2009 Ninth IEEE International Conference on Bioinformatics and Bioengineering

Analysis of Autism Prevalence and Neurotoxins Using Combinatorial Fusion and Association Rule Mining

Christina Schweikert[1], Yanjun Li[1], David Dayya[2], David Yens[3], Martin Torrents[2], D. Frank Hsu[1]
[1] Laboratory for Informatics and Data Mining, Dept. of Computer & Information Science,
Fordham University, Bronx, NY 10458, USA
[2] St. Barnabas Hospital, Bronx, NY 10457, USA
[3] New York College of Osteopathic Medicine,
New York Institute of Technology, Old Westbury, NY 11568, USA

Abstract - **The increase in autism prevalence has been the motivation for much research which has produced various theories for its causation. Genetic and environmental factors have been investigated. An area of focus is the affect of exposure to neurotoxins, such as mercury and lead, during critical stages in a child's early development. In this study we apply Combinatorial Fusion Analysis (CFA) and Association Rule Mining (ARM) to autism prevalence, mercury, and lead data to generate hypotheses and explore possible associations.**

Keywords- Combinatorial Fusion Analysis (CFA), Rank-Score Characteristic (RSC) graph, Multiple Scoring Systems, Information Fusion, Association Rule Mining, Data Mining, autism, neurotoxins, lead, mercury

I. INTRODUCTION AND BACKGROUND

The rise of autism rates in the United States is cause for great concern among the medical community and families of individuals suffering with autism and autism spectrum disorders. The Centers for Disease Control and Prevention's Autism and Developmental Disabilities Monitoring Network indicates that in 2007 about 1 in 150 8-year-old children in multiple areas of the United States has an autism spectrum disorder (ASD) and it is estimated that up to 560,000 individuals between birth and age 21 have an ASD. Though it is difficult to obtain an exact number of autism cases, it is commonly acknowledged that autism prevalence has been increasing over the past ten years [25].

Research has shown that autism may have a strong genetic component and has identified genes that may be linked to autism [13, 17, 20]. There are indications that some autism spectrum disorders may be attributed to a combination of certain genetic susceptibilities, such as reduced ability to excrete mercury, and exposure to mercury at critical developmental stages [8]. In a recent study, it has been demonstrated that, for some autistic children, a genetic predisposition may increase vulnerability to lead toxicity during pre-natal and post-natal neurodevelopment [18]. Several studies have contributed to the evidence that environmental toxicity may play a role in autism [15], and have more specifically investigated the role of mercury to autism [3, 4, 7].

The following two regional ecological studies, though limited in scope, have contributed to the debate on the affect of environmental factors on autism rates. An ecological study on the association between environmentally released mercury and autism rates in Texas purports that for each 1000 lb of environmentally released mercury, there is a 61% increase in the rate of autism [16]. A study of the relationship between autism spectrum disorders and the distribution of hazardous air pollutants (HAP) in the San Francisco bay area suggests that areas with higher ambient levels of HAP, metals and chlorinated solvents in particular, during the pre-natal period or early childhood may be associated with a moderately increased risk of autism [21].

In this study, we apply Combinatorial Fusion Analysis (CFA) and Association Rule Mining (ARM) to autism prevalence and neurotoxin data to analyze relationships and discover interesting trends. In the following sections we discuss the datasets, methods (CFA and ARM), and conclusions of our study.

II. DATASETS

Autism prevalence rates are collected under federal law requiring the reporting of autism for all children in public schools under the Individuals with Disabilities and Education Act (IDEA) [23]. Live birth data used to calculate autism prevalence was retrieved from the CDC National Center for Health Statistics [24]. In this study, the autism cases reported in the 3-5 age groups are used. The prevalence per thousand is calculated as number of cases in year t divided by the number of live births in the year t-4 multiplied by one thousand to account for the prenatal period as well as the three postnatal years.

The National Atmospheric and Deposition Program Mercury Deposition Network (NADP-MDN) dataset represents wet deposition and precipitation levels of mercury concentration in the ambient environment [26]. The Environmental Protection Agency's Air Quality Systems (EPA-AQS) dataset contains concentration levels of mercury and lead in air particulate matter. For this study, the particulate matter 2.5 micron was chosen since this size is more likely to persist longer in the atmosphere, potentially increasing the interval of exposure and the amount of environmental deposition and contamination. The 2.5 micron size is also known to migrate further into the respiratory system, leading to more systemic absorption than larger sized particles [27, 28].

E-mail addresses: cschweikert@cis.fordham.edu (C. Schweikert), yli@cis.fordham.edu (Y. Li), david_dayya@stbarnabas-ny.org (D. Dayya), dyens@nyit.edu (D. Yens), marinotorrents@hotmail.com (M. Torrents), hsu@trill.cis.fordham.edu (D.F. Hsu)

978-0-7695-3656-9/09 $25.00 © 2009 IEEE
DOI 10.1109/BIBE.2009.69

III. COMBINATORIAL FUSION ANALYSIS

A. *Multiple Scoring Systems*

CFA is a data-driven method for analyzing the combination of multiple scoring systems and has been widely applied to various domains such as information retrieval, pattern recognition, target tracking, protein structure prediction, drug design and discovery, virtual screening, and biomedical informatics [10, 11, 12, 22]. CFA can be applied to the many sources of, sometimes diverse, information that are being explored as possible links to the increase in autism incidence. In this study, the scoring systems considered are the rates of autism prevalence in children ages 3 to 5, lead (Pb) concentration in air, mercury (Hg) precipitation concentration, and mercury (Hg) deposition in soil.

B. *Rank-Score Function as a Diversity Measure*

Each scoring system: autism prevalence from 2000 to 2006 (*A*), autism prevalence in 2006 (*B*), mercury precipitation concentration (*H*); mercury deposition (*G*); and concentration of lead particulate matter in air (*P*) from 1996 to 2006, consists of a score function and rank function. The score function *s(d)* assigns a real number to each state, *d,* in the set $D = \{d_1, d_2, ..., d_n\}$ where *n* is the number of states. For scoring systems: *A*, *H*, *G*, and *P* the score function for each state is the slope of the linear regression line of the particular data over time. The autism prevalence for each state in the year 2006 is the score function of *B*.

The score functions of each scoring system are normalized before they can be compared and/or combined. Each score function $s(d)\colon D \rightarrow R$ is transformed to $s^*(d)\colon D \rightarrow [0,1]$ where $s^* = \frac{s(d) - s_{\min}}{s_{\max} - s_{\min}}, d \in D$ and $s_{max} = max\{s(d)|d\in D\}$ and $s_{min} = min\{s(d)|d\in D\}$. The rank function *r(d)* from *D* to $N = \{1, 2, ..., n\}$ assigns a rank to each state after sorting the array of scores *s(d)* in descending order.

A Rank-Score Characteristic (RSC) graph, as defined by Hsu et al [9], visualizes the rank/score function of a scoring system. A rank-score function is defined as $f : N \rightarrow R$ such that $f(i) = (s \circ r^{-1})(i) = s(r^{-1}(i))$. Diversity between two scoring systems can be measured by the distance between their rank-score functions [9]. If the distance is sufficiently large, the two systems are considered diverse. RSC graphs visualize a system's scoring behavior and can be used to measure diversity between multiple systems.

The performance of combined multiple scoring systems will be enhanced if each of the individual systems has relatively high performance and the scoring systems are diverse [22]. If the RSC graph for two scoring systems indicates diverse scoring behavior, we combine their score and rank functions to generate new information. The Rank-Score Characteristic graph for mercury precipitation concentration and deposition is depicted in *Fig. 1* and the rank/score graph of lead particulate matter in air and mercury deposition in soil in *Fig. 2*.

The rank-score characteristic graph for mercury precipitation concentration and mercury deposition indicates that these systems are not diverse since there is not a large distance between their rank/score functions. However, the rank/score graph for mercury deposition and lead concentration in air indicates that there is greater diversity between these two scoring systems since there is a larger distance between their rank/score functions.

Since these two scoring systems are diverse, we combine or "fuse" the score functions $s_G(d)$ and $s_P(d)$ and rank functions $r_G(d)$ and $r_P(d)$. The score function of the score combination for mercury deposition and lead particulate matter in air is defined as: $c_S(d) = \frac{1}{2}[s_G(d) + s_P(d)]$. The score function of the rank combination is then: $c_R(d) = \frac{1}{2}[r_G(d) + r_P(d)]$. The rank and score correlation matrix for the 20 state subset in *Table 1* is obtained by correlating the score and rank functions of *A* and *B* with *P*, *H*, *G*, and *P/G* combinations with Pearson's r (score) and Spearman's Rho (rank).

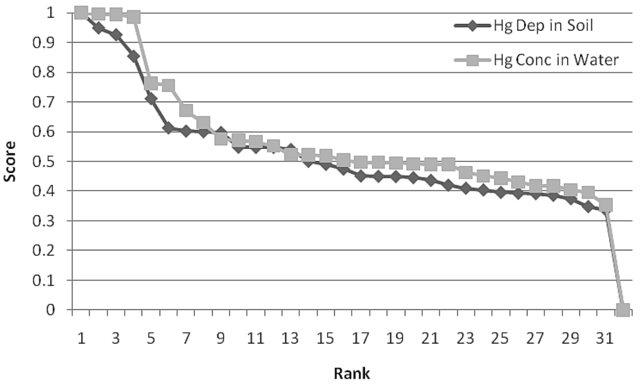

Figure 1. Rank-Score Characteristic Graph for Hg Conc in Water (H) and Hg Dep in Soil (G) for 33 States.

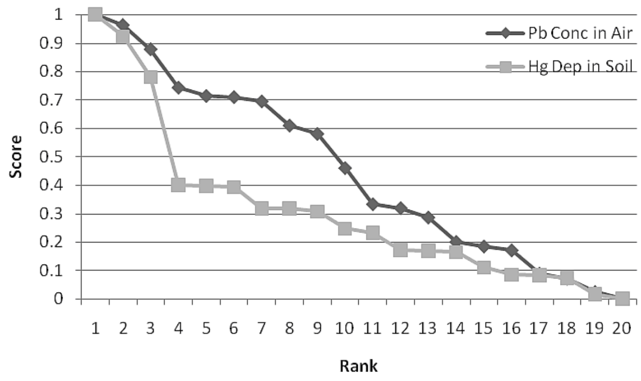

Figure 2. Rank-Score Characteristic Graph of Pb Conc in Air (P) and Hg Dep in Soil (G) for 20 States.

TABLE 1. RANK AND SCORE CORRELATION MATRIX FOR 20 STATESUBSET

	P (lead particulate matter in air)	*H (mercury precipitation concentration)*	*G (mercury deposition)*	*Combination of P and G*
A *(autism prev. 2000-2006)*	$0.121\ (s_A(d), s_P(d))$ $0.205\ (r_A(d), r_P(d))$	$-0.239\ (s_A(d), s_H(d))$ $-0.083(r_A(d), r_H(d))$	$0.121\ (s_A(d), s_G(d))$ $0.041\ (r_A(d), r_G(d))$	$0.162\ (s_A(d), c_S(d))$ $0.159\ (r_A(d), c_R(d))$
B *(autism prev. 2006)*	$0.207\ (s_B(d), s_P(d))$ $0.220\ (r_B(d), r_P(d))$	$-0.186\ (s_B(d), s_H(d))$ $-0.072(r_B(d), r_H(d))$	$0.227\ (s_B(d), s_G(d))$ $0.215\ (r_B(d), r_G(d))$	$0.292\ (s_B(d), c_S(d))$ $0.307(r_B(d), c_R(d))$

Preliminary analysis of raw data indicates the possibility of a correlation between autism prevalence and lead concentration in air and, interestingly, possible correlation between autism prevalence and the combination (rank and score) of lead and mercury. We find that the rank combination of P (concentration of lead particulate matter in air) and G (mercury deposition) has a higher correlation with autism prevalence than either one individually.

IV. ASSOCIATION RULE MINING

Association rule mining is a popular data mining method for discovering interesting relations between variables in large databases. Association rules are employed today in many application areas including Web mining, intrusion detection, and bioinformatics [6, 14, 19]. Data in a data set are considered transactions which consist of a set of items (variables) $V = \{v_1, v_2, ..., v_n\}$. Association rules are statements of the form $X => Y$ where $\{X,Y\} \subseteq$ V, and $X \cap Y = \{\ \}$, which means that if we find all of X in a transaction, then there is a good chance of finding Y in the same transaction [1, 2].

To select interesting rules from a set of all possible rules, various measures of significance and interest can be used. The first two are minimum thresholds on support and confidence [1]. The support of an itemset, *supp(X)*, is defined as the proportion of transactions in the data set which contain the itemset. The confidence of a rule, *confidence(X=>Y)*, is defined as the ratio between $supp(X \cup Y)$ and *supp(X)*, i.e.,

$$confidence(X=>Y) = supp(X \cup Y)/supp(X).$$

For example, a rule has a confidence of 0.2 / 0.4 = 0.5 in the data set, which means that for 50% of the transactions containing X, the rule is correct. Confidence can be interpreted as an estimate of the probability of finding Y in transactions under the condition that these transactions also contain X. The higher the confidence value, the more likely that X and Y have an interesting relation.

Another measure is the lift [5] of a rule, *lift(X=>Y),* which is defined as the ratio of the probability that X and Y occur together to the multiple of the two individual probabilities for X and Y, i.e.,

$$lift(X=>Y) = supp(X \cup Y)/(supp(X)*supp(Y)).$$

If *lift(X=>Y)* is 1, then X and Y are independent. The higher the lift value, the more likely that the co-existence of X and Y in a transaction is not by chance but because of the relation between them.

A. Preprocessing of Data Sets

In the raw data sets, we have autism prevalence (A) of the 3-5 age group of every state for each year from 2000 to 2006 and mercury precipitation concentration (H), mercury deposition (G), and lead air concentration (P,) which are monitored and collected several times a month at stations throughout most states for the years 1996 to 2006. The average Hg concentration, Hg deposition, and Pb concentration readings for each month are calculated and the value of a given year in the data sets H, G, and P is the average of the monthly values for that year (years with less than 6 months of data are removed).

We want to explore whether there is an interesting relation between autism prevalence and mercury/lead concentration in the environment during the life of the child and also during the pre-natal period. Since the autism patients are reported at age of 3-5, we focus on the autism reporting year in addition to the 4 prior years.

We define a data grid M(n, m) consisting of states $D=\{d_1, d_2, ..., d_i, ..., d_n\}$ and years $T=\{t_1, t_2, ..., t_j, ..., t_m\}$ in which each $M(i, j)$ entry has associated values for A, H, G, and P.

The trends we are interested in for values of A, H, G, and P for state d_i and year t_j are represented in the following matrix for entry $M(i, j)$ where the function

$d(x,y):R\rightarrow\{-1, 0, 1\}$ is defined as $d(x, y) = \begin{cases} 1 \text{ if } x > y \\ 0 \text{ if } x=y \\ -1 \text{ if } x<y \end{cases}$.

We then created three data sets, A vs. H, A vs. G, and A vs. P. Since there is a lot of missing data, we only keep those transactions which have at least one item for H, G, or P trends and an item for A. In total, there are 150 transactions in data set A vs. H, 151 transactions in data set A vs. G, and 157 transactions in data set A vs. P.

TABLE 2. ENTRY FOR STATE d_i AND YEAR t_j IN DATA GRID $M(n, m)$

A	*H*	*G*	*P*
$d(A_{t_j}, A_{t_{j-1}})$	$d(H_{t_j}, H_{t_{j-1}})$	$d(G_{t_j}, G_{t_{j-1}})$	$d(P_{t_j}, P_{t_{j-1}})$
	$d(H_{t_{j-1}}, H_{t_{j-2}})$	$d(G_{t_{j-1}}, G_{t_{j-2}})$	$d(P_{t_{j-1}}, P_{t_{j-2}})$
	$d(H_{t_{j-2}}, H_{t_{j-3}})$	$d(G_{t_{j-2}}, G_{t_{j-3}})$	$d(P_{t_{j-2}}, P_{t_{j-3}})$
	$d(H_{t_{j-3}}, H_{t_{j-4}})$	$d(G_{t_{j-3}}, G_{t_{j-4}})$	$d(P_{t_{j-3}}, P_{t_{j-4}})$
	$d(H_{t_j}, H_{t_{j-4}})$	$d(G_{t_j}, G_{t_{j-4}})$	$d(P_{t_j}, P_{t_{j-4}})$

B. Data Mining Results and Interpretation

After applying association rule mining to these three data sets, we find an interesting relation between the increase of mercury concentration in the environment of a given year and the increase of the autism prevalence three years later.

In data set *A* vs. *H*, the interesting itemset is $\{X(d(H_{t_{j-3}}, H_{t_{j-4}})=1), Y(d(A_{t_j}, A_{t_{j-1}})=1)\}$. The support of this itemset is 33%; the confidence of rule $X \Rightarrow Y$ is 0.96, which means that for any state, if the mercury concentration in the water increases in a year, there is 96% chance that three years later, the number of autism cases of children at age 3-5 increases. The lift of this itemset is 1.015, which is larger than 1. This tells us that it is likely the existence of *X* and *Y* together in a transaction is not just a random occurrence, but because there is a relation between them. For comparison, we also check the relation between $X'(d(H_{t_{j-3}}, H_{t_{j-4}})= -1)$ and *Y*.

The results show that the support of itemset *{X', Y}* is lower than that of itemset *{X, Y}*. The confidence of rule *X'=>Y* is also lower than that of rule *X=>Y*, and the lift of itemset (*X',Y)* is less than 1 (0.976), which means *X'* and *Y* are less likely to occur together than they are independent. The result supports the conclusion that *X* and *Y* are not independent. Therefore, *X=>Y* is a strong rule. If the size of the data set is larger, we could have more convincing results to support this interesting pattern.

In data set *A* vs. *G*, we find a similar pattern. The interesting itemset is:

$$\{X\ (d(G_{t_{j-3}}, G_{t_{j-4}})=1),\ Y\ (d(A_{t_j}, A_{t_{j-1}})=1)\}.$$

We also compare the relation between $X'\ (d(G_{t_{j-3}}, G_{t_{j-4}})= -1)$ and *Y*. The measure results of these two itemsets are listed in *Table 4*.

The interesting findings in these two data sets show that the mercury concentration in the environment has a special effect on the autism prevalence three years later. Since the child is diagnosed and reported as having autism when he/she is 3 to 5 years old, the interesting patterns suggest that when the child is 1 year old or younger, the increase in mercury concentration in the environment may play a role in the development of autism.

V. CONCLUDING REMARKS

Autism research has suggested that the cause of autism may be in part genetic. In addition to heredity, environmental factors, such as exposure to neurotoxins, are also believed to play a role. Several studies have concluded that genetic susceptibility during critical stages of development may be vital in determining the effects of mercury and lead exposure. In this study, we apply Combinatorial Fusion Analysis (CFA) and Association Rule Mining (ARM) to autism, lead, and mercury data.

The novel approach of CFA for analysis of multiple scoring systems was applied to autism prevalence, mercury, and lead data. The rank-score characteristic graph visualizes the scoring behavior and diversity among the systems. The rank and score functions of diverse systems, in this case mercury deposition and concentration of lead particulate matter in air can then be combined. The CFA analysis revealed a higher correlation between autism prevalence and the rank combination of mercury and lead than with either system alone.

Association Rule Mining is applied to large databases to extract trends. In this study, we are interested in finding a relationship between an increase in mercury or lead and an increase in autism. We discovered a trend where an increase in mercury was strongly related to an increase in autism prevalence three years later.

Application of information fusion and data mining techniques, such as combinatorial fusion analysis and association rule mining, to autism prevalence and neurotoxin research enriches existing knowledge that will be valuable in furthering and refining our interpretation and understanding of the relationships between autism and neurotoxins and their combinations. The results of our preliminary analysis are in accord with other studies that propose environmental neurotoxin levels may be related to autism prevalence in some cases.

Since this study used raw measured lead and mercury data, we faced some limitations due to sparse and missing data. Further investigations will include application of these techniques to complete modeled neurotoxin data and incorporation of additional confounders. Combinatorial fusion analysis and data mining techniques can also be applied to other information related to autism research for further exploration.

TABLE 3. TRANSACTION OF FLORIDA 2001 IN DATA SET - A VS. G

d_i	t_j	$d(A_{t_j}, A_{t_{j-1}})$	$d(G_{t_j}, G_{t_{j-1}})$	$d(G_{t_{j-1}}, G_{t_{j-2}})$	$d(G_{t_{j-2}}, G_{t_{j-3}})$	$d(G_{t_{j-3}}, G_{t_{j-4}})$	$d(G_{t_j}, G_{t_{j-4}})$
FL	2001	1	1	1	-1	1	1

TABLE 4. ARM RESULTS FOR DATA SETS A VS. H AND A VS. G

(support, confidence, lift)	**X => Y**	**Y =>X**	**X' => Y**	**Y=> X'**
A vs. H X: $d(Ht_{j-3}, Ht_{j-4})=1$ X': $d(Ht_{j-3}, Ht_{j-4})= -1$ Y: $d(At_j, At_{j-1})=1$	(0.33, 0.961, 1.015)	(0.33, 0.352, 1.015)	(0.26, 0.930, 0.976)	(0.26, 0.282, 0.976)
A vs. G X: $d(Gt_{j-3}, Gt_{j-4})=1$ X': $d(Gt_{j-3}, Gt_{j-4})= -1$ Y: $d(At_j, At_{j-1})=1$	(0.34, 0.963, 1.017)	(0.34, 0.363, 1.017)	(0.25, 0.927, 0.978)	(0.25, 0.266, 0.978)

REFERENCES

[1] Agrawal, R., Imielinski, T., and Swami, A. "Mining association rules between sets of items in large databases," In Proceedings of the ACM SIGMOD International Conference on Management of Data, pages 207-216, Washington D.C., May 1993.

[2] Agrawal, R. and Srikant, R. "Fast algorithms for mining association rules in large databases," In Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago, Chile, September 1994.

[3] Bernard S, Enayati A, Redwood L, Roger H, Binstock T. Autism: a novel form of mercury poisoning. Med Hypotheses 2001; 56 : 462-71.

[4] Bernard S, Enayati A, Roger H, Binstock T, Redwood L. The role of mercury in the pathogenesis of autism. Mol Psychiatry 2002; 7 (Suppl 2) : S42-3.

[5] Brin, S., Motwani, R., Ullman, J.D., and Tsur, S. "Dynamic itemset counting and implication rules for market basket data," In Proceedings ACM SIGMOD International Conference on Management of Data, pages 255-264, Tucson, Arizona, USA, May 1997.

[6] Creighton, C. and S. Hanash. Mining Gene Expression Databases for Association Rules. Bioinformatics Vol 19 no. 1, pp 79-86. 2003.

[7] Geier, D.A., Geier, M.R. A Prospective Assessment of Porphyrins in Autistic Disorders:A Potential Marker for Heavy Metal Exposure. Neurotoxicity Research, 2006, VOL. 10(1). pp. 57-64

[8] Geier, D.A., King, P.G., Sykes, L.K., Geier, M.R. A comprehensive review of mercury provoked autism. Indian J Med Res 128, October 2008, pp 383-411.

[9] Hsu, D.F., Chung, Y.S., & Kristal, B.S. Combinatorial fusion analysis: methods and practice of combining multiple scoring systems, in: H.H. Hsu (Ed.), Advanced Data Mining Technologies in Bioinformatics, Idea Group Inc. 2006.

[10] Hsu, D.F. & Taksa, I. Comparing rank and score combination methods for data fusion in information retrieval. Information Retrieval, 8(3), 2005, 449-480.

[11] Lin, K.-L.; Chun-Yuan Lin; Chuen-Der Huang; Hsiu-Ming Chang; Chiao-Yun Yang; Chin-Teng Lin; Chuan Yi Tang; Hsu, D.F. Feature Selection and Combination Criteria for Improving Accuracy in Protein Structure Prediction, NanoBioscience, IEEE Transactions on Volume 6, Issue 2, June 2007 Page(s):186 – 196.

[12] Lyons, D.M. & Hsu, D.F. Combining multiple scoring systems for target tracking using rank–score characteristics, Information Fusion. 2008.

[13] Marshall, C.R., Noor, A., Vincent, J.B., et al, Structural Variation of Chromosomes in Autism Spectrum Disorder, The American Journal of Human Genetics, Volume 82, Issue 2, 477-488, 17 January 2008.

[14] Mobasher, B., Dai, H., Luo, T. and Nakagawa, M. Effective personalization based on association rule discovery from web usage data. In Proceedings of the 3rd ACM Workshop on Web Information and Data Management (WIDM01), Atlanta, Georgia, November 2001.

[15] Nataf, R., Skorupka, C., Amet, L., Lam, A., Springbett, A., and Lathe, R. (2006) Porphyinuria in childhood autistic disorder: implications for environmental toxicity. Toxicol. Appl. Pharmcol. 214, 99-108.

[16] Palmer, R.F., Blanchard, S., Stein, Z., Mandell, D., and Miller, C. Environmental mercury release, special education rates, and autism disorder: an ecological study of Texas. Health & Place. 2006; 12:203–209.

[17] Ravinesh, K. A.; Samer, K.; Jyotsna, S., et al. Recurrent 16p11.2 microdeletions in autism. Human Molecular Genetics:Volume 17(4)14 February 2008pp 628-638.

[18] Rose, S., Melnyk, S., Savenka, A., Hubanks, A., Jernigan, S., Cleves, M., and James, S. J. The Frequency of Polymorphisms affecting Lead and Mercury Toxicity among Children with Autism. American Journal of Biochemistry and Biotechnology 4 (2): 85-94, 2008.

[19] Treinen, J.J. & Thurimella, R. A framework for the application of association rule mining in large intrusion detection infrastructures. In RAID 2006: Proceedings of the 9th Annual Symposium on Recent Advances in Intrusion Detection, pages 1--18. Springer Berlin/Heidelberg, 2006.

[20] Weiss, L.A., Shen, Y., Korn, J.M., et al. Association Between Microdeletion and Microduplication at 16p11.2 and Autism, New England Journal of Medicine, 2008; 358:667-675.

[21] Windham, G.C., Zhang, L., Gunier, R., Croen, L.A., and Grether, J.K. Autism spectrum disorders in relation to distribution of hazardous air pollutants in the San Francisco Bay area. Environ Health Perspect. 2006 September; 114(9): 1438–1444.

[22] Yang, J.M., Chen, Y.F., Shen, T.W., Kristal, B.S., and Hsu, D.F. Consensus scoring for improving enrichment in virtual screening. Journal of Chemical Information and Modeling. 45 (2005) 1134-1146.

[23] Individuals with Disabilities Education Act (IDEA) Data. https://www.ideadata.org/PartBChildCount.asp

[24] U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, http://www.cdc.gov/nchs/births.htm

[25] U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Autism Information Center, http://www.cdc.gov/ncbddd/autism/

[26] National Atmospheric Deposition Network. Mercury Deposition Network. Champaign, Ill. http://nadp.sws.uiuc.edu/mdn/.

[27] U.S. Department of Environmental Protection. Air Quality System. Research Triangle Park, N.C. http://www.epa.gov/ttn/airs/airsaqs/index.htm

[28] Visibility Information Exchange Web System. AQS Fine Speciation. http://vista.cira.colostate.edu/views/