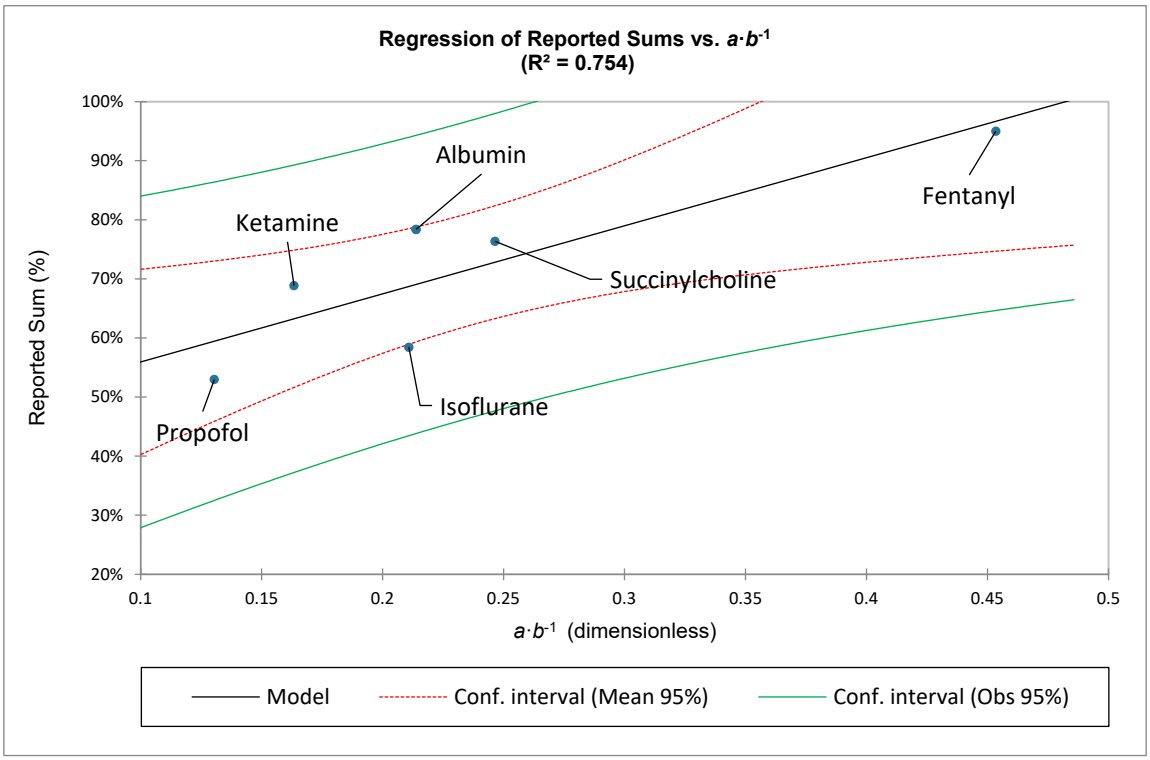


Figure 5. The ratio of $a \cdot b^{-1}$ correlates with the reported sum for each medication.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 April 2026 doi:10.20944/preprints202604.0494.v1

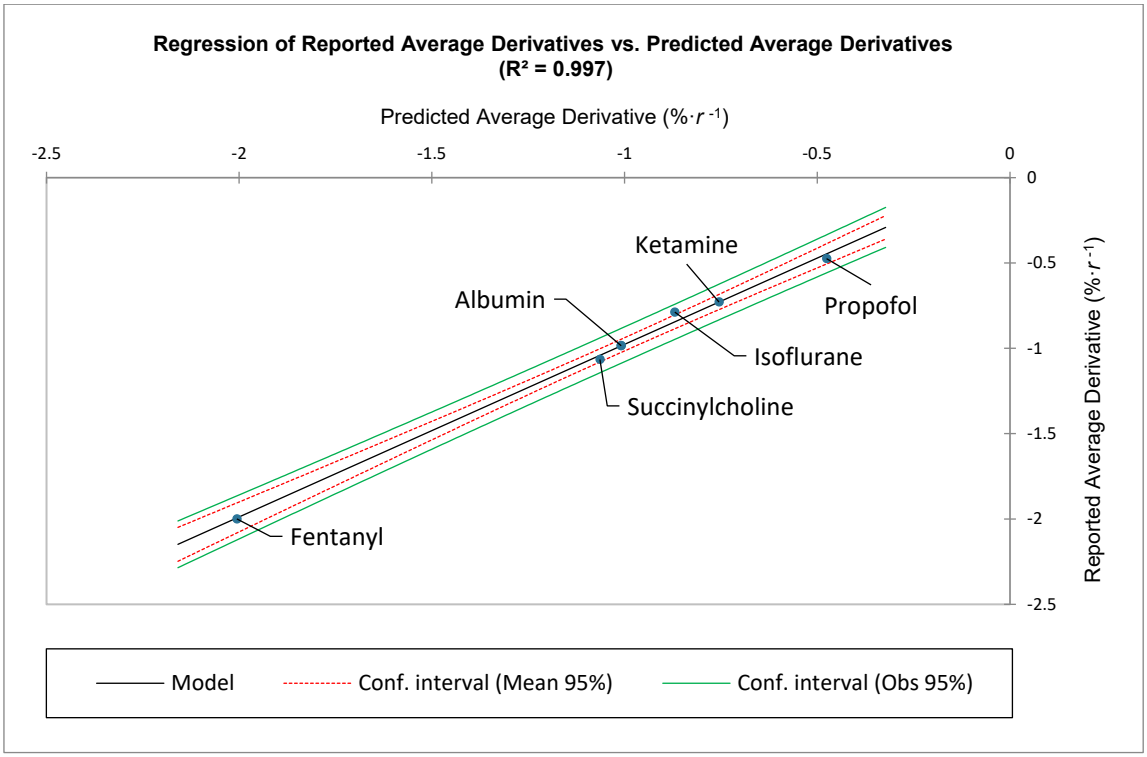


Figure 6. Each reported average derivative correlates with its predicted average derivative.

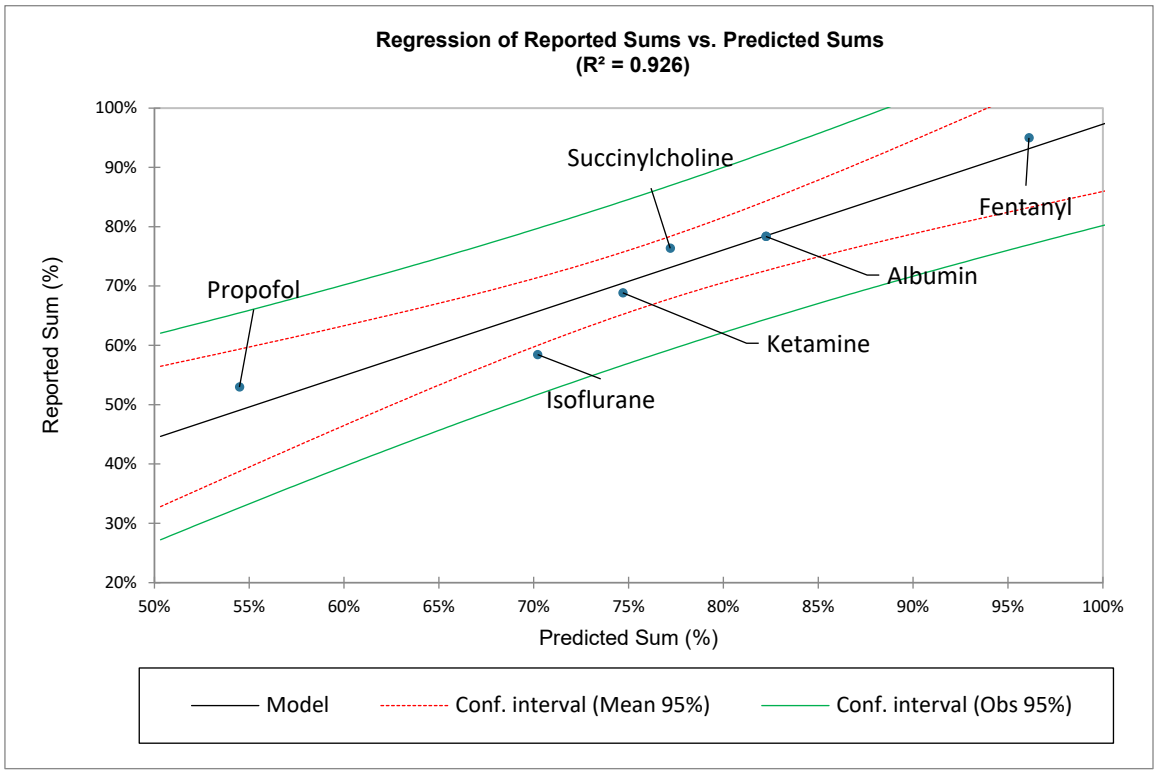


Figure 7. Correlation of reported and predicted sums.

 doi:10.20944/preprints202604.0494.v1

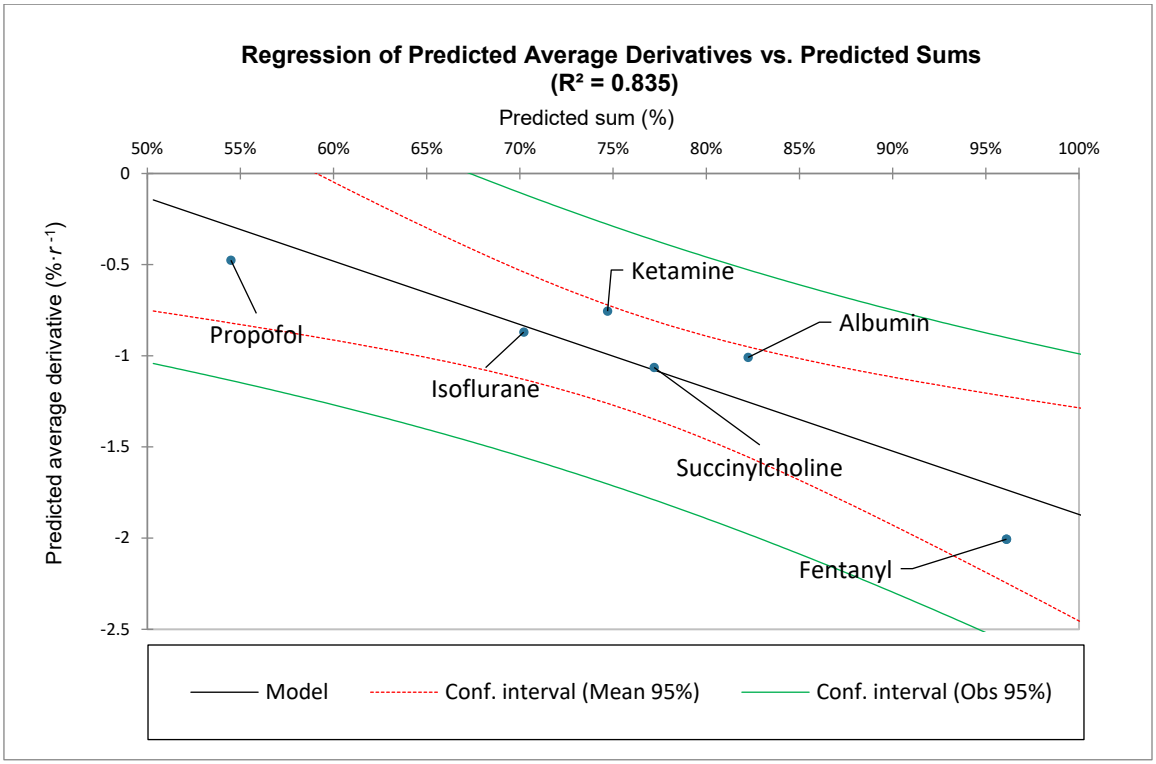


Figure 8. Predicted average derivatives correlate with predicted sums.

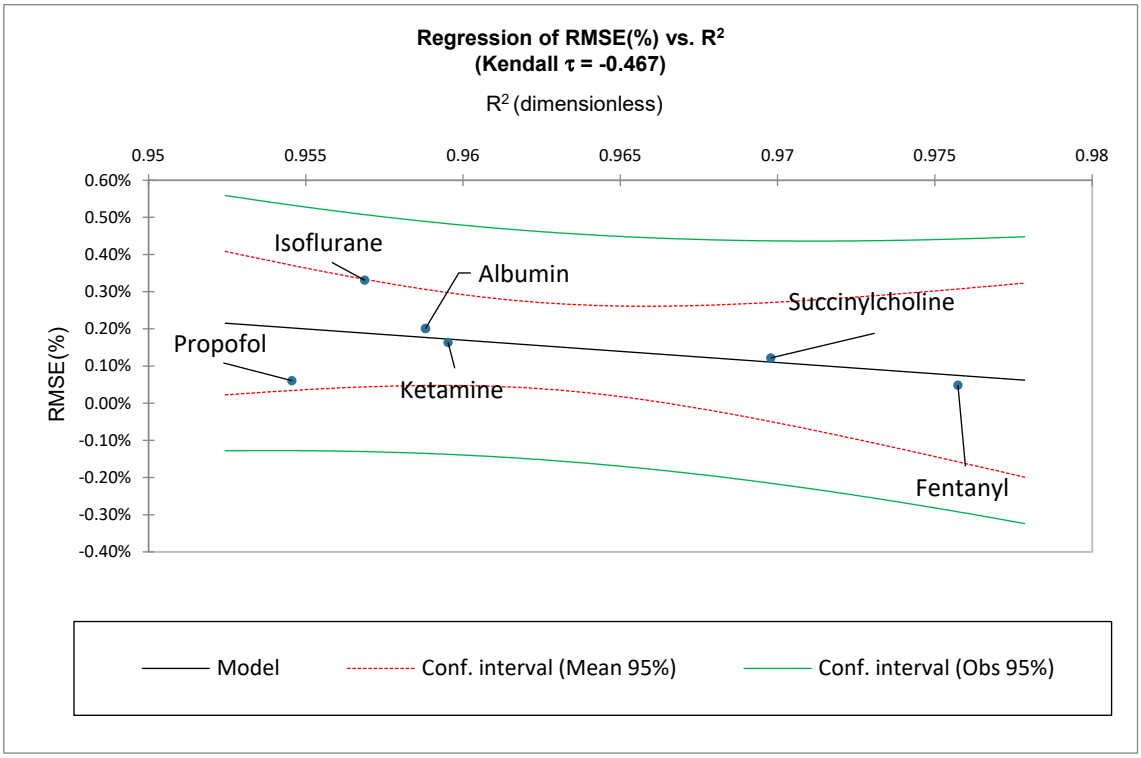


Figure 9. Correlation of medication-specific root mean square error using percentage, RMSE(%), and its respective coefficient of determination, R^2. Note that RMSE(%) is calculated using the non-linear ZM model whereas R^2 is calculated using the natural logarithmic model.

Table 1. Results of nonlinear and natural logarithmic curve fitting. * ($p < 0.0001$).

	Coefficients		Ratio	Nonlinear Model	Natural Logarithmic Model*
Medication	***a***	***b***	$a \cdot b^{-1}$	**RMSE(%)**	**R^2**
Fentanyl	1.211	2.670	0.453	0.05	0.976

 doi:10.20944/preprints202604.0494.v1

Albumin	1.096	5.128	0.214	0.20	0.959
Succinylcholine	1.152	4.674	0.246	0.12	0.970
Ketamine	1.070	6.547	0.163	0.02	0.960
Isoflurane	1.149	5.450	0.211	0.33	0.957
Propofol	1.120	8.586	0.130	0.06	0.955

Table 2. Reported and predicted sum and average derivatives. Note that these are based upon the first ten ranks.

	Based upon the first ten ranks			
Medication	**Reported Sum (%)**	**Predicted Sum (%)**	**Reported Average Derivative (%·r^{-1})**	**Predicted Average Derivative (%·r^{-1})**
Fentanyl	95.650	96.106	-1.999	-2.006
Albumin	78.370	82.240	-0.984	-1.009
Succinylcholine	76.370	77.197	-1.065	-1.064
Ketamine	68.864	74.696	-0.728	-0.755
Isoflurane	58.434	70.202	-0.788	-0.870
Propofol	52.980	54.485	-0.474	-0.476

Discussion

This paper is a preliminary demonstration of the applicability of the ZM law to mathematically model US FDA ADRs. This may provide a reasonable "first step" for comparing overall clinical complication rates associated with various medications.

As demonstrated, medications which have a distribution of their ADRs with a less steep or shallower average slope as well as a reduced sum may be potentially safer than those with steeper slopes and greater sums. Note that these values are determined over the initial range of each medication's first ten ranks.

Thus, a flatter distribution of ADRs may be associated with potentially safer medications as a result of a more even dispersal of medication-based complications. Further research would be necessary to investigate the potential clinical utility of this application of the ZM law.

Moreover, it should be noted that the dimensionless ratio $a \cdot b^{-1}$, was also found to correlate with both the average derivative and the sum of adverse events associated with the first ten ranks. Specifically, $a \cdot b^{-1}$, negatively correlated with the average slope whereas $a \cdot b^{-1}$ positively correlated with the sum. As stated, both correlations are based upon the adverse events associated with the first ten ranks of the ADRs of each medication.

Therefore, a lower value of $a \cdot b^{-1}$ results in a flatter or a more even distribution of ADRs. Thus, this ratio may function as a useful index of comparison when evaluating ADRs from different medications. Lastly, inspection of Table 1 and Figure 3 demonstrates that changes $a \cdot b^{-1}$ are mainly due to changes in b.

ADRs are a leading cause of both morbidity and mortality among hospitalized patients. It has been estimated that approximately 10.9% of all inpatients will experience an ADR with 2.1% experiencing a serious event and 0.19% having a fatal event [12]. In addition, inpatients who were older, female, and who took more medications had a greater risk of having an ADR. Moreover, a greater length of stay was also associated with an increased likelihood of experiencing an ADR among inpatients [13].

ADRs which occur outside of a hospital and lead to hospitalizations have also been examined. Specifically, the overall incidence of ADRs in the primary care setting is estimated as 8.32 percent with many of these being considered preventable [14].

While the ZM law appears to be readily applicable to modeling FDA-reported ADRs as a function of rank, it should be noted that the quantification of the economic effects of ADRs is a more difficult process which is often incomplete or inaccurate. However, there have been numerous efforts to perform this type of analysis as well as recommendations for more efficacious investigations [15].

 doi:10.20944/preprints202604.0494.v1

In the early 1990s, a team from LDS Hospital evaluated three years of data at their institution and concluded that an ADR increased length of stay by an average of 1.91 days at an increase in cost for care by an average of $2,262 [16].

In the same decade, a team led by Bates et al. produced similar results, concluding that an ADR increased length of stay by 2.2 days and increased cost of care by $3,244 [17].

It is important to note that there is likely great discrepancy between institutions, patient populations, and geographic locations regarding ADR-related economic costs. Specifically, ADRs were found to extend length of stay by an average of 14.1 days in a pediatric patient population in Japan in 2021. This was also associated with an increase in expenditures of $8,258 [18]. However, another team led by Fernandez et al., in a pharmacy setting, estimated a cost of $218 per likely ADR [19].

With the advent of electronic healthcare records, clinicians are frequently made aware of potential ADRs. However, these alerts are commonly overridden. Slight et al. evaluated the cost of ADRs related to inappropriate medication-related alert overrides in a United States inpatient setting. Using data from Brigham and Women's Hospital, they utilized a regression model to estimate that 5.5 million medication alerts may be inappropriately overridden annually in the United States. This estimate could therefore potentially yield 196,600 ADRs, likely costing between $871 million and $1.8 billion total [20].

Conclusion

This paper has preliminarily demonstrated that the ZM law is applicable to the mathematical modeling of US FDA ADRs. This was supported using six commonly utilized hospital-based medications and applying non-linear curve fitting and a natural logarithmic transformation to each medication-specific distribution.

In addition, the ratio of the two coefficients of the ZM law, $a \cdot b^{-1}$, may function as a useful index which summarizes both the flatness and sum of each distribution based upon the ADRs associated with the first ten ranks. Thus, medications with a lower $a \cdot b^{-1}$ ratio may be potentially safer owing to a more even dispersal of their ADRs.

Lastly, the applicability of the ZM law may be beneficial in reducing ADRs in a hospital setting. This could potentially reduce morbidity and mortality as well as healthcare costs. With the widespread utilization of electronic medical records, further examination and implementation of the ZM law may allow for improved medication-based decision making.

Appendix 1

Reported ADRs: Ranks 1 through 10.

Fentanyl			
Category	**Number of Cases**	**Rank**	**Percentage**
Death	13,254	1	20.82
Toxicity To Various Agents	9,703	2	15.24
Wrong Technique In Product Usage Process	7,675	3	12.06
Overdose	6,663	4	10.47
Drug Ineffective	6,378	5	10.02
Product Adhesion Issue	4,566	6	7.17
Product Quality Issue	3,806	7	5.98
Pain	3,708	8	5.82
Therapeutic Product Effect Decreased	2,286	9	3.59
Nausea	2,217	10	3.48
		Total	**94.65%**

Albumin

 doi:10.20944/preprints202604.0494.v1

Category	Number of Cases	Rank	Percentage
Pyrexia	86	1	15.90
Hypotension	63	2	11.65
Dyspnea	55	3	10.17
Urticaria	36	4	6.65
Chills	34	5	6.28
Stevens-Johnson Syndrome	32	6	5.91
Pruritus	31	7	5.73
Tachycardia	31	8	5.73
Toxic Epidermal Necrolysis	29	9	5.36
Dermatitis	27	10	4.99
		Total	**78.37%**

Succinylcholine			
Category	**Number of Cases**	**Rank**	**Percentage**
Hyperthermia Malignant	260	1	12.70
Anaphylactic Shock	238	2	11.62
Drug Ineffective	226	3	11.04
Hypotension	198	4	9.67
Cardiac Arrest	171	5	8.35
Anaphylactic Reaction	100	6	4.88
Bronchospasm	99	7	4.83
Fetal Exposure During Pregnancy	99	8	4.83
Tachycardia	88	9	4.30
Bradycardia	85	10	4.15
		Total	**76.37%**

Ketamine			
Category	**Number of Cases**	**Rank**	**Percentage**
Drug Ineffective	772	1	15.71
Off Label Use	442	2	8.99
Anaphylactic Shock	411	3	8.36
Drug Abuse	366	4	7.45
Hypotension	297	5	6.04
Toxicity To Various Agents	265	6	5.39
Product Use In Unapproved Indication	225	7	4.58
Agitation	224	8	4.56
Hyperhidrosis	203	9	4.13
Hallucination	179	10	3.64
		Total	**68.86%**

Isoflurane			
Category	**Number of Cases**	**Rank**	**Percentage**
Hyperthermia Malignant	430	1	18.60
Hypotension	203	2	8.78
Drug Ineffective	124	3	5.36
Cardiac Arrest	103	4	4.46
Hepatitis	90	5	3.89
Post Procedural Complication	88	6	3.81
Maternal Exposure During Pregnancy	87	7	3.76

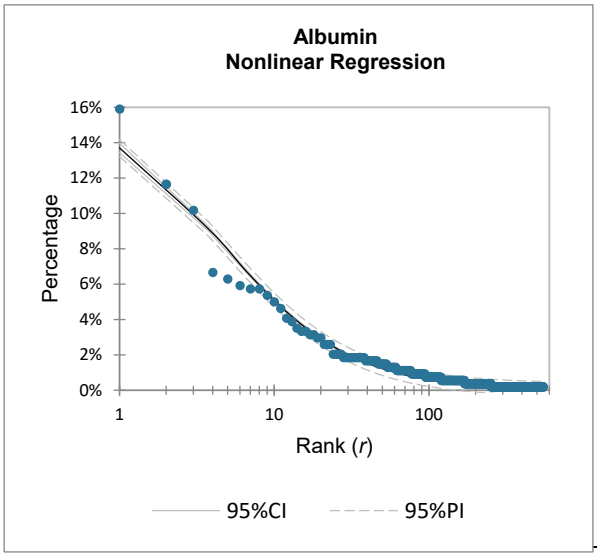
 doi:10.20944/preprints202604.0494.v1

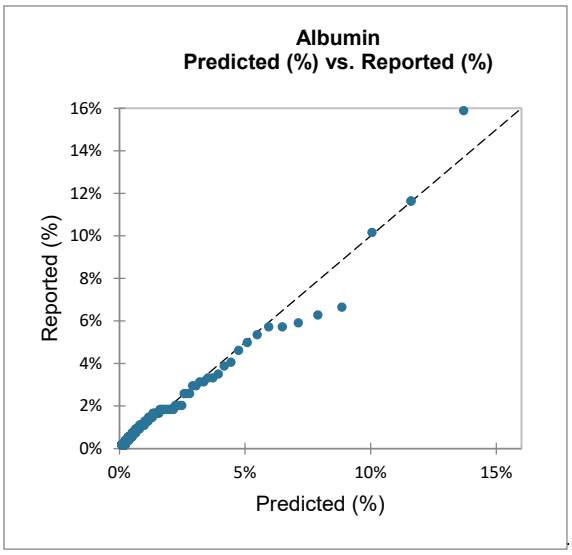
Pyrexia	79	8	3.42
Bradycardia	74	9	3.20
Anesthetic Complication Neurological	73	10	3.16
		Total	**58.43%**

Propofol			
Category	**Number of Cases**	**Rank**	**Percentage**
Hypotension	1,681	1	9.28
Anaphylactic Shock	1,337	2	7.38
Drug Ineffective	1,292	3	7.13
Cardiac Arrest	1,006	4	5.55
Anaphylactic Reaction	945	5	5.22
Drug Interaction	760	6	4.20
Off Label Use	725	7	4.00
Bradycardia	710	8	3.92
Tachycardia	589	9	3.25
Rhabdomyolysis	549	10	3.03
		Total	**52.98%**

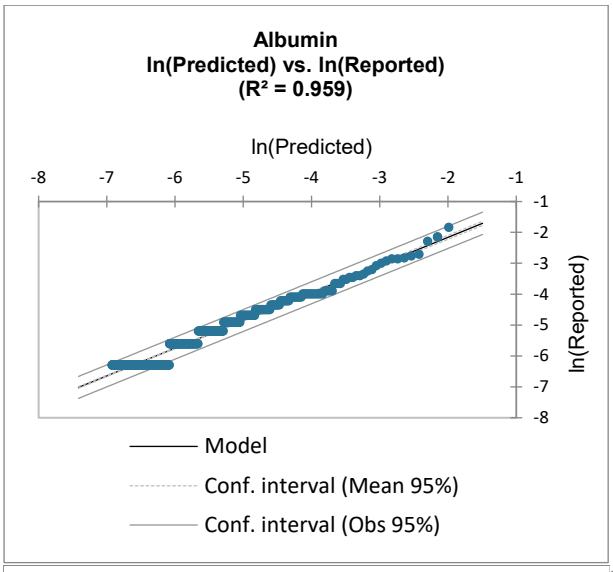
Appendix 2

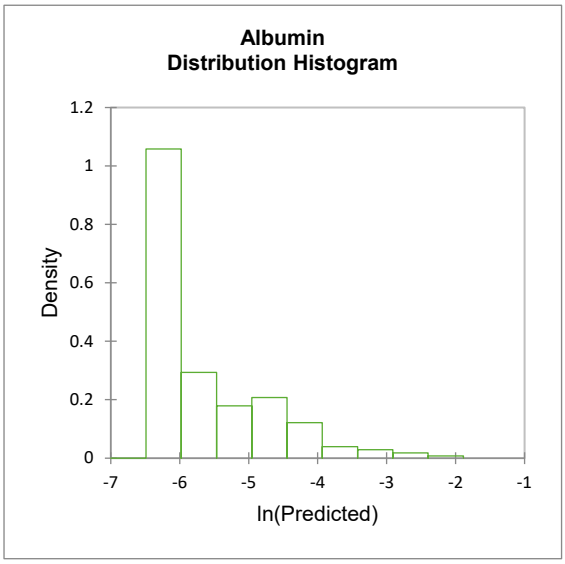
Graphical summaries of US FDA ADRs utilizing nonlinear regression and natural logarithmic (ln) transformations

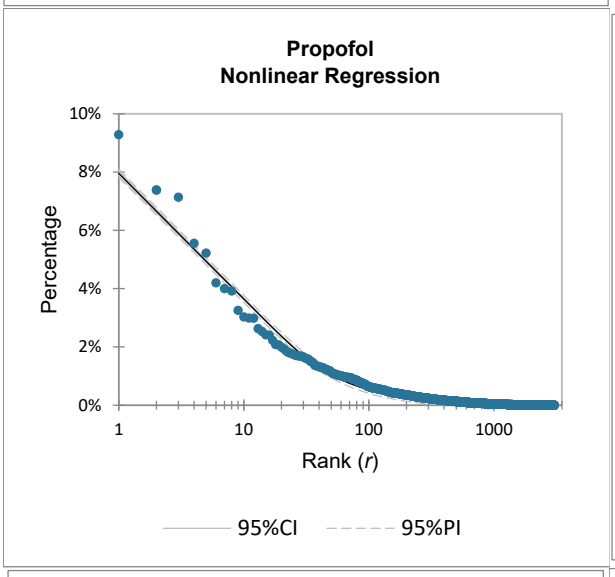


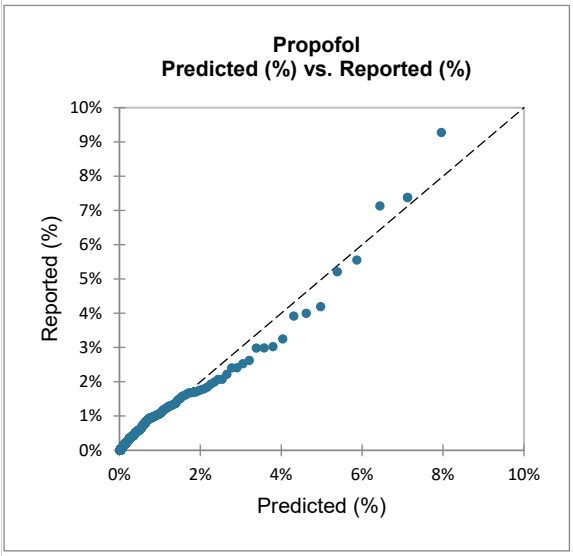


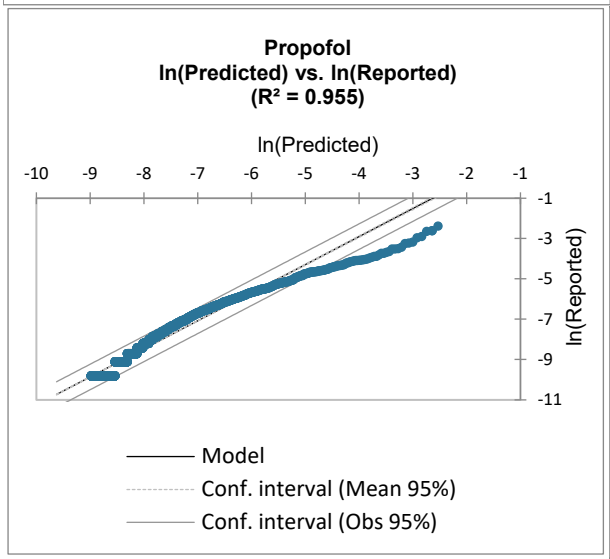
Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 April 2026 doi:10.20944/preprints202604.0494.v1

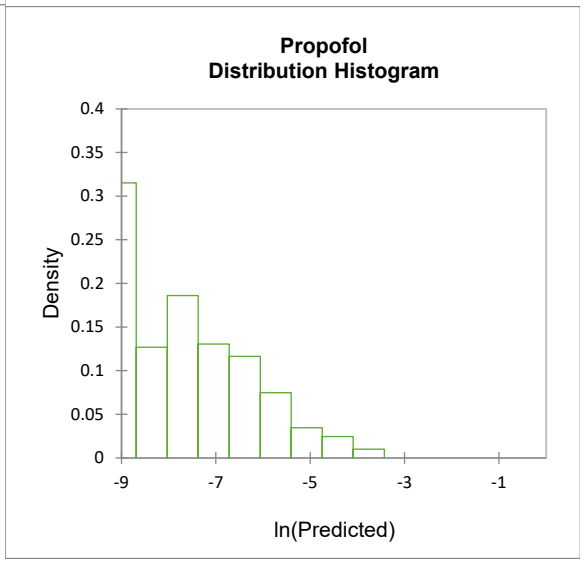




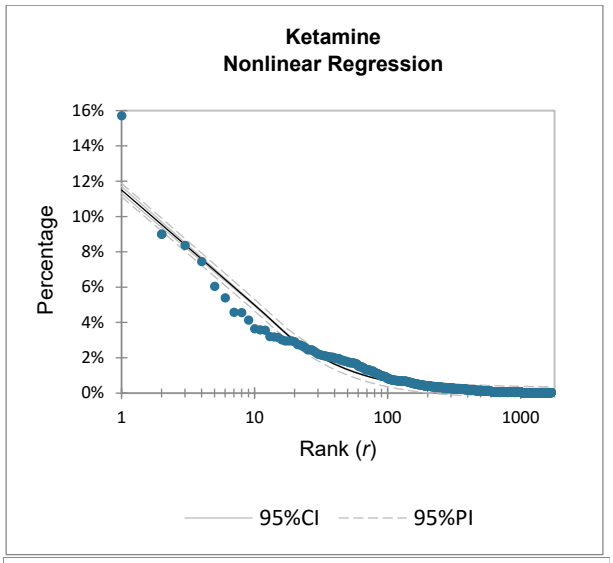


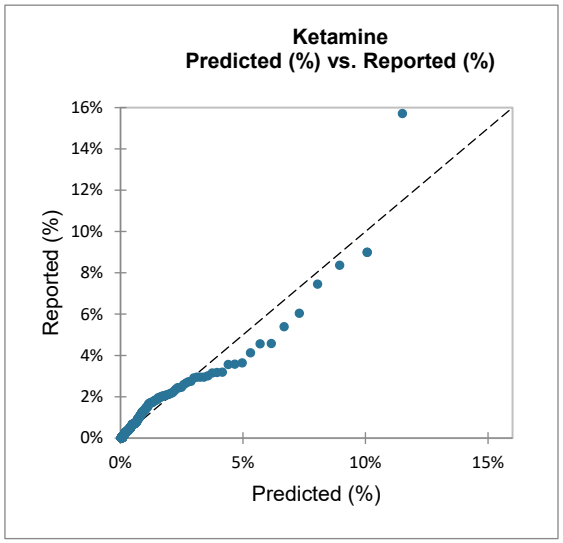


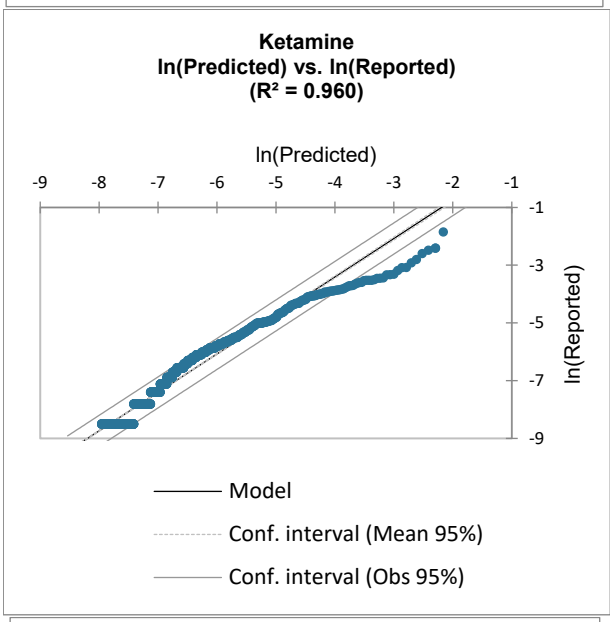


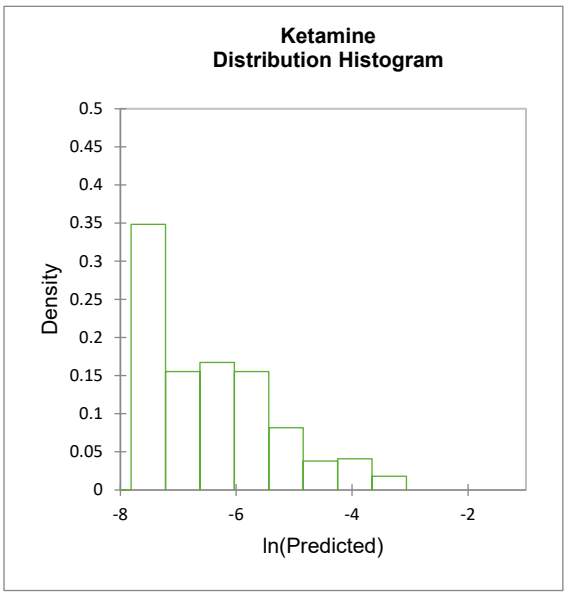


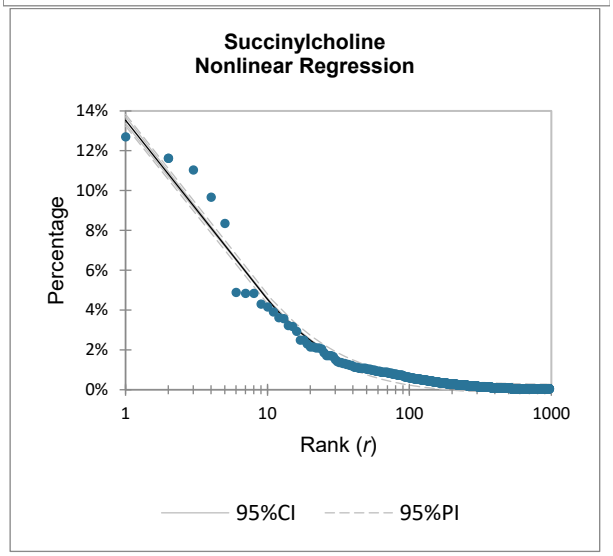
Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 April 2026 doi:10.20944/preprints202604.0494.v1

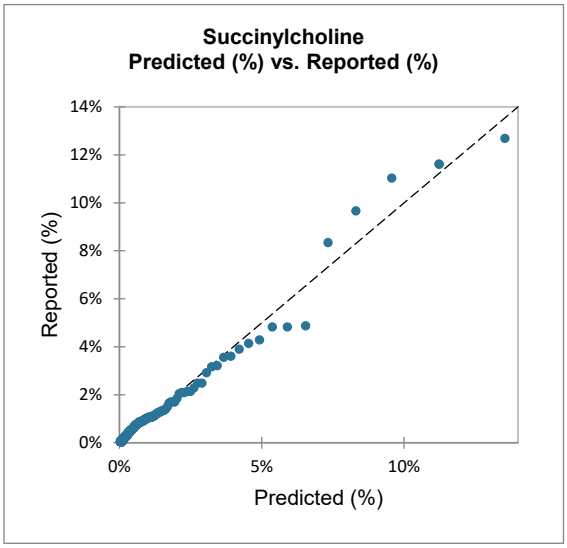




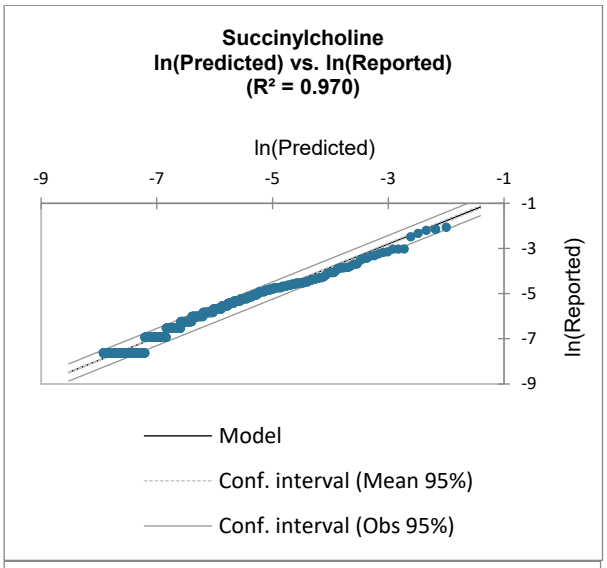


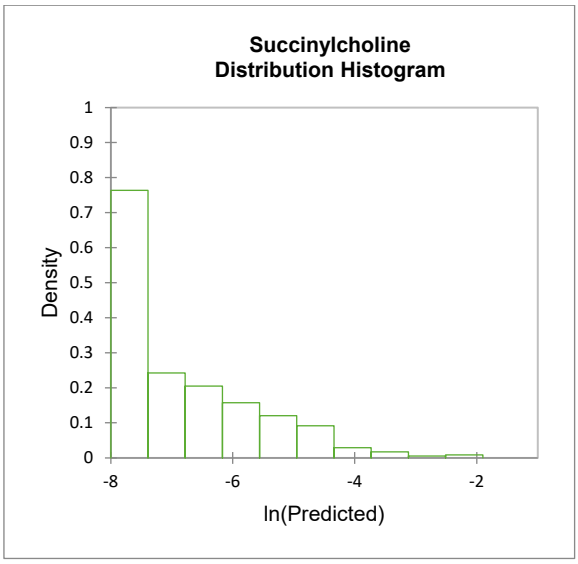


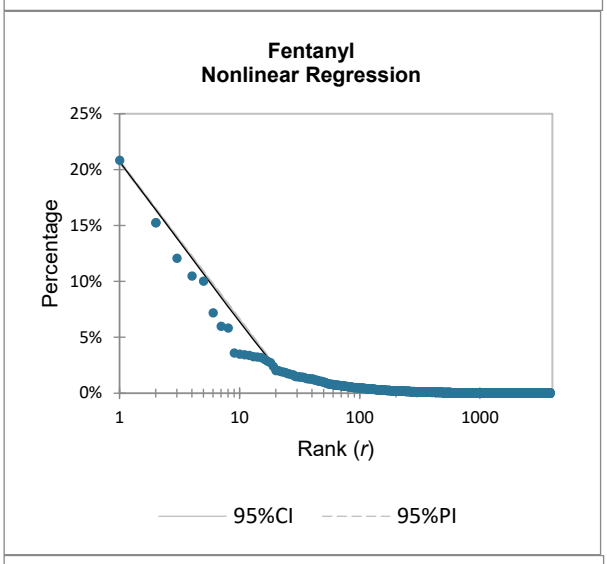


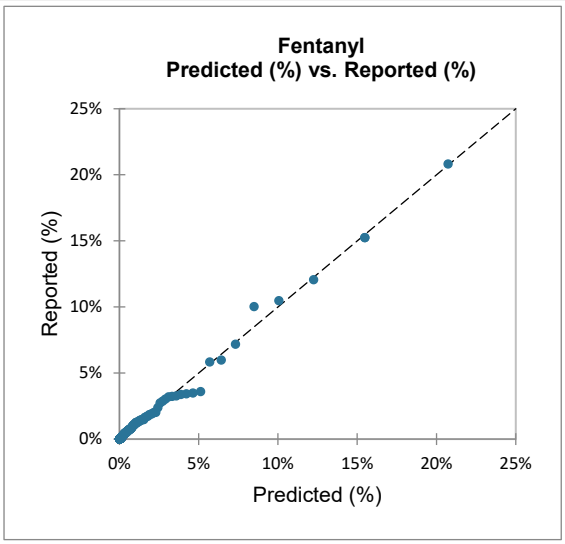


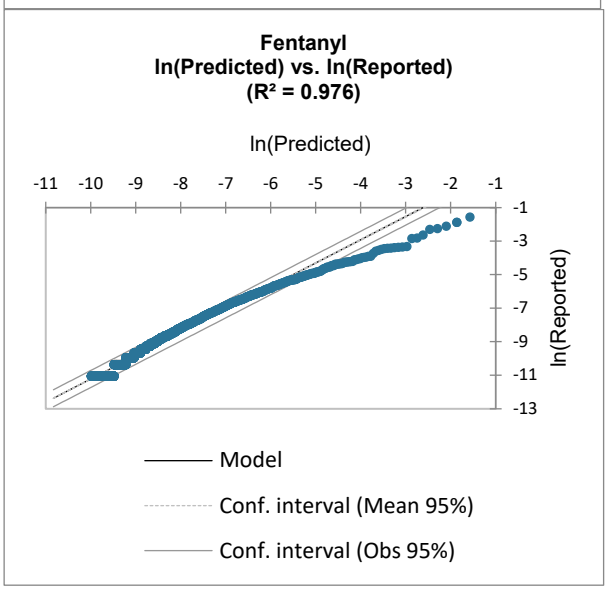
Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 April 2026 doi:10.20944/preprints202604.0494.v1

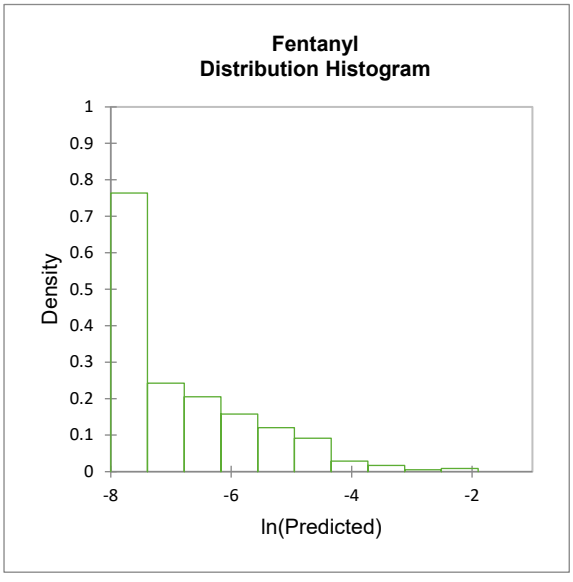




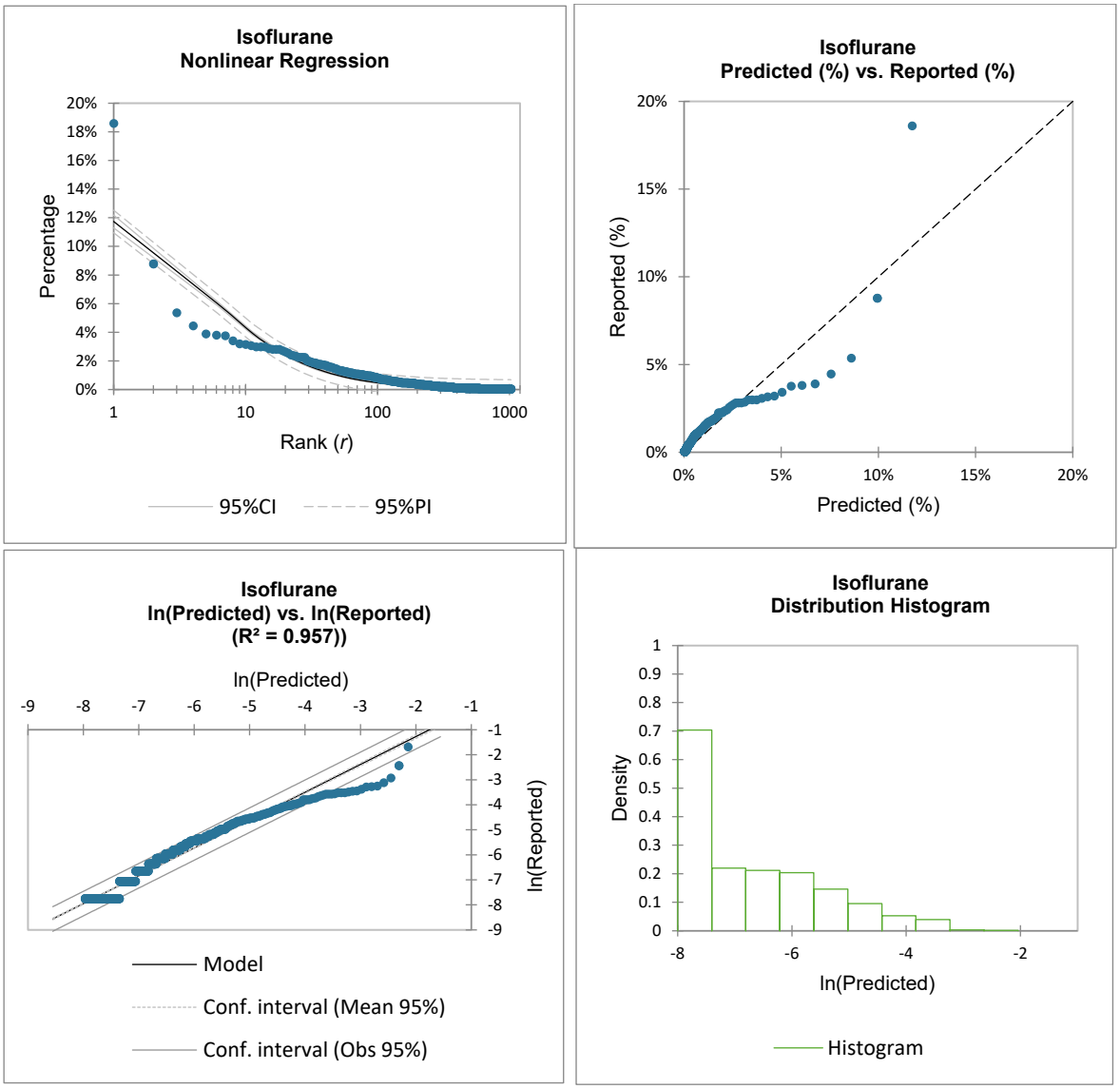








 doi:10.20944/preprints202604.0494.v1



Appendix 3

The derivation of RSME(%) using RSME(decimal form)

Mean square error, MSE(decimal form), is defined as:

$$\text{MSE(decimal form)} = \sum_{j=1}^{N} \frac{\left(X_{r_j} - X_{p_j}\right)^2}{N}. \tag{A1}$$

Where X_{r_j} and X_{p_j} are the dimensionless rank-specific reported and predicted ADRs in decimal form, respectively.

In an analogous manner, MSE(%) is defined using the reported and predicted values expressed as a function of percentages:

$$\text{MSE}(\%) = \sum_{j=1}^{N} \frac{\left(AX_{r_j} - AX_{p_j}\right)^2}{N}, \qquad A = 100. \tag{A2}$$

Expanding the numerator of the above equation yields:

$$\left(AX_{r_j} - AX_{p_j}\right)^2 = A^2X_{r_j}^2 - 2A^2X_{r_j}X_{p_j} + A^2X_{p_j}^2 = A^2\left(X_{r_j} - X_{p_j}\right)^2. \tag{A3}$$

Therefore:

$$\text{MSE}(\%) = A^2 \sum_{j=1}^{N} \frac{\left(X_{r_j} - X_{p_j}\right)^2}{N} = A^2 \cdot \text{MSE(decimal form)}. \tag{A4}$$

Note that MSE(%) has units of "percent squared." Whereas root mean square error, RMSE(decimal form), is determined by calculating the square root of the dimensionless MSE(decimal form):

$$\text{RMSE(decimal form)} = \sqrt{\sum_{j=1}^{N} \frac{\left(X_{r_j} - X_{p_j}\right)^2}{N}} = \sqrt{\text{MSE(decimal form)}}. \tag{A5}$$

 doi:10.20944/preprints202604.0494.v1

RMSE(%) is subsequently defined in an analogous manner:

$$\mathrm{RMSE}(\%) = \sqrt{\mathrm{MSE}(\%)} = \mathrm{A}\cdot\sqrt{\sum_{j=1}^{N}\frac{\left(X_{r_j} - X_{p_j}\right)^2}{N}} = A\cdot\sqrt{\mathrm{MSE(decimal\ form)}}. \tag{A6}$$

Thus:

$$\mathrm{RMSE}(\%) = 100\cdot\mathrm{RMSE(decimal\ Form)} = 100\cdot\sqrt{\mathrm{MSE(decimal\ form)}}. \tag{A7}$$

Note that RMSE(%) has units of percentage

References

1. Miguel A, Azevedo LF, Araújo M, Pereira AC. Frequency of adverse drug reactions in hospitalized patients: a systematic review and meta-analysis. Pharmacoepidemiology and drug safety. 2012 Nov;21(11):1139-54.
2. Bouvy JC, De Bruin ML, Koopmanschap MA. Epidemiology of adverse drug reactions in Europe: a review of recent observational studies. Drug safety. 2015 May;38(5):437-53.
3. Piantadosi ST. Zipf's word frequency law in natural language: A critical review and future directions. Psychonomic bulletin & review. 2014 Oct;21(5):1112-30.
4. Dalton DB, Dweck R, Elinon M, Davidson J. Modeling Insurance Frequency with the Zipf-Mandelbrot Distribution. InCAS E-Forum 2022 Sep 26.
5. Silagadze ZK. Citations and the Zipf-Mandelbrot's law. arXiv preprint physics/9901035. 1999 Jan 21.
6. Adamic LA, Huberman BA. Zipf's law and the Internet. Glottometrics. 2002 Jun;3(1):143-50.
7. Gabaix X. Power laws in economics and finance. Annu. Rev. Econ.. 2009 Sep;1(1):255-94.
8. Hackmann A, Klarl T. The evolution of Zipf's Law for US cities. Papers in Regional Science. 2020 Jun 1;99(3):841-52.
9. Newman ME. Power laws, Pareto distributions and Zipf's law. Contemporary physics. 2005 Sep 1;46(5):323-51.
10. https://www.fda.gov/drugs/fda-adverse-event-monitoring-system-aems/fda-adverse-event-monitoring-system-aems-public-dashboard
11. Ni J, Tang X, Chen L. Medication overdose data analysis: a review of medication error reports in the FDA adverse event reporting system (FAERS). BMC pharmacology and toxicology. 2023 Aug 4;24(1):41.
12. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. Jama. 1998 Apr 15;279(15):1200-5.
13. Davies EC, Green CF, Taylor S, Williamson PR, Mottram DR, Pirmohamed M. Adverse drug reactions in hospital in-patients: a prospective analysis of 3695 patient-episodes. PLoS one. 2009 Feb 11;4(2):e4439.
14. Insani WN, Whittlesea C, Alwafi H, Man KK, Chapman S, Wei L. Prevalence of adverse drug reactions in the primary care setting: A systematic review and meta-analysis. PLoS One. 2021 May 26;16(5):e0252161.
15. Ghabri S, Dawoud D, Drummond M. Methods for including adverse events in economic evaluations: suggestions for improvement. Value in Health. 2024 Jul 1;27(7):936-42.
16. Classen DC, Pestotnik SL, Evans RS, Lloyd JF, Burke JP. Adverse drug events in hospitalized patients: excess length of stay, extra costs, and attributable mortality. Jama. 1997 Jan 22;277(4):301-6.
17. Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, Small SD, Sweitzer BJ, Leape LL. The costs of adverse drug events in hospitalized patients. Jama. 1997 Jan 22;277(4):307-11.
18. Iwasaki H, Sakuma M, Ida H, Morimoto T. The burden of preventable adverse drug events on hospital stay and healthcare costs in Japanese pediatric inpatients: the JADE study. Clinical Medicine Insights: Pediatrics. 2021 Feb;15:1179556521995833.

 doi:10.20944/preprints202604.0494.v1

19. Fernández EV, Warriner CL, David T, Gordon E, Twigg G, Carroll NV. Potential cost savings by prevention of adverse drug events with a novel medication review program. Journal of the American Pharmacists Association. 2020 May 1;60(3):462-9.
20. Slight SP, Seger DL, Franz C, Wong A, Bates DW. The national cost of adverse drug events resulting from inappropriate medication-related alert overrides in the United States. Journal of the American Medical Informatics Association. 2018 Sep;25(9):1183-8.