6-2012

A Knowledge-based Clinical Toxicology Consultant for Diagnosing Single Exposures

Joel D. Schipper
Embry-Riddle Aeronautical University, schippej@erau.edu

Douglas D. Dankel II
University of Florida

A. Antonio Arroyo
University of Florida

Jay L. Schauben
Florida Poison Information Center - Jacksonville, Shands Jacksonville Medical Center/University of Florida Health Science Center.

Scholarly Commons Citation

Schipper, J. D., Dankel II, D. D., Arroyo, A. A., & Schauben, J. L. (2012). A Knowledge-based Clinical Toxicology Consultant for Diagnosing Single Exposures. *Artificial Intelligence in Medicine, 55*(2). https://doi.org/10.1016/j.artmed.2012.03.006

Published in *Artificial Intelligence in Medicine*, Volume 55, Issue 2, June 2012, Pages 87-95 (http://dx.doi.org/10.1016/j.artmed.2012.03.006).

A knowledge-based clinical toxicology consultant for diagnosing single exposures

Joel D. Schipper[a,*], Douglas D. Dankel II[b], A. Antonio Arroyo[c], Jay L. Schauben[d]

[a] Electrical and Computer Engineering, Embry-Riddle Aeronautical University, 3700 Willow Creek Road, Prescott, AZ 86301, USA

[b] Computer and Information Science and Engineering, Box 116120, University of Florida, Gainesville, FL 32611, USA

[c] Electrical and Computer Engineering, Box 116200, University of Florida, Gainesville, FL 32611, USA

[d] Florida Poison Information Center – Jacksonville, Shands Jacksonville Medical Center/University of Florida Health Science Center – Jacksonville, 655 West 8th Street, Box C23, Jacksonville, FL 32209, USA

[*] Corresponding author. Tel.: +1 928 777 4205; Fax: +1 928 777 6945. *E-mail address*: joel.schipper@erau.edu (J. Schipper).

Correspondence Address:
E-mail: joel.schipper@erau.edu

Joel Schipper
Electrical & Computer Engineering
3700 Willow Creek Road
Prescott, AZ 86301
USA

Abstract

Objective: Every year, toxic exposures kill twelve hundred Americans. To aid in the timely diagnosis and treatment of such exposures, this research investigates the feasibility of a knowledge-based system capable of generating differential diagnoses for human exposures involving unknown toxins.

Methods: Data mining techniques automatically extract prior probabilities and likelihood ratios from a database managed by the Florida Poison Information Center. Using observed clinical effects, the trained system produces a ranked list of plausible toxic exposures. The resulting system was evaluated using 30,152 single exposure cases. In addition, the effects of two filters for refining diagnosis based on a minimum number of exposure cases and a minimum number of clinical effects were also explored.

Results: The system achieved accuracies (calculated as the percentage of exposures correctly identified in top 10% of trained diagnoses) as high as 79.8% when diagnosing by substance and 78.9% when diagnosing by the major and minor categories of toxins.

Conclusions: The results of this research are modest, yet promising. At this time, no similar systems are currently in use in the United States and it is hoped that these studies will yield an effective medical decision support system for clinical toxicology.

KEYWORDS
Knowledge-based systems;
Decision support systems;
Data mining;
Differential diagnosis;
Clinical toxicology

1. Introduction

Toxicology is the study of poisons and their effects on living organisms. One prominent societal benefit of toxicology is the development of poison control centers (PCCs). Thousands of people call PCCs daily for free consultation and information regarding chemicals and drugs. In 2007, the American Association of Poison Control Centers (AAPCC) consisted of 61 PCCs serving all 50 of the United States and handling more than 2.4 million human poison exposure cases [1]. The AAPCC has compiled a database containing the details of nearly 46 million human poison exposure cases from the calls received and documented by its constituent PCCs [1]. A medical database of this magnitude represents a remarkable opportunity for data mining and knowledge-based system (KBS) research.

1.1. System overview and design principles

The goal of this research is to create a KBS for clinical toxicology that automatically learns relationships from PCC databases. A brief overview of the system is given in abstracts [2,3], which summarize the KBS and its results in minimal detail. The system mines the database of the Florida Poison Information Center (FPIC), extracting associations between the clinical effects (CEs) (i.e., signs and symptoms) observed in a patient and the final diagnosis. The system has a hybrid design, containing elements of data mining, case-based reasoning, rule-based systems, and uncertainty management. Data mining techniques clean and extract relevant information from the database. The case-based reasoning component uses the example cases obtained from data mining to generate a collection of likelihood ratios (LRs) based on composite observations. The system calculations performed with these LRs are effectively a set of simple rules running in parallel, and the LRs provide a means of handling uncertainty.

Throughout development, we sought to produce a system with specific characteristics, including simplicity, understandability, automatic system generation, and incremental updates. The characteristic of simplicity is of utmost importance. Holsheimer et al. [4] show that success in extracting information from databases does not require complex algorithms. In fact, "simpler, even trivial, processes are better than complicated ones if they are enough for the job of discovery" [5]. Generally, systems with simple representations and algorithms are more efficient, require less processing power, are more portable, and provide scalability, which is extremely important given the size and continual growth of the FPIC database. Mining large databases is challenging [6,7] and as the FPIC database grows to include more cases, the system's simple mathematical representation will become essential for scalability. Furthermore, simplicity of design gives the system inherent understandability.

The understandability of system results was another chief concern during development. If physicians understand the method by which the system obtains its answers, they are more likely to trust the system and use it within the spectrum of its intended purpose. According to Atzmueller et al. [8], "understandability and interpretability of…learned models is of prime importance" and "ideally, the learning method constructs knowledge in the same representation the human expert favors." For this reason, our system makes use of pre-test probabilities (prior probabilities) and LRs, which are commonly used throughout the medical field, and presents its results to the user as a differential diagnosis. By using these familiar approaches, physicians should find the system to be relevant, understandable, and easy to operate. Additionally, the methods used in the system's mathematics are similar to medical case studies seeking to identify patterns of clinical syndromes, which should help the system gain acceptance. Medical mathematics is not only more understandable to users in the medical field, it communicates

information that is more relevant for medical diagnosis than other traditional measurements [9,10].

Automatic system generation was another primary objective. Although Atzmueller et al. [8] state that "pure automatic learning methods are usually not good enough to reach a quality comparable to manually built knowledge bases," automatic methods offer certain advantages that should not be overlooked. Automatically trained systems fully bypass the need to interview experts, which is the knowledge acquisition bottleneck of traditional KBS development [11], as well as relieving the knowledge engineer of the burden of acquiring extensive knowledge about toxicology to implement the system. Furthermore, Kononenko et al. [12] have shown that, in many domains, systems that automatically generate their own diagnostic rules are capable of performing with a higher degree of accuracy than individual physicians, when given identical information. The system presented in this paper was developed by an engineer with no expertise in the area of toxicology and no guidance from toxicologists regarding specific diagnostic approaches.

The final desired attribute of the system is the ability to perform incremental updates. As the FPIC database grows in size, additional information becomes available for aiding in diagnosis. Although the system could recompile all the data from 1996 to the present with every update, such an operation would be inefficient and could require significant processing time. By storing basic summary data from the previous update, LRs can be recalculated without the need to reprocess every case in the database.

The current system is a proof-of-concept prototype capable of generating a differential diagnosis for exposures to a single toxin. The system seeks to serve as a consultant to all medical personnel that may encounter toxic exposure cases. It is not meant to replace humans, which

have intuition and highly developed senses that are difficult for computers to replicate. Rather, it supplements human intelligence by providing case-based diagnostic information that a human would be hard-pressed to obtain without the use of a computer. Ultimately, the healthcare professional makes the final decision regarding the treatment the patient should receive.

1.2. Applicability of knowledge-based systems to clinical toxicology

Medicine is a continually expanding field where diagnoses often must be made using incomplete data. These traits make medicine an ideal domain for KBSs because of their adaptability to dynamic domains and ability to cope with uncertainty. Beyond the general complications common to all fields of medicine, toxicology faces two specific challenges for which KBSs are well tailored.

The first challenge is making pertinent information readily available at the time of a poisoning. Since toxicology is a narrow specialization within medicine, only a limited number of experts, known as toxicologists, exist. Although KBSs cannot replace toxicologists, KBSs can aid physicians in diagnosis by offering ubiquitous access to toxicological information when these rare experts are unavailable.

The second toxicological challenge is ensuring the rapid diagnosis and treatment of exposures. In 2007, 1239 people died of toxic exposures [1]. When treating potentially harmful or lethal exposures to drugs or poisons, time is of the essence. The sooner an informed initial therapeutic decision can be made, the better the prognosis. KBSs provide rapid aid in the assessment and management of toxic exposures by offering physicians an initial differential list of diagnoses to consider without having to wait for expert consultations or for the analysis of blood or urine.

1.3. Knowledge-based systems in medicine

The application of rule-based expert systems and KBSs to medicine began in the early 1970's with the development of MYCIN [13-14] and INTERNIST [15-18]. Numerous other approaches to medical diagnosis and data mining have been implemented, including case-based reasoning [19-25], set covering [23-26], Bayesian belief networks [27-29], fuzzy logic [30-31], rough sets [32-34], genetic algorithms [21,35], and artificial neural networks [22,36].

Although many KBSs exist for predictive toxicology to determine the toxicity of chemicals based on structure (e.g., [37-40]), there are surprisingly few systems in the field of clinical toxicology, which focuses on the diagnosis and treatment of toxic exposures. In fact, according to Darmoni et al. [41], in 1995 "Toxline and Toxlit [showed] that less than ten computer-aided decision support systems [had] been developed in clinical toxicology." Of these systems, two in particular stand out from the rest. The first is a French system called SETH [41,42] that was developed for use in Rouen University Hospital. SETH's inference engine is a rule-based, forward chaining system that utilizes the Rete algorithm for pattern matching and set theory for diagnosing cases involving multiple drugs. The system began experimental use in 1992 and was used in the diagnosis of over 2000 drug intoxication cases. Due to lack of objective criterion, SETH only received an internal evaluation from experts at Rouen University Hospital. The second system, known as MEDICOTOX-CONSILIUM [43], was developed for use in Bulgarian hospitals as a diagnostic system for first aid clinical toxicology. MEDICOTOX-CONSILIUM uses frame structures, rules, and scores provided by experts for diagnosis. The system contains 1000 rules and facts that use 47 syndrome and 134 symptom definitions to identify poisons and supply the user with information about the appropriate cure from any of 86 treatments and 55 antidotes. Users of MEDICOTOX-CONSILIUM responded positively, however, only a few

example cases are offered as evidence of system functionality rather than any data from extensive testing.

More recently, another system was created for use by the Russian Toxicology Information and Advisory Center in Moscow [20]. The system uses the Inreca (Induction and Reasoning from Cases) approach, which involves case-based reasoning using a pre-compiled decision tree based on a Russian database. Although preliminary tests yielded accuracies ranging from 78.5% to 96.0%, the system was only tested on eight types of acute poisonings.

2. Source data

Since 1996, the FPIC has collected data on every call received. In 2008 alone, the FPIC received over 117 thousand calls and made more than 46 thousand follow-up calls related to human exposures [44]. For this research, the FPIC provided access to four years of cases recorded in the Jacksonville database, comprising more than 160 thousand toxic exposure cases.

The database supplied by the FPIC conforms to the standards set forth by the National Poison Data System (NPDS) and its predecessor, the Toxic Exposure Surveillance System (TESS). These national standards are defined by the AAPCC and regulate the mandatory fields contained within the database of each PCC, guaranteeing that every PCC records CEs and final diagnoses for each case. Following these standards ensures that these entries in the database have discrete values that are easy to process with a computer algorithm, increasing the portability of a system designed for the FPIC to other PCCs around the country.

3. System development

In generating the system, data mining techniques are used to clean the records and extract the appropriate information from the FPIC database. Informational calls are removed so that only exposure cases remain. These exposure cases are filtered so that only cases with CEs followed to

a known outcome remain. Although this reduces the size of the dataset to 30,152 single exposure cases, the filtering process ensures that only significant representative cases with the best documentation are used to train the system. The database contains flags indicating whether the CEs observed in a patient were "related," "unknown if related," or "not related" to the substance involved in the exposure. The CEs marked as "not related" are removed from the database whereas those marked as "related" and "unknown if related" are used for system training.

After cleaning, tables of prior probabilities and LRs are calculated for each toxin. A prior probability, *P*, representing the likelihood of a particular substance being involved given that a toxic exposure has occurred, is calculated as:

$$P = \frac{Cases}{Total}, \tag{1}$$

where *Cases* is the number of cases involving a particular substance and *Total* is the total number of exposure cases in the database [45].

When calculating LRs, the system treats each CE as an independent diagnostic test used to detect the presence of a toxic substance. LRs represent the odds that an observed CE is caused by a particular toxin versus the odds that the CE is the result of exposure to any other toxin. The LR corresponding to positive test results, LR^+, is calculated as:

$$LR^+ = \frac{\left(\frac{TP}{TP + FN}\right)}{\left(\frac{FP}{FP + TN}\right)}, \tag{2}$$

where *TP* represents true positives, *TN* represents true negatives, *FP* represents false positives, and *FN* represents false negatives [46]. An exhaustive table of LRs relating every individual CE to every possible substance exposure is the primary resource utilized by the system in creating a

differential diagnosis. The advantage of LRs over other medical measurements (i.e., sensitivity, specificity, positive and negative predictive values, etc.) is that LRs can be easily combined through multiplication. Moreover, LRs are easily calculated, can account for disorder prevalence by including the prior probabilities, and characterize many cases with a single number, making the system scalable to large databases while ensuring a rapid response time.

Although the LR has its advantages, it inherently contains the drawbacks of every mathematical ratio, the possibility of evaluating to zero or causing a divide-by-zero error. An LR of zero only occurs when $TP = 0$. Since every CE is treated as an independent test for detecting the presence of a toxic substance, when the system combines LRs by multiplication, any individual LR equal to zero forces the combined LR to be zero. To the contrary, the absence of cases in the database associating a substance with a CE is not a definitive indication that the substance cannot cause that CE. Furthermore, even if the substance truly cannot cause the CE, patients may have unassociated CEs caused by other ailments.

The divide-by-zero error occurs if $TP + FN = 0$ or $FP = 0$. (Note that although $FP + TN = 0$ causes an error, addressing $FP = 0$ also prevents that error from occurring.) The sum of true positives and false negatives ($TP + FN$) is the total number of cases where a particular substance is involved. Due to the structure of the database and its queries, a substance with no recorded cases in the database is ignored and not included in the system. As a result, $TP + FN$ never equals zero. The second divide-by-zero error, $FP = 0$, occurs when a CE in the database is only associated with one particular substance. In reality, however, no single substance is the only possible cause for any CE in the system and the problem only occurs due to a lack of sufficient data.

The preliminary system used a simple-minded approach to solving the multiplication by zero and divide-by-zero problems. Multiplication by zero was handled by replacing all zero-valued LRs with a value of one. Although this prevents the system from gaining any knowledge about a substance from a CE not associated with the substance, it prevents that CE from destroying the knowledge gained from other CEs. The divide-by-zero error was solved by examining the data set and manually modifying the offending records so CEs lacking sufficient data were omitted when computing LRs. The LRs calculated using the method described in this paragraph are referred to as "non-adjusted" LRs from this point forward.

The current system utilizes a generalized equation developed to replace LR^{+}:

$$LR^{+}_{Adj} = \frac{\left(\frac{TP + \Delta}{TP + \Delta + FN + \Delta}\right)}{\left(\frac{FP + \Delta}{FP + \Delta + TN + \Delta}\right)}, \tag{3}$$

where Δ is a small, positive constant, often called a pseudocount [47]. *TP*, *TN*, *FP*, and *FN* are the four possible outcomes of a diagnostic test. By adding Δ to each outcome, the equation states that any of these outcomes is a possibility, even if no supporting cases exist in the database. The end result is a stable equation that closely approximates (2), avoids the difficulties of multiplying by zero, prevents the divide-by-zero error, and converges to the same value as (2) as the number of cases increases. A variety of Δ values were compared, including 1.0, 0.1, 0.01, and 0.001. Due to the sparse nature of the toxic exposure data, frequently only a single true positive exists in the database, resulting in $TP = 1$. As a result, it was discovered that using a Δ of 1.0, which essentially doubles the value of *TP*, adversely affected the accuracy of the calculation. Ultimately, a Δ of 0.01 was selected as it altered the value of single true positives (or any other

outcome) by a negligible amount of 1%, yielding a suitable substitute for (2). Equation (3) with $\Delta = 0.01$ is referred to as the "adjusted" LR from this point forward.

Several other terms and techniques exist for producing similar results to the pseudocount, including small-sample correction [48], smoothing [48], Laplace estimators [49], and M-estimators [50]. The method selected in this research closely resembles [47], which also utilizes a pseudocount value of 0.01.

4. System operation

As discussed in the previous section, the system utilizes two tables of calculations to create a differential diagnosis, a table of prior probabilities for every substance and a table of LRs relating every individual CE to every possible substance exposure. When supplied with a set of CEs, the system calculates a combined LR, including the prior probability, for every potential single exposure diagnosis. The results are then sorted and presented as a differential diagnosis in the form of a ranked list. (Note: The odds-ratio form of Bayes' theorem [46] can be used interchangeably with the current system calculations. The only difference is that Bayes' theorem uses prior odds rather than the prior probability, so both calculations produce an identical ordering of diagnoses on the ranked list.)

The system's user interface organizes CEs into nine categories defined by TESS: cardiovascular, dermal, gastrointestinal, heme/hepatic, neurological, ocular, renal/GU, respiratory, and miscellaneous. The user is also able to select whether to diagnose the exposure by substance, major and minor categories, or major category. Thus far, we have discussed the research only in terms of diagnosing exposures to a single toxic substance; however, each substance (e.g. black widow spider poison) belongs to a minor category (e.g. spider bites), which in turn belongs to a major category (e.g. bites and envenomations). In the same manner that the

system is trained to diagnose individual substances, it can be trained to diagnose based on major and minor categories or even solely based on major category. Giving physicians a general idea of the drug categories to be considered can prove as valuable as attempting to directly diagnose a specific substance.

If desired, users may also adjust the number of "Minimum Exposure Cases" (MC) and "Minimum CE Occurrences" (MCE) used by the system. MC and MCE serve as data filters that eliminate diagnoses and CEs with poor representative sampling sizes. The MC filter allows the user to specify the minimum required number of cases for a diagnosis. If a diagnosis does not have at least this many cases in the database, the diagnosis does not appear on the resulting differential diagnosis list. The MCE filter allows the user to set the minimum number of times a CE must appear in the database. If a CE does not appear in the database at least that many times, the CE is ignored when calculating the LR even if the CE is selected by the user.

Evaluating the exposure based on the selected CEs displays a differential diagnosis list similar to the one shown in Fig. 1. The list contains the calculated LR on the left and the associated diagnosis on the right. The results in the figure indicate bacterial food poisoning, with an LR of 148.9, is by far the most likely cause of abdominal pain, dehydration, and diarrhea. The second most likely cause is mushrooms, with an LR of 3.32. It should be noted that, although LRs are helpful for indicating the strength of support for various diagnoses, rank within the list is more important. Physicians should consider the top ten substances before making their final diagnosis.

5. System testing and results

During the various stages of testing, the system's prior probabilities and LRs were trained and tested using 10-fold cross-validation. For each toxic exposure in the database, the rank of the

correct diagnosis on the system's differential diagnosis list was saved to a summary table. The results of testing are discussed below.

The first stage of testing compared the effectiveness of adjusted versus non-adjusted LRs. Both LR calculations were tested at all three diagnostic levels: diagnosing by substance, diagnosing by major and minor categories, and diagnosing by major category alone. Additionally, the settings for the MC and MCE filters were varied to produce multiple points of comparison. While maintaining a constant MCE value of 10, MC was tested at 10, 25, and 100. Likewise, while maintaining a constant MC value of 25, MCE was tested at 0, 10, and 50. Furthermore, four levels of medical outcomes were tested against the system: all exposures with a minor severity or worse, moderate severity or worse, major severity or worse, and a severity level where the outcome was death. These tests yielded sixty resultant sets for both adjusted and non-adjusted LRs.

After generating these results, the accuracy of the sixty adjusted sets was compared to the accuracy of the sixty non-adjusted sets. For our purposes, we define accuracy as the percentage of correct diagnoses, which is calculated in three ways: the percentage of exposures appearing as the top diagnosis, the percentage of exposures appearing in the top ten diagnoses, and the percentage of exposures appearing in the top 10% of the trained diagnoses. Comparing adjusted accuracies with non-adjusted accuracies, it was determined that adjusted LRs appear to be a good approximation of non-adjusted LRs, with adjusted calculations yielding a higher accuracy 90% of the time. Of the 180 accuracy calculations, there were eighteen exceptions where non-adjusted calculations outperformed adjusted calculations. Ten of these exceptions involved the outcome of death. There are a few explanations for this anomaly. First, the database contains very few recorded death cases, making it more likely that a random variation might favor one system

approach over another. Second, death cases may often display CEs that are not normally associated with a particular toxic exposure. The reason is that as the body systems begin to shut down, extreme failures begin to cause cascading effects. In such cases, it becomes impossible to reliably compare two diagnostic systems. The accuracies of the remaining eight exceptions were within 0.5% of the corresponding adjusted performances. This nominal gain is more than compensated for by the 127 instances where adjusted calculations outperformed non-adjusted calculations on test cases not limited to the outcome of death. Additionally, a system based on adjusted calculations is much easier to generate automatically than one based on non-adjusted calculations because it does not require any manual intervention by the system designer. Having established that the adjusted LR is a valid substitute for the traditional LR, the remainder of the research results is discussed in terms of adjusted calculations.

The next step in system development was to determine the best values for the MCE and MC filters. Beginning with a constant MC value of 25, MCE was tested at values of 0, 2, 5, 10, and 50. For each of these values, the adjusted system was also tested at the three diagnosis levels of substance, major and minor categories, and major category alone. Each of the three diagnosis levels yielded a system with a significantly different number of trained diagnoses. To enable comparisons between the three diagnosis levels, the percentage of exposures appearing in the top 10% of the trained diagnoses was used as the accuracy measurement. Table 1 shows the accuracy of the system when diagnosing by substance, by major and minor categories, and by major category alone for various MCE values. Looking at the diagnosis of minor severity cases by substance, it can be seen that varying MCE has no effect on the accuracy of the system. Under moderate and major severity, the accuracy decreases from 74.4% to 74.2% and 77.8% to 77.6%, both negligible changes. Likewise, looking at the other data in Table 1 it becomes obvious that

varying MCE causes little to no change for minor, moderate, and major severities. Once again, the exception is the severity where the outcome is death, which is most likely due to a small sampling size. For example, when diagnosing by major and minor categories the 5.1% increase in accuracy observed in exposures with an outcome of death is a difference of only four additional cases being diagnosed in the top 10%. Prior to these tests, it was believed that using too low of an MCE cutoff might create falsely high or low LRs in some substances, decreasing diagnosis accuracy. However, based on these results, it is reasonable to conclude that filtering by MCE yields negligible changes in system accuracy. Using an adjusted LR with $\Delta = 0.01$ already mitigates the potential problem, thus, the filter can be removed from the system.

The second filter to be examined was the MC filter. Using adjusted calculations with a constant MCE value of 10, MC was tested at values of 0, 2, 5, 10, 25, 50, and 100. Again, to enable comparisons between the three diagnosis levels of substance, major and minor categories, and major category alone, the percentage of exposures appearing in the top 10% of the trained diagnoses was used as the accuracy measurement. Additionally, since varying MC directly affects the number of trained diagnoses in the system, it was hoped that the 10% accuracy measurement would enable comparisons between systems generated by different MC filter values. Table 2 shows the accuracy of the system when diagnosing by substance, by major and minor categories, and by major category alone for various MC values. Looking at the accuracies for minor, moderate, and major severities when diagnosing by substance or by major and minor categories, it is readily apparent that accuracy generally decreases as MC increases. Diagnosing by major category alone has the same tendency as MC steps from 10 to 25 and 50 to 100, but seems to plateau for MC values from 0 to 10 and 25 to 50. At first it might appear that using a lower MC yields a more accurate system, and, therefore, the MC filter should be removed.

However, such a conclusion fails to account for the purpose of the MC filter. As MC decreases, more possible diagnoses with less supporting cases are added to the system. As more diagnoses are added to the system, the accuracy calculation based on the top 10% includes substances that are ranked lower on the differential diagnosis. It turns out that the number of diagnoses that are added to the top 10% outweighs the number of new exposure cases being tested against the system. As a result, the lower the MC value, the more accurate the system appears. The plateaus observed when diagnosing by major category alone are also accounted for by this explanation because the top 10% of cases evaluate to the same number for MCs of 0, 2, 5, and 10 as well as for MCs of 25 and 50.

Since comparing MC values using an accuracy based on the top 10% of trained diagnoses failed to demonstrate proper system training, a second accuracy measurement was calculated using the correct diagnoses appearing in the top ten slots of the differential diagnosis. From a user standpoint, this accuracy measurement is more appropriate because the list size that a user can process without being overwhelmed is not dependent on the number of trained substances. Looking at the minor, moderate, and major severity rows in Table 3, it can be seen that as MC increases, accuracy also increases. The data tell little about selecting a value for MC because they indicate what is expected of any system: as more cases are used to define each substance, system accuracy should increase. Another contributor to the increase in accuracy is that fewer substances are trained as MC increases. With fewer substances, the top ten substances become a larger portion of the available diagnoses. Even random guessing would experience an increase in accuracy under these circumstances.

To prove that the system is training properly requires the accuracies based on correct diagnoses in the top ten to be normalized. In the first attempt to normalize accuracies, a ratio of

the data in Table 3 versus the accuracy of diagnosing by random guessing was calculated. However, it was found that the ratio suffered from problems similar to the accuracies calculated in Table 2. Lowering MC increases the number of trained diagnoses in the system, adversely affecting random guessing. As a result, the ratio falsely indicated that a lower MC cutoff would yield better results. A second attempt at normalizing the accuracies calculated the ratio of the data in Table 3 against a system that selected its top ten choices based on prior probabilities alone.

Fig. 2 shows a graph of the ratio for minor, moderate, and major severities when diagnosing by substance. Likewise, Fig. 3 displays the ratio for diagnosing by major and minor categories and Fig. 4 the ratio for diagnosing by major category alone. The graphs indicate that as MC increases, ensuring better representative likelihood calculations, the system tends to perform better. The increase appears to be almost linear. There is no evidence of any breakpoints that would yield a superior MC cutoff. These results indicate that the adjusted LR is training correctly and that the exact value used for MC is unimportant. However, a reasonable MC value of at least ten should be chosen to ensure that outliers do not excessively influence the diagnosis.

Comparing Fig. 2, Fig. 3, and Fig. 4, it can be seen that the slopes and the ratios are higher for diagnosis by substance than diagnosis by major and minor categories, which in turn are higher than diagnosis by major categories. The reason is that the number of diagnoses trained for diagnosing by substance (around 200 to 600) is significantly more than diagnosing by major and minor categories (around 100 to 200), which is more than diagnosing by major category alone (around 50 to 60). With more possible diagnoses, the problem becomes more difficult to place a diagnosis in the top ten without intelligence. Thus, the system's performance ratio improves as more substances are added. Additionally, the curves indicate that the system scales well to a

large number of diagnoses since the ratios steadily increase as the available diagnoses increase. Another notable characteristic of the curves is that they indicate better performance in cases that are more severe. The primary reason is that more severe cases generally have more associated CEs. With more CEs, the system has more information to properly differentiate between various diagnoses, yielding a higher accuracy. The good news is that the most important cases are the most severe ones, and this is precisely where the system performs best.

6. Discussion

6.1. Typical System Performance

The first three columns of Table 4 exhibit typical system performance data. These representative accuracies use MC = 10 to ensure that outliers do not excessively influence the results and discard the MCE filter since it has little to no effect on system performance. To enable comparison between the various forms of diagnosis, the percentage of exposures appearing in the top 10% of trained diagnoses is used as the accuracy calculation. The results reiterate the fact that the system performs better on more severe cases. Once again, death is the exception due to limitations in the data and system failures in the body leading to cascading CEs. Moreover, the difficulties associated with the cases involving death make it fruitless to discuss trends for that severity. For major and moderate severities, diagnosing by substance performs best, followed by major and minor categories and finally by major category alone. The converse is true for minor severity cases, where diagnosing by major category alone performs best. Though not universally observed in the test runs, this accuracy inversion is not uncommon and is most likely due to the lack of CEs in minor severity cases. With minimal CEs it is easier to classify the general major category of a toxin than to identify the specific toxin involved.

The accuracy calculations in the first three columns of Table 4 show a high value of 79.8%, which occurs when diagnosing major severity cases by substance. These accuracy calculations include a large number of cases involving only a single CE, which would be difficult for even the most experienced expert to diagnose without additional information. To better demonstrate system functionality, the accuracies from the first three columns of Table 4 are recalculated in the last three columns to include only cases with at least three recorded CEs. A large improvement in system accuracy is observed, particularly in minor severity cases where accuracies are boosted from the range of 67.4% to 68.6% into the 74.0% to 75.1% range. Additionally, the accuracy of diagnosing major severity cases by substance and by major and minor categories is raised above 80%. Further system improvements could be achieved by removing categories that do not add value, such as the "unknown drug" diagnosis, and consolidating nearly redundant substances, such as "aspirin: pediatric formulation," "aspirin: unknown if adult or pediatric formulation," and "aspirin: adult formulation." However, one purpose of the research presented here is to bypass the need for expert input when generating a system and making such improvements would assume knowledge of the domain.

6.2. System Training and Response Time

An important aspect of system usability is the amount of processing time required to train the system and the response time of the user interface to diagnostic queries. System calculations were intentionally kept simple to enable scalability, rapid system generation, and a low response time. For research purposes, the system was developed using Microsoft Access 2002 on a Compaq Presario 2100 laptop with a 2.4GHz processor and 320MB of RAM. Training the system on four years of data required less than three minutes. Running a diagnosis under worst case conditions takes approximately three seconds when the program is first queried. Once

loaded into RAM, however, the diagnosis runtime is cut in half. Obviously, porting the program to the dedicated SQL server used by the FPIC would offer further speed improvements.

As the number of cases in the system increases, system training time could increase significantly, though there should be a minimal impact on diagnosis time due to the architecture of the system. Rather than beginning anew each update, the system can maintain key information about current values and incorporate the information from the latest cases into its calculations. Although incremental updates have not been implemented because the system is not directly linked to the central database, the use of LRs makes their implementation straightforward. To calculate LRs, a count of true positives, true negatives, false positives, and false negatives must be determined for each CE. By saving a table of these four values with their corresponding substance, updating the system simply involves querying the new data for a count of each of the four values, adding the results to the old table, and recalculating the LRs. See [51] for a comparison of three common incremental update algorithms and [7] for a discussion sufficient statistics that increase update efficiency.

6.3. Limitations

Although the FPIC database is a valuable resource, the data it contains include biases. Obviously, since the database only includes information on toxic exposures, a system trained on such data is only capable of suggesting diagnoses assuming a toxic exposure has occurred. The system is completely incapable of recognizing or speculating about non-toxicological disorders that may cause identical CEs. Additionally, substance abuse and illicit drug use often go unreported, while abusers who are hospitalized may lie in an attempt to conceal their activities. Furthermore, physicians and nurses may not fully recount all the important details of a case when reporting to the PCC or even intentionally omit CEs they deem unimportant. In spite of

anomalies inherent to the FPIC database, the accuracies achieved by the system give credence to the practical usefulness of the data and the addition of data to the case base may help dilute some biases in the data.

7. Conclusion

This paper presents research performed to create a prototype KBS for diagnosing toxic exposures, which will be expanded to include multiple exposures as the research progresses. A major goal was to bypass the knowledge acquisition bottleneck of traditional KBSs by using data mining to automatically generate the system. Because system generation assumes no knowledge about the field of toxicology, lower accuracy percentages are to be expected; however, future research can build on this foundation and intelligently modify substance groupings to improve performance. Additionally, expanding the system's case base to include data from the other PCC's in Florida or the AAPCC national database should lead to improved accuracies. Another important aspect of the system is the use of adjusted LRs. LRs are mathematical calculations that are commonly known and used throughout the medical field. In this research, traditional LRs are adjusted by adding a fractional possibility to every potential outcome. The result is a robust equation that mitigates multiply-by-zero and divide-by-zero errors while rapidly converging to the same value as traditional LRs. Ultimately, the system is intended to serve as a diagnostic consultant by providing differential diagnoses for toxic exposure cases based on observed CEs. The system enables physicians to tap into the knowledge stored in PCC databases, giving decision support information in a simple, understandable format.

This paper focuses on the development of a system for diagnosing single exposure cases. The research explored the effects of two different filters for refining diagnosis based on a minimum number of exposure cases and a minimum number of CEs. System accuracy reached

as high as 79.8% and increased above 80% when test cases were required to involve more than one CE. Furthermore, system testing performed by toxicologists within the FPIC yielded a positive response.

The research performed on this system offers a number of contributions to both the field of KBSs and medicine. First, being automatically generated, the system bypasses the knowledge acquisition bottleneck of traditional KBSs. A second contribution is the application of intelligent systems to the field of toxicology. At the present time, no American diagnostic systems exist for the field of clinical toxicology. Although systems have been implemented for France, Bulgaria, and Russia, they use different methods and are not readily available to assist American physicians. Finally, the use of LRs serves to bridge the gap between intelligent systems and the medical field. Too often, intelligent systems fail because they use methods that are unknown and distrusted by the medical community. The adjusted LR utilizes mathematics commonly accepted in medicine with a slight modification that creates a robust calculation without losing the essence of the original equation.

Acknowledgements

The authors thank Dausear "Dar" McRae and the Florida Poison Information Center in Jacksonville for their willingness to provide data, technical support, and consultation.

References

[1] A. Bronstein, D. Spyker, L. Cantilena, J. Green, B. Rumack, and S. Heard, 2007 annual report of the American Association of Poison Control Centers' National Poison Data System (NPDS): 25th annual report, *Clinical Toxicology 46* (2008) 927-1057.

[2] J. D. Schipper, J. L. Schauben, D. D. Dankel, A. A. Arroyo, and D. R. Sollee, Differential toxicological diagnoses using a computerized knowledge-based model, *Clinical Toxicology 46* (2008) 612.

[3] J. D. Schipper, J. L. Schauben, D. D. Dankel, A. A. Arroyo, and D. R. Sollee, A knowledge-based consultant for human toxic exposures, *Clinical Toxicology 47* (2009), 465-466.

[4] M. Holsheimer, M. Kersten, H. Mannila, and H. Toivonen, A perspective on databases and data mining, in: U. Fayyad and R. Uthurusamy, eds., *Proceedings of the 1st International Conference on Knowledge Discovery and Data Mining* (AAAI Press, Menlo Park, CA, 1995) 150-155.

[5] R. Valdes-Perez, Principles of human-computer collaboration for knowledge discovery in science, *Artificial Intelligence 107* (1999) 335-346.

[6] P. Bradley, U. Fayyad, and C. Reina, Scaling clustering algorithms to large databases, in: R. Agrawal and P. Stolorz, eds., *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining* (AAAI Press, Menlo Park, CA, 1998) 9-15.

[7] G. Graefe, U. Fayyad, and S. Chaudhuri, On the efficient gathering of sufficient statistics for classification from large SQL databases, in: R. Agrawal and P. Stolorz, eds., *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining* (AAAI Press, Menlo Park, CA, 1998) 204-208.

[8] M. Atzmueller, J. Baumeister, and F. Puppe, Inductive learning of simple diagnostic scores, in: P. Perner, R. Brause, H.-G. Holzhutter, eds., *Medical Data Analysis: 4th International Symposium* (Springer, Berlin, 2003) 23-30.

[9] K. Cios and G. Moore, Uniqueness of medical data mining, *Artificial Intelligence in Medicine 26* (2002) 1-24.

[10] N. Lavrac, Selected techniques for data mining in medicine, *Artificial Intelligence in Medicine 16* (1999) 3-23.

[11] E. Feigenbaum, Themes and case studies of knowledge engineering, in: D. Michie, ed., *Expert Systems in the Micro-Electronic Age* (Edinburgh University Press, Edinburgh, Scotland, 1979) 3-25.

[12] I. Kononenko, I. Bratko, and M. Kukar, Application of machine learning to medical diagnosis, in: R. Michalski, I. Bratko, and M. Kubat, eds., *Machine Learning and Data Mining: Methods and Applications* (John Wiley & Sons, Inc., New York, NY, 1998) 389-428.

[13] B. Buchanan and E. Shortliffe, Uncertainty and evidential support, in: B. Buchanan and E. Shortliffe, eds., *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (Addison-Wesley Publishing Company, Reading, MA, 1984) 209-232.

[14] V. Yu, L. Fagan, S. Bennett, W. Clancey, A. Scott, J. Hannigan, B. Buchanan, and S. Cohn, An Evaluation of MYCIN's advice, in: B. Buchanan and E. Shortliffe, eds., *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (Addison-Wesley Publishing Company, Reading, MA, 1984) 589-596.

[15] H. Pople, The formation of composite hypotheses in diagnostic problem solving: An exercise in synthetic reasoning, in: *Proceedings of the 5th International Joint Conference on Artificial Intelligence* (Morgan Kaufmann Publishers Inc., San Francisco, 1977) 1030-1037.

[16] H. Pople, CADUCEUS: An experimental expert system for medical diagnosis, in: P. Winston and K. Prendergast, eds., *The AI Business: The Commercial Uses of Artificial Intelligence* (The MIT Press, Cambridge, MA, 1985) 67-80.

[17] H. Pople, Evolution of an expert system: From Internist to Caduceus, in: I. De Lotto and M. Stefanelli, eds., *Proceedings of the International Conference on Artificial Intelligence in Medicine* (North-Holland Publishing Company, Amsterdam, 1985) 179-208.

[18] R. Miller, H. Pople, and J. Myers, INTERNIST-I, an experimental computer-based diagnostic consultant for general internal medicine, *The New England Journal of Medicine 8* (1982) 468-476.

[19] P. Koton, Reasoning about evidence in causal explanations, in: T. Mitchell and R. Smith, chairmen, *Proceedings of the 7th National Conference on Artificial Intelligence* (AAAI Press, Menlo Park, CA, 1988) 256-261.

[20] K. Althoff, R. Bergmann, S. Wess, M. Manago, E. Auriol, O. Larichev, A. Bolotov, Y. Zhuravlev, and S. Gurov, Case-based reasoning for medical decision support tasks: The Inreca approach, *Artificial Intelligence in Medicine 12* (1998) 25-41.

[21] E. Golobardes, X. Llora, M. Salamo, and J. Marti, Computer aided diagnosis with case-based reasoning and genetic algorithms, *Knowledge-Based Systems 15* (2002) 45-52.

[22] S. Abidi and S. Manickam, Leveraging XML-based electronic medical records to extract experiential clinical knowledge: An automated approach to generate cases for medical case-based reasoning systems, *International Journal of Medical Informatics 68* (2002) 187-203.

[23] J. Baumeister, M. Atzmueller, and F. Puppe, Inductive learning for case-based diagnosis with multiple faults, in: S. Craw and A. Preece, eds., *Advances in Case-Based Reasoning: 6th European Conference* (Springer, Berlin, 2002) 28-42.

[24] M. Atzmueller, J. Baumeister, and F. Puppe, Evaluation of two strategies for case-based diagnosis handling multiple faults, in: M. Nick and K.-D. Althoff, eds., *Proceedings of the 2nd German Workshop on Experience Management* (CEUR Workshop Proceedings, Aachen, Germany, 2003).

[25] M. Atzmueller, J. Baumeister, F. Puppe, W. Shi, and J. Barnden, Case-based approaches for diagnosing multiple disorders, in: V. Barr and Z. Markov, eds., *Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference* (AAAI Press, Menlo Park, CA, 2004) 154-159.

[26] J. Baumeister, D. Seipel, and F. Puppe, Incremental development of diagnostic set-covering models with therapy effects, in: G. Kern-Isberner, T. Lukasiewicz, and E. Weydert, eds., *Proceedings of the KI-2001 Workshop on Uncertainty in Artificial Intelligence* (Linkping University Electronic Press, Linkoping, Sweden, 2002).

[27] A. Onisko, M. Druzdzel, and H. Wasyluk, Extension of the HEPAR II model to multiple-disorder diagnosis, in: M. Klopotek, M. Michalewicz, and S. Wierzchon, eds., *Intelligent Information Systems: Proceedings of the IIS'2000 Symposium* (Physica-Verlag, Heidelberg, 2000) 303-313.

[28] A. Onisko, M. Druzdzel, and H. Wasyluk, Learning Bayesian network parameters from small data sets: Application of Noisy-OR gates, *International Journal of Approximate Reasoning 27* (2001) 165-182.

[29] F. Diez, J. Mira, E. Iturralde, and S. Zubillaga, DIAVAL, a Bayesian expert system for echocardiography, *Artificial Intelligence in Medicine 10* (1997) 59-73.

[30] M. Delgado, D. Sanchez, M. Martin-Bautista, and M. Vila, Mining association rules with improved semantics in medical databases, *Artificial Intelligence in Medicine 21* (2001) 241-245.

[31] K. Boegl, K.-P. Adlassnig, Y. Hayashi, T. Rothenfluh, and H. Leitich, Knowledge acquisition in the fuzzy knowledge representation framework of a medical consultation system, *Artificial Intelligence in Medicine 30* (2004) 1-26.

[32] A. Kusiak, J. Kern, K. Kernstine, and B. Tseng, Autonomous-decision making: A data mining approach, *IEEE Transactions on Information Technology in Biomedicine 4* (2000) 274-284.

[33] A. Kusiak, I. Law, and M. Dick, The G-algorithm for extraction of robust decision rules—Children's postoperative intra-atrial arrhythmia case study, *IEEE Transactions on Information Technology in Biomedicine 5* (2001) 225-235.

[34] S. Tsumoto, Automated discovery of positive and negative knowledge in clinical databases, *IEEE Engineering in Medicine and Biology 19* (2000) 56-62.

[35] S. Vinterbo and L. Ohno-Machado, A genetic algorithm approach to multi-disorder diagnosis, *Artificial Intelligence in Medicine 18* (2000) 117-132.

[36] Y. Hayashi, Neural expert system using fuzzy teaching input and its application to medical diagnosis, *Information Sciences - Applications 1* (1994) 47-58.

[37] W. Muster, A. Breidenbach, H. Fischer, S. Kirchner, L. Muller, and A. Pahler, Computational toxicology in drug development, *Drug Discovery Today 13* (2008) 303-310.

[38] T. W. Schultz, M. Cronin, and T. Netzeva, The present status of QSAR in toxicology, *Journal of Molecular Structure: THEOCHEM 622* (2003) 23-38.

[39] N. Greene, Computer systems for the prediction of toxicity: an update, *Advanced Drug Delivery Reviews 54* (2002) 417-431.

[40] E. Benfenati and G. Gini, Computational predictive programs (expert systems) in toxicology, *Toxicology 119* (1997) 213-225.

[41] S. Darmoni, P. Massari, J. Droy, T. Blanc, and J. Leroy, Functional evaluation of Seth: An expert system in clinical toxicology, in: P. Barahona, M. Stefanelli, and J. Wyatt, eds., *Artificial Intelligence in Medicine: 5th Conference on Artificial Intelligence in Medicine Europe* (Springer, Berlin, 1995) 231-238.

[42] S. Darmoni, P. Massari, J. Droy, N. Mahe, T. Blanc, E. Moirot, and J. Leroy, SETH: An expert system for the management on acute drug poisoning in adults, *Computer Methods and Programs in Biomedicine 43* (1994) 171-176.

[43] A. Monov, I. Iordanova, P. Zagorchev, V. Vassilev, M. Nissimov, R. Kojuharov, R. Tconev, and V. Damianov, MEDICOTOX CONSILIUM – An expert system in clinical toxicology, in: K. Lun, P. Degoulet, T. Piemme, and O. Rienhoff, eds., *Proceedings of the 7th World Congress on Medical Informatics* (North-Holland Publishing Company, Amsterdam, 1992) 610-614.

[44] Florida Poison Information Center Network (2011, Jan.). FPIN statewide annual reports, calendar year (Jan-Dec) 2008: General call summary report. [Online]. Available: http://fpicn.jax.ufl.edu/Data/Reports/Calls_State_2008.pdf (Accessed: 14 October 2011).

[45] R. Duda, P. Hart, and D. Stork, Bayesian Decision Theory, *Pattern Classification*, 2nd ed. (John Wiley & Sons, Inc., New York, NY, 2001).

[46] D. Owens and H. Sox, Medical decision-making: Probabilistic medical reasoning, in: E. Shortliffe and L. Perreault, eds., *Medical Informatics: Computer Applications in Health Care and Biomedicine* (Springer-Verlag, New York, NY 2001) 76-131.

[47] R. Maxion and T. Townsend, Masquerade Detection Using Truncated Command Lines, in: *Proceedings of the International Conference on Dependable Systems and Networks* (IEEE Computer Society, Los Alamitos, CA, 2002) 219-228.

[48] D. Meretakis and B. Wuthrich, Extending Naive Bayes Classifiers Using Long Itemsets, in: U. Fayyad, S. Chaudhuri, D. Madigan, chairmen, *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York, 1999) 165-174.

[49] H. Jeffreys, *Theory of Probability*, 2nd ed. (The Claredon Press, Oxford, 1948).

[50] P. Huber, Robust Estimation of a Location Parameter, *The Annals of Mathematical Statistics 35* (1964) 73-101.

[51] M. Zhang, B. Kao, and C. Yip, A comparison study on algorithms for incremental update of frequent sequences, in: V. Kumar, S. Tsumoto, N. Zhong, P. Yu, and X. Wu, eds., *Proceedings of the 2002 IEEE International Conference on Data Mining* (IEEE Computer Society, Los Alamitos, CA, 2002) 554-561.

RESULTS_SS_ADJ : Table

LR	SubDesc
148.949829478	BACTERIAL FOOD POISONING: UNKNOWN TYPE
3.32482311621	UNKNOWN MUSHROOM
0.31018510633	MULTI-BOTANICAL WITHOUT MA HUAN OR CITRUS AURANTIUM
0.29315333996	CARDIAC GLYCOSIDE
0.18296274892	LITHIUM
0.13801692951	ORGANOPHOSPHATE
0.11248568761	SUSPECTED FOOD POISONING-UNKNOWN TYPE-PATIENT SYMPTOMATIC
0.06391201604	ACETAMINOPHEN WITH HYDROCODONE
0.06183397575	MULTI VITAMIN-TABLET: CHILD WITH IRON (NO FLUORIDE)
0.05855603720	LAXATIVE

Record: 1 of 265

Figure 1 The derived differential diagnosis for a toxic exposure inducing the gastrointestinal CEs of abdominal pain, dehydration, and diarrhea.

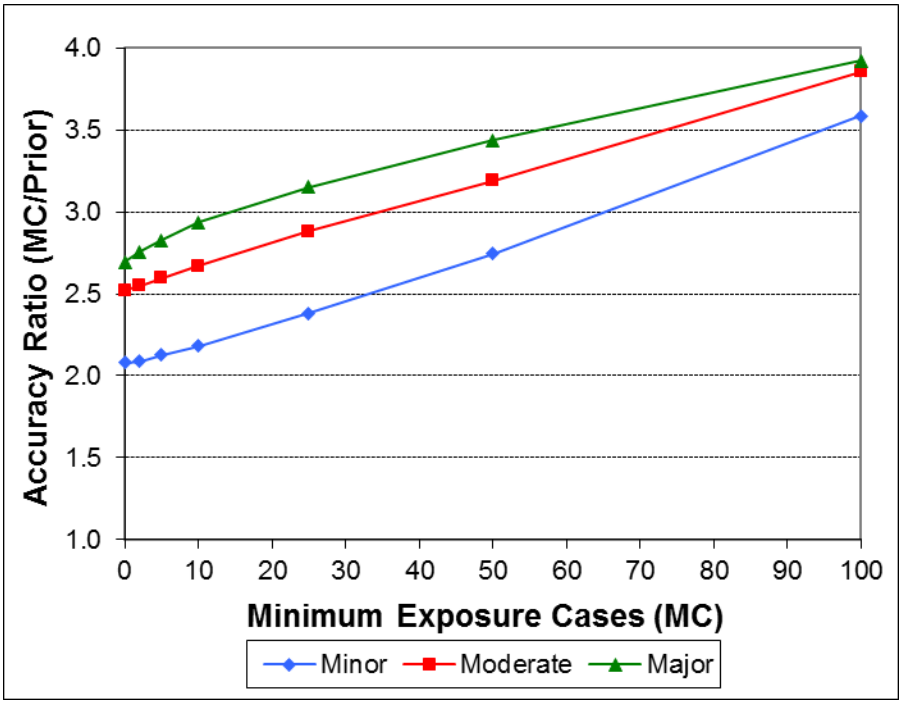

Figure 2 Accuracy ratios by substance.

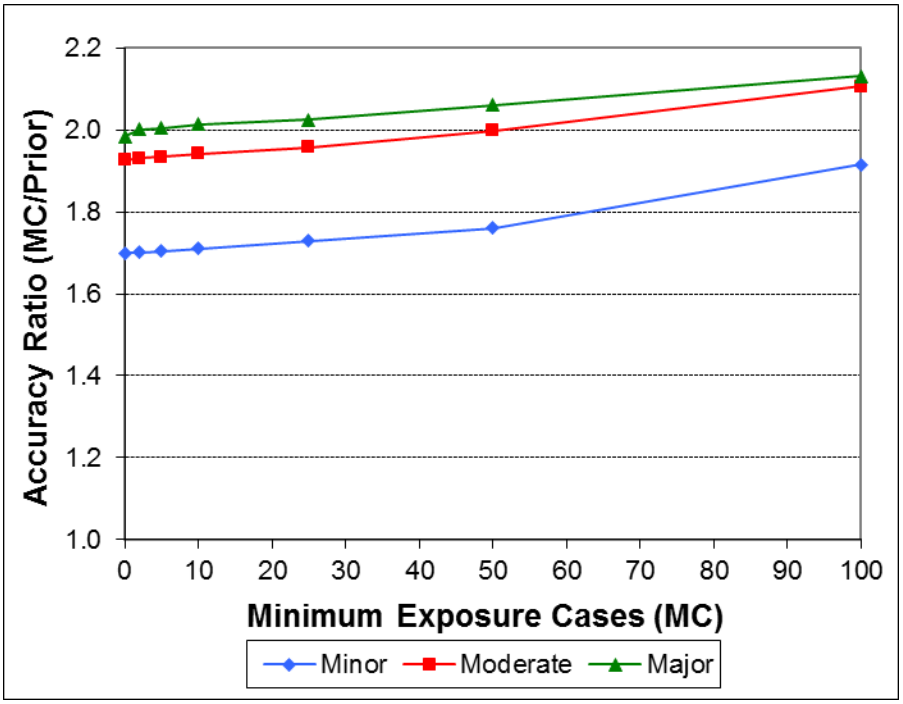

Figure 3 Accuracy ratios by major and minor categories.

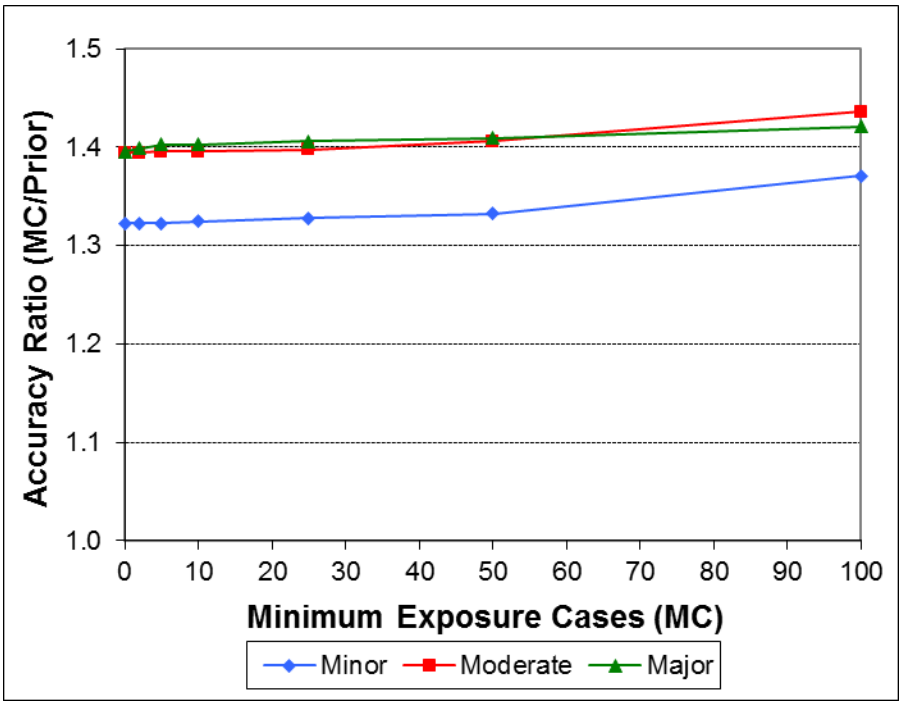

Figure 4 Accuracy ratios by major category.

Table 1

Accuracy of the system diagnosing exposures in the top 10% for various MCE when diagnosing by substance, major and minor categories, and major category alone with MC = 25.

Diagnosis by		Substance Minimum CE Occurrences (MCE)					Major & Minor Categories Minimum CE Occurrences (MCE)					Major Category Minimum CE Occurrences (MCE)				
		0	2	5	10	50	0	2	5	10	50	0	2	5	10	50
Severity	Minor	64.7%	64.7%	64.7%	64.7%	64.7%	64.1%	64.1%	64.1%	64.1%	63.9%	63.9%	63.9%	63.8%	63.8%	63.8%
	Moderate	74.4%	74.4%	74.4%	74.4%	74.2%	72.7%	72.7%	72.7%	72.6%	72.4%	70.0%	70.0%	69.9%	69.9%	69.8%
	Major	77.8%	77.8%	77.7%	77.7%	77.6%	75.4%	75.4%	75.4%	75.1%	75.4%	70.5%	70.5%	70.4%	70.0%	70.5%
	Death	62.2%	62.2%	62.2%	62.2%	58.1%	58.2%	58.2%	58.2%	58.2%	63.3%	55.7%	55.7%	54.4%	54.4%	54.4%

Table 2

Accuracy of the system diagnosing exposures in the top 10% for various MC when diagnosing by substance, major and minor categories, and major category alone with MCE = 10.

Diagnosis by		Substance Minimum Exposure Cases (MC)							Major & Minor Categories Minimum Exposure Cases (MC)							Major Category Minimum Exposure Cases (MC)						
		0	2	5	10	25	50	100	0	2	5	10	25	50	100	0	2	5	10	25	50	100
Severity	Minor	74.4%	72.9%	69.6%	67.4%	64.7%	62.9%	58.9%	71.7%	70.1%	68.9%	67.5%	64.1%	63.0%	58.4%	68.5%	68.5%	68.5%	68.6%	63.8%	64.1%	60.3%
	Moderate	80.3%	79.7%	78.0%	76.6%	74.4%	71.6%	64.8%	79.1%	78.0%	76.9%	75.6%	72.6%	71.8%	66.7%	73.4%	73.4%	73.4%	73.5%	69.9%	70.2%	66.4%
	Major	80.6%	81.7%	81.3%	79.8%	77.7%	75.1%	68.6%	81.1%	80.3%	79.7%	78.5%	75.1%	74.2%	69.2%	73.9%	73.9%	74.0%	74.0%	70.0%	70.3%	65.2%
	Death	62.0%	65.8%	63.3%	62.8%	62.2%	61.2%	46.8%	67.1%	68.4%	68.4%	67.1%	58.2%	59.2%	49.3%	58.2%	58.2%	58.2%	58.2%	54.4%	56.4%	51.3%

Table 3

Accuracy of the system diagnosing exposures in the top ten diagnoses for various MC when diagnosing by substance, major and minor categories, and major category alone with MCE = 10.

Diagnosis by		Substance Minimum Exposure Cases (MC)							Major & Minor Categories Minimum Exposure Cases (MC)							Major Category Minimum Exposure Cases (MC)						
		0	2	5	10	25	50	100	0	2	5	10	25	50	100	0	2	5	10	25	50	100
Severity	Minor	41.2%	41.4%	42.1%	43.2%	47.2%	54.4%	71.0%	63.0%	63.1%	63.2%	63.4%	64.1%	65.3%	71.0%	79.5%	79.5%	79.5%	79.6%	79.8%	80.1%	82.4%
	Moderate	50.0%	50.5%	51.4%	52.9%	57.1%	63.2%	76.4%	71.4%	71.6%	71.7%	72.0%	72.6%	74.1%	78.1%	83.8%	83.8%	83.9%	83.9%	84.0%	84.5%	86.3%
	Major	53.4%	54.6%	56.0%	58.2%	62.4%	68.1%	77.7%	73.6%	74.2%	74.3%	74.7%	75.1%	76.4%	79.0%	83.9%	84.1%	84.3%	84.3%	84.5%	84.7%	85.4%
	Death	35.4%	39.2%	41.8%	44.9%	45.9%	50.7%	57.4%	58.2%	58.2%	58.2%	58.2%	58.2%	60.5%	61.3%	69.6%	70.9%	72.2%	72.2%	72.2%	71.8%	71.8%

Table 4

Accuracy of the system diagnosing exposures in the top 10% when diagnosing by substance, major and minor categories, and major category alone with MC =10 and MCE = 0.

Trained on		All Exposure Cases			Cases with 3 or more Clinical Effects		
Diagnosis by		Substance	Major & Minor Categories	Major Category	Substance	Major & Minor Categories	Major Category
Severity	Minor	67.4%	67.5%	68.6%	75.1%	74.6%	74.0%
	Moderate	76.6%	75.7%	73.5%	78.4%	77.2%	75.0%
	Major	79.8%	78.9%	74.8%	81.0%	80.5%	75.9%
	Death	62.8%	67.1%	59.5%	69.4%	71.4%	66.7%