Identifying combinatorial biomarkers by association rule mining in the CAMD Alzheimer's database

Balázs Szalkai[a], Vince K. Grolmusz[c], Vince I. Grolmusz[a,b,*], Coalition Against Major Diseases[1]

[a] PIT Bioinformatics Group, Eötvös University, H-1117 Budapest, Hungary
[b] Uratim Ltd., H-1118 Budapest, Hungary
[c] 2nd Department of Internal Medicine, Semmelweis University, Budapest, Hungary.

ARTICLE INFO

ABSTRACT

The concept of combinatorial biomarkers was conceived when it was noticed that simple biomarkers are often inadequate for recognizing and characterizing complex diseases. Here we present an algorithmic search method for complex biomarkers which may predict or indicate Alzheimer's disease (AD) and other kinds of dementia. We show that our method is *universal* since it can describe any Boolean function for biomarker discovery. We applied data mining techniques that are capable to uncover implication-like logical schemes with detailed quality scoring. The new SCARF program was applied for the Tucson, Arizona based Critical Path Institute's CAMD database, containing laboratory and cognitive test data for 5821 patients from the placebo arm of clinical trials of large pharmaceutical companies, and consequently, the data is much more reliable than numerous other databases for dementia. The results of our study on this larger than 5800-patient cohort suggest beneficial effects of high B12 vitamin level, negative effects of high sodium levels or high AST (aspartate aminotransferase) liver enzyme levels to cognition. As an example for a more complex and quite surprising rule: Low or normal blood glucose level with either low cholesterol or high serum sodium would also increase the probability of bad cognition with a 3.7 multiplier. The source code of the new SCARF program is publicly available at http://pitgroup.org/static/scarf.zip.

1. Introduction

Dementia is a major problem presently of high-income countries and also an increasing concern of low-income nations worldwide. Though sporadic before the age of 60, its occurrence is doubled by every five years of age thereafter (Bermejo-Pareja et al., 2008; Carlo et al., 2002). About 40 percent of the population over 90 are affected, and up to 20 percent of those between 75 and 84 suffer from this condition (Prince & Jackson, 2009; Wortmann, 2012). The most common cause of dementia is Alzheimer's disease (AD). The earliest symptoms of AD include memory problems; disorientation in time and space; and difficulty with calculation, language, concentration and judgment. As the disease evolves, patients may develop severe behavioral abnormalities and may even become psychotic. In the final stages of the disease the sufferers are incapable of self-care and become bed-bound, for years or even decades.

The diagnosis of AD in the great majority of the cases is done by clinical criteria, using standardized questionnaires (Mossello et al., 2010). Generally accepted evidences show that neuropathological damage begins more than 20 years before those clinical signs (Jack et al., 2009), and by the time it is diagnosed, a large part of the neurons are already irreversibly lost.

In the last years, by the combination of cerebrospinal fluid analysis, clinical signs and neuroimaging techniques a quite reliable diagnostic method emerged (Dubois et al., 2007). The method, however, is prohibitively expensive, is not an early warning-type biomarker, and does not seem to be applicable for wide-scale screening of the senior population.

Very recently, using the combination of usual clinical laboratory data, cognitive impairment questionnaires and blood-based proteomics assays was reported to reliably diagnose AD, without neuroimaging or cerebrospinal fluid assays (O'Bryant et al., 2010, 2011). However, early warning biomarkers are still need to be found.

The final goal of ours is finding new combinatorial biomarkers for Alzheimer's disease. In this paper we report our results that may be used to reach this final goal; but presently we are able to show only that

certain sets of laboratory data may make dementia (and not AD) more probable, and certain other sets may make dementia less probable.

There are several large databases on Alzheimer's disease available for researchers. The quality of their data obviously depends on the methodology of the research that produced the database in question. The most well-organized, strictly overseen and rigorously documented experiments are perhaps conducted by the order of large pharmaceutical companies in hospitals and clinics in phase 1, 2 and 3 drug trials. Unfortunately, the detailed results of those trials are seldom published (especially those corresponding to unsuccessful drug trials) since they are owned by the companies that ordered the trials.

In their Alzheimer's disease database the Tucson, Arizona based Critical Path Institute made available the results of the placebo arm of numerous multi-million dollar clinical trials conducted by the order of large pharmacological companies (Rogers et al., 2012; Romero et al., 2009, 2010). The data of the placebo line of the trials does not contain proprietary information concerning the effects of the novel drugs under trial, but it does contain reliable, well-organized laboratory and cognitive test-data, presumably in much higher quality than other, larger, but perhaps less strictly conducted and controlled studies for AD.

Data used in the preparation of this article have been obtained from the Coalition Against Major Diseases (CAMD) database (Romero et al., 2009). In 2008, Critical Path Institute, in collaboration with the Engelberg Center for Health Care Reform at the Brookings Institution, formed the Coalition Against Major Diseases (CAMD). The Coalition brings together patient groups, biopharmaceutical companies, and scientists from academia, the US Food and Drug Administration (FDA), the European Medicines Agency (EMA), the National Institute of Neurological Disorders and Stroke (NINDS), and the National Institute on Aging (NIA). Coalition Against Major Diseases (CAMD) includes over 200 scientists from member and non-member organizations. The data available in the CAMD database have been volunteered by both CAMD member companies and non-member organizations.

The data of the CAMD database were anonymized by an HIPAA Safe Harbor compliant method (Romero et al., 2009).

In contrast with more conservative statistical methods, we applied data mining techniques for data analysis and combinatorial biomarker search. Data mining, as defined in Hand, Mannila, and Smyth (2001), is the analysis of large observational sets of data for finding new, still unsuspected relations with novel, usually high-throughput methods. Frequently, data mining uses large data sets originally collected for uses other than the data mining analysis (Hand et al., 2001).

Simple biomarkers (e.g., the high level of glucose in diabetes) show a physiological condition, related to the appearance or the status of a disease. The concept of combinatorial biomarkers appeared around 2010, and numerous authors simply use the term in the following sense: If—say—the high concentration of all the molecules *A*, *B* and *C* characterizes a certain condition *X* well (and the high concentration of any subset of the set {*A*, *B*, *C*} would not), then they say that {*A*, *B*, *C*} is a combinatorial biomarker of the condition *X* (Wu et al., 2012). In O'Bryant et al. (2011), by applying proteomics assays, a 30-protein set was identified as a combinatorial biomarker of AD.

We intend to discover more involved combinatorial biomarkers that may contain clinical laboratory data and psychiatric test data. We describe a *universal* method in the following sense:

Suppose binary variables $x_1, x_2, ..., x_n$ correspond to Boolean clinical or laboratory data; that is, $x_i = 1$ if the value of a measurement, denoted by M_i, is in a given interval, and $x_i = 0$ if the value of M_i is outside of the given interval. Now, a Boolean function $f(x_1, x_2, ..., x_n)$ that takes values either 0 or 1, can describe *any* functional relation between the set of the clinical or laboratory data that corresponds to variables $x_1, x_2, ..., x_n$ and the presence or absence of any clinical or biological condition denoted by A: $f(x_1, x_2, ..., x_n) = 1$ corresponds to the presence of the clinical/biological condition A, and $f(x_1, x_2, ..., x_n) = 0$ corresponds to the absence of the clinical/biological condition A.

Since *every* Boolean function f could be given by logical conjunctions (ANDs) and logical disjunctions (ORs) of variables $x_1, x_2, ..., x_n$ or their negations, our method SCARF is universally able to find any such conjunctions or disjunctions. That is the novelty of our method.

Note that applying more than one binary x_i variable for the same measurement M_i, one can describe the distribution of the values of M into more than one intervals, therefore our method is applicable for quantities described by non-Boolean variables as well.

We remark that simple biomarkers relate a single—not necessarily Boolean—variable x_1 to the presence or absence of condition A. In combinatorial biomarkers, like in O'Bryant et al. (2011), a 30-protein set was identified as a combinatorial biomarker of AD, that is, f was just the conjunction of 30 variables x_1 through x_{30}, corresponding to the presence of 30 proteins in measurable concentrations. Our SCARF program is capable to discover any combinations of conjunctions and disjunctions and negations, therefore, it can describe an arbitrary Boolean function on variables $x_1, x_2, ..., x_n$.

We make use of an adequately modified association rule mining (Hand et al., 2001) procedure, and we also apply a new methodology that discovers complex combinatorial biomarkers only if these complex biomarkers have stronger implications than the simpler biomarkers. Therefore, our program SCARF will not produce artificially complex biomarkers just for the sake of complexity: the more complex is the new biomarker, the more valid is the new implication.

1.1. Association rule mining

Our research group was among the firsts applying association rule mining in molecular biology (Ivan, Szabadka, & Grolmusz, 2007). Recently, association rule mining has been gaining applications in drug discovery (Galustian & Dalgleish, 2010), in the design of clinical trials (Epstein, 2009), and most recently, also in image analysis in Alzheimer's research (Chaves et al., 2011).

Association rule mining is a field of data mining (Hand et al., 2001) developed by marketing experts for discovering implication-like rules in uncovering customer behavior (Agrawal, Imielinski, & Swami, 1993), without *a priori* assumptions on this behaviour.

The form of an association rule is as follows: $X \rightarrow Y$, where X and Y are disjoint subsets of a ground-set of attributes or "properties" S, and $X \rightarrow Y$ means that the set of attributes X implies the set of attributes Y. Both X and Y are just two disjoint lists of properties, no logical operators or clauses are allowed.

We intended to apply a generalization of this method for laboratory and cognitive test data from the CAMD database (Romero et al., 2009).

We analyzed how the presence/absence/severity of cognitive impairment could be detected from combinations of known biomarkers, demographic information and measurements of vital signs. As an example, consider this expression:

$$\text{sodium} = \text{high} \& (\text{protein} = \text{high or age} \geq 60) \rightarrow \text{mmset otal} \leq 15 \quad (1)$$

Here & stands for logical AND or conjunction. This rule states that if blood sodium is high, AND urine protein is high OR age is at least 60, then the total MMSE (Mini Mental State Examination) score will be at most 15 out of 30. Let us call the left-hand side of the expression (abbreviated by LHS) a combinatorial marker of the right-hand side (abbreviated by RHS). Thus the statement above can be reformulated as follows: high serum sodium combined with either high urine protein or age of at least 60 is a marker of a total MMSE score less than or equal to 15.

An expression consists of *elementary clauses* combined by logical operators. These elementary clauses may include equalities and inequalities. By substituting all elementary clauses with some wildcard, we can obtain the *pattern* of an expression. For example, the expression above is of the following pattern:

$$\square \; \& \; (\square \text{ or } \square) \rightarrow \square \quad (2)$$

During our analysis we started with a given pattern like the one above. Then we considered all the possible logical expressions according to this pattern, and assigned numerical values to them that indicated the reliability and validity of the logical rules. Then we filtered and sorted the vast amount of possible rules according to these numerical criteria, and selected the best ones. We changed a simpler rule to a more complex rule only if the more complex rule had higher reliability/validity than the simpler rule (see the next section for the exact definitions).

1.2. Comparison of SCARF with other methods

In Nanavati, Chitrapura, Joshi, and Krishnapuram (2001) a method is presented for mining generalized disjunctive association rules. In their work, the authors are searching for association rules of the form

$$\square \rightarrow (\square \text{ or } \square \dots \text{or } \square) \tag{3}$$

by using a heuristic approach: their algorithm does not necessarily return all the valid implications of the given form. In comparison,

- SCARF discovers association rules of the form (2) that is, we are interested in rules uncovering "what causes dementia" and not, as in Nanavati et al. (2001), "what is caused by dementia";
- If SCARF contained a disjunctive normal form (an or of & s, DNF) on the left hand side of (2), then the rule could easily be decomposed into several valid rules. This is the reason that SCARF mines CNFs (conjunctive normal forms, i.e., an & of "or" s).
- While it is possible to convert a rule of the form (3) to a rule of the form (2) by using logic operations, converting the numerical quality criteria (support, confidence, lift) do not seem to be possible with an algorithmically tractable method;
- SCARF is not a heuristic algorithm: it finds *every* rule of the given form, satisfying the numerical quality criteria; the algorithm described in Nanavati et al. (2001) is heuristic.

The scopes of logic regression analysis (e.g., Ruczinski, Kooperberg, & LeBlanc, 2003) and association rule mining are quite different. In logic regression, one intends to approximate a function (binomial- or multinomial) on its whole domain, and the quality of the approximation is measured by some distance-related measure on the whole domain. Therefore, logic regression analysis targets the ambitious goal of approximating, for example, the "demented" or "not-demented" multivariate function on the whole domain of study. In SCARF, we intend to find combinatorial biomarkers that if present, imply the demented status. Additionally, logic regression analysis never finds a biomarker that is true only on the 1% of the domain: the associated distance metric will be negligible. In association rule mining, however, it is possible to find a risk factor that holds only for 1% of the cohort, but there, strongly implies the dementia or other disease.

2. Materials and methods

Our data source, which will be referred to as CAMD from now on (Romero et al., 2009), was provided by the Coalition Against Major Diseases, and consisted of the placebo arm of several drug trials. Over 5800 subjects participated in these trials including demented and not demented people of various age and sex (see Table 1) for basic statistics). Standard laboratory data that have been collected for the subjects included about 300 different values in blood or urine altogether. These values were generally measured multiple times per subject (on different visit days), though each person was tested for only about 30 different values. The cognitive and psychological status of the subjects was measured at different times by standardized questionnaires ADAS-COG, ADCS-ADL, MMSE, NPI and SIB. In addition, some genetic tests have been performed, e.g., ApoE and MTHFR genotypes were recorded. Vital sign measurements (BP, pulse rate, respiratory rate and body

Table 1
Basic statistics on the subjects of the CAMD data.

Age distribution		Gender distribution		MMSE distribution	
A: up to 65 years	1093	Female	3315	A: severe cog. impairment	255
B: 66–75 years	2070	Male	2653	B: moderate cog. impairment	611
C: 76–85 years	2408			C: mild cog. impairment	3224
D: more than 85	397			D: normal cognition	1352

temperature) have also been taken. Results concerning this dataset will be described in greater detail below.

We transformed this large dataset into a conveniently processable form. The CAMD database contained several rows describing one person and these were scattered between multiple data tables. So we collected the essential data from CAMD into one single table: this simplified table contained only one row for each subject.

If a subject was tested on different visit days, then we took the average of these test results. The resulting main table for CAMD consisted of around 170 columns (record fields) and 5821 rows (entries).

Our main method of processing the resulting table was association rule mining. First, we took a given pattern like $\square \& (\square \text{ or } \square) \rightarrow \square$. Notice that the LHS (Left Hand Side) is in conjunctive normal form here (multiple OR clauses ANDed together). This pattern matches all statements of the following kind: "if property A is present and property B or property C is present, then property D is present".

Since we are interested in implication-like association rules that indicate factors implying normal or demented mental state, we made restrictions on which data columns can occur on the LHS (left hand side) and the RHS (right hand side). Laboratory data and sex were allowed on the LHS, and columns directly indicating mental status on the RHS. Then we gave numerical constraints on the "goodness" of a rule—thus introducing an ordering on the rules. Finally we tried to fill in all the void boxes in all possible ways to find the best rules.

If done without any optimization, this process would have yielded a vast amount of different rules that would have needed to be evaluated "by hand". Even just enumerating all the possible matches to this pattern would have required enormous computational resources. Consequently, we needed to make the computation feasible: we used a *branch-and-bound* approach similar to the Apriori Algorithm (Agrawal & Srikant, 1994): if certain values for the first two boxes made a rule fail our constraints—regardless of what would be written in the third box—, then we threw out the rule and did not bother checking all the possible values for the third box.

More exactly, SCARF uses a branch-and-bound algorithm for finding association rules. The LHS is filled from left to right. When a disjunctive clause is completed, we calculate the universe, support and LHS support of the rule. If one of these is already too small, we will not continue building the rule, but advance to the next value for the lastly filled elementary clause. Thus we eliminate the rest of that branch.

The possible rules can be represented as the leaves of a tree graph. The root is the blank pattern itself, each inner node is an incomplete rule (where only some blanks at the beginning are filled with equalities), and the children of a node are those—possibly incomplete—rules where one more blank is substituted with an equality.

The algorithm performs a depth-first search (DFS) on this tree. We start from the root node. In each step, if the current node has any unvisited children, then we descend to one of them in the tree. Otherwise, we ascend to the parent node. This is done until we have traversed the whole tree and arrived back to the root.

In our case, the DFS means that the blanks are filled with attribute-value pairs, one after the other, in a recursive manner. When we reach the last blank, the rule is evaluated, and another equality is tried for the

last blank, until we exhausted all the possibilities. After that, we ascend to the second last blank, and try another equality for that. Then again we descend to the last one, and try all equalities once again. This goes on until we visited all the nodes in the tree.

Branch-and-bound means that we need not visit all the leaves, but rather we are able to determine at some inner nodes that the corresponding subtree will not yield good enough rules. If this occurs, we decide to ascend immediately, without visiting the descendant nodes. This is done by calculating bounds for the parameters of descendant nodes. If those bounds imply that the parameters will not be good enough for the rules to appear in the top N list, then we do not traverse the subtree.

This technique saved a considerable computational time, and made possible this study on a 5821 cohort.

The association rule mining was done with our own new program written in the C++ programming language, named SCARF (Simple Combinatorial Association Rule Finder). We calculated various standard numerical values for all association rules, which would indicate their validity. First, we defined the *universe* of a rule: this is the set of the database rows where all columns present in the rule have a known value. As we mentioned before, not all subjects were tested for everything, so our database contained a large amount of N/A entries. For testing the validity of a rule, only those rows could be taken into account, where there is no N/A written to any of the columns participating in the rule.

For evaluating the validity of a rule, we continued to work with only its universe and temporarily discarded all other rows in the database. Next, we calculated the *LHS support*, *RHS support* and *support* of a rule. The *LHS support* is the number of the rows where the LHS is true, the *RHS support* is the number of the rows where the RHS is true, and the *support* is the number of the rows where both the LHS and the RHS are true.

Then, we calculated the *confidence*, *lift*, *leverage* and χ^2-statistic for a rule. The *confidence* is defined as the conditional probability of the RHS, assuming that the LHS is true. If one has high serum sodium combined with high urine protein or age at least 60 in our example, then confidence describes the chance of having a low MMSE score. The *lift* shows how many times the presence of the LHS increases the probability of RHS. Generally it indicates how big a risk factor the LHS is—though it is not certain that the LHS *causes* the RHS, as they both may be only consequences of some background phenomenon (Hand et al., 2001).

The *leverage* is the difference between the observed probability of both the LHS and RHS being true, and the estimated probability we get by assuming that the LHS and RHS are independent events. It indicates the level of dependency between the LHS and the RHS in a way. Finally, the χ^2-statistic is a well-known measure of the estimated dependence of the indicator variables of the LHS and RHS. The p-value output by SCARF comes from this χ^2 test.

The E-value (also calculated by SCARF) equals to the p-value multiplied by the total number of possible rules. The E-value is a more useful measure of randomness, since if we examine many rules, there is a high probability that the p-value will be small enough, while the E-value is insensitive for this kind of artifact.

The following table formalizes some of the above definitions. Here P denotes the probability measure, and $P(A|B)$ denotes the conditional probability of event A on condition B:

Confidence $= P(\text{RHS}|\text{LHS})$

Lift $= \dfrac{P(\text{RHS} \mid \text{LHS})}{P(\text{RHS})}$

Leverage $= P(\text{RHS} \wedge \text{LHS}) - P(\text{RHS})P(\text{LHS})$

For the CAMD database the acceptable values were set as follows: universe ≥ 500, support ≥ 50, confidence ≥ 0.5, lift ≥ 1.2, p-value ≤ 0.05. In particular, we recorded rules on data that were measured on at least 500 subjects. We defined the *goodness* of a rule to be

equal to its lift.

Therefore we listed association rules of lift at least 1.2, i.e., only those rules were listed where the LHS increased the probability of RHS with at least 20%.

One of the most significant novelties in our approach was filtering out those rules which were too complicated. The SCARF program threw out elementary clauses from the LHS as long as the overall goodness (i.e. the lift) of the rule did not decrease by more than 2%. Then it deleted the whole rule if its numerical values dropped below our constraints during the simplification process. In other words, we sacrificed some of the lift for simplicity, to avoid overfitting.

Having listed the best rules, we also determined whether the elementary clauses (like *lb _ a st = h*, *lb _ f olate = l*, etc.) have positive or negative effect on mental state. Therefore we counted their appearances on LHS, and classified these occurrences by the nature of the RHS: does it indicate normal cognition or rather dementia? We counted how many times an elementary clause occurred on the LHS of a rule when the RHS indicated a positive mental state, and how many times it occurred in rules where the RHS showed a negative state. Thus, in addition to mining rules whose LHS could probably serve as good combinatorial risk factor of dementia, we estimated the contribution of the *individual* clauses, for example "protein = *high*" to the onset of cognitive impairment.

For an elementary clause, *Positive score* was the number of rules with positive RHS, and *Negative score* was the number of rules with negative RHS. Then we compared *Positive score* with *Negative score*: by subtracting the negative score from the positive score we got a value called simply the *score* of the clause. Those elementary clauses whose score was positive were called *positive* clauses, and similarly, those where the score was negative were called *negative* clauses.

To summarize our method: we searched for combinatorial biomarkers using a branch-and-bound algorithm for association rule mining; then made statistical analysis regarding elementary clauses.

The source code of the new SCARF program is publicly available at http://pitgroup.org/static/scarf.zip.

3. Results

The program outputs 725 rules from the CAMD database. Selected rules, ordered by lift (i.e. "goodness") decreasing are listed in Table 2. The whole set of rules is presented as Table S1 of the online supporting material at http://uratim.com/CAMD.

On the LHS, clauses concerning biomarkers end in "= l", "= h", "= n", or combinations of these. Here l means low, h means high and n means normal. If there are multiple letters (such as *nh*), then the corresponding equality states that the value is either high or normal. In other words, single letters correspond to a value category, while multiple letters mean the union of these categories.

For example, the second rule in Table 2 was that of the second best lift. It can be interpreted in the following way: It is likely that if serum sodium level is elevated, and serum glucose level is either too low or normal, then the total MMSE score will be less than 15. Note that it is true for all rules of ours that there is not necessarily a causal relation between the LHS and RHS, as both the LHS and RHS can be consequences of an unknown process in the background.

The third rule states that "if serum sodium level is elevated, and calcium level is either low or normal, then MMSE orientation sub-score will be at most 2". The seventh rule in Table 2 states that "if serum sodium level is elevated, and body temperature is too low, then total MMSE score will be less than 15".

From these selected rules we can conclude that elevated sodium combined with various other factors (not too high glucose, not too high calcium, low temperature) might be a good indicator (or even the cause) of mental decline.

Elementary clauses with the greatest positive effect on normal cognition are listed in Table 5.

Table 2

Several association rules of the highest lift. The lift value describes the multiplication factor, increasing the probability of the right hand side (RHS) if the left hand side is true. For example, our best rule (the first below) is saying that one can have the a bad result of a cognitive test with four times higher probability if one has high serum sodium and either low cholesterol or low or normal blood glucose level. We refer to Table 4 for the legends applied in this table, and to subsection "Association Rule Mining" in the Introduction for the simple implication-like formalism that describes the association rules found.

```
(lb_sodium=h) & (lb_chol=l or lb_gluc=ln) ->mm_ori=B
Universe: 2783, LHS support: 87, RHS support: 401, Support: 50
Confidence: 0.574713, Lift: 3.98859, Leverage: 0.0134618, p-value:
    0, E-value: 0
3.98859

(lb_gluc=ln) & (lb_chol=l or lb_sodium=h) ->mm_ori=B
Universe: 2783, LHS support: 105, RHS support: 401, Support: 57
Confidence: 0.542857, Lift: 3.76751, Leverage: 0.0150451, p-value:
    0, E-value: 0
3.76751

(lb_sodium=h) & (lb_hct=l or lb_gluc=ln) ->mm_ori=B
Universe: 2926, LHS support: 95, RHS support: 420, Support: 51
Confidence: 0.536842, Lift: 3.74, Leverage: 0.0127695, p-value: 0,
    E-value: 0
3.74

(lb_sodium=h) & (bpsys=ln or lb_gluc=ln) ->mm_ori=B
Universe: 3091, LHS support: 102, RHS support: 425, Support: 52
Confidence: 0.509804, Lift: 3.70777, Leverage: 0.0122858, p-value:
    0, E-value: 0
3.70777

(lb_gluc=ln) & (lb_creat=l or lb_sodium=h) ->mm_ori=B
Universe: 3091, LHS support: 99, RHS support: 425, Support: 50
Confidence: 0.505051, Lift: 3.6732, Leverage: 0.0117722, p-value:
    0, E-value: 0
3.6732

(lb_sodium=h) & (age=D or lb_gluc=ln) ->mm_ori=B
Universe: 3091, LHS support: 101, RHS support: 425, Support: 51
Confidence: 0.50495, Lift: 3.67248, Leverage: 0.0120068, p-value:
    0, E-value: 0
3.67248

(lb_gluc=ln) & (lb_ast=l or lb_sodium=h) ->mm_ori=B
Universe: 3091, LHS support: 101, RHS support: 425, Support: 51
Confidence: 0.50495, Lift: 3.67248, Leverage: 0.0120068, p-value:
    0, E-value: 0
3.67248
```

Elementary clauses with the greatest negative effect on normal cognition are listed in Table 6.

4. Discussion

Among the 725 rules identified, 513 had lift values exceeding 2.00. Most of the rules exceeding even the 3.00 lift value had one thing in common: the LHS contained the premise *lb_sodium=h*.

4.1. Liver function

The rules found suggest that having high serum levels of AST (aspartate aminotransferase), as well as having low or high serum levels of ALT (alanine aminotransferase) may predispose to an impaired cognition characterized by low mini mental state examination (MMSE) scores. It should be noted that low ALT was much more rare in the CAMD database than high ALT, so the negative effect should be attributed mainly to high ALT. However, serum ALP (alkaline phosphatase) levels seem to have a controversial effect on mental status.

AST, ALT and ALP levels derive from the liver. Elevated ALP might indicate bile duct obstruction. AST or ALT may elevate in a number of cases of liver injury or damage, spreading from acute or chronic viral infections to alcohol induced or non-alcoholic steatohepatitis. It is interesting to note that elevated serum levels of AST (more than those of

Table 3

Some association rules involving serum cholesterol level. We refer to Table 4 for the legends applied in this table, and Section 1.1 for the simple implication-like formalism that describes the association rules found.

```
(lb_sodium=h) & (lb_chol=l or lb_gluc=ln) - - - > mm_ori=B
Universe: 2783, LHS support: 87, RHS support: 401, Support: 50
Confidence: 0.574713, Lift: 3.98859, Leverage: 0.0134618, p-value: 0, E-value: 0
3.98859

(lb_gluc=ln) & (lb_chol=l or lb_sodium=h) - - - > mm_ori=B
Universe: 2783, LHS support: 105, RHS support: 401, Support: 57
Confidence: 0.542857, Lift: 3.76751, Leverage: 0.0150451, p-value: 0, E-value: 0
3.76751

(lb_sodium=h) & (lb_chol=ln or lb_gluc=ln) ->mm_ori=B
Universe: 2783, LHS support: 106, RHS support: 401, Support: 55
Confidence: 0.518868, Lift: 3.60102, Leverage: 0.0142747, p-value:
    0, E-value: 0
3.60102

(lb_chol=h) & (lb_cl=h or lb_sodium=h) ->mm_ori=B
Universe: 1420, LHS support: 71, RHS support: 304, Support: 51
Confidence: 0.71831, Lift: 3.35526, Leverage: 0.0252113, p-value:
    2.22045e-016, E-value: 1.88773e-007
3.35526

(lb_chol=h) & (lb_monole=l or lb_sodium=h) ->mm_total=AB
Universe: 1364, LHS support: 73, RHS support: 325, Support: 58
Confidence: 0.794521, Lift: 3.33454, Leverage: 0.02977, p-value:
    1.51101e-013, E-value: 0.00012846
3.33454

(lb_sodium=h) & (lb_monole=h or lb_chol=h) ->mm_total=AB
Universe: 1364, LHS support: 66, RHS support: 325, Support: 51
Confidence: 0.772727, Lift: 3.24308, Leverage: 0.0258608, p-value:
    5.9952e-015, E-value: 5.09687e-006
3.24308

(lb_chol=h) & (lb_hbsag=h or lb_sodium=h) ->mm_attcal=B
Universe: 1164, LHS support: 67, RHS support: 312, Support: 50
Confidence: 0.746269, Lift: 2.78416, Leverage: 0.0275268, p-value:
    6.2725e-011, E-value: 0.0533262
2.78416

(lb_sodium=h) & (lb_bun=h or lb_chol=h) ->mm_attcal=B
Universe: 1387, LHS support: 61, RHS support: 429, Support: 52
Confidence: 0.852459, Lift: 2.75609, Leverage: 0.023888, p-value:
    8.87168e-012, E-value: 0.00754232
2.75609

(lb_sodium=h) & (lb_ca=l or lb_chol=h) ->mm_attcal=B
Universe: 1420, LHS support: 61, RHS support: 460, Support: 51
Confidence: 0.836066, Lift: 2.5809, Leverage: 0.0219996, p-value:
    2.52266e-011, E-value: 0.0214466
2.5809

(lb_sodium=h) & (lb_cl=h or lb_chol=h) ->mm_attcal=B
Universe: 1420, LHS support: 66, RHS support: 460, Support: 55
Confidence: 0.833333, Lift: 2.57246, Leverage: 0.0236759, p-value:
    1.65421e-010, E-value: 0.140634
2.57246
```

ALT) have been associated with impaired mental status. Although mild elevations in serum levels of AST and ALT are nonspecific to the etiology of liver injury, certain alteration patterns in these parameters may reflect the nature of the hepatic disease. For instance, the value of the AST/ALT ratio—also known as the De Ritis ratio—is approximately 0.8 in normal subjects, a ratio exceeding 2.00 being suggestive to alcoholic hepatitis.

Therefore we scanned the subjects with high AST values for higher than 2 AST/ALT ratio: we have only found 10 subjects satisfying both conditions. In addition, only 2 rules had AST/ALT on the left-hand side. Consequently, we may assume that high serum AST in the study subjects is not typically accompanied with high De Ritis ratio (i.e. probable alcoholic hepatitis).

The association of impaired liver function with mental decline can be illuminated in two perspectives. On one hand, impaired liver function might be insufficient to prevent the brain from the effects of certain neurotoxins, e.g. ammonia. This happens in the case of hepatic

Table 4
The legends for Tables 2 and 3. The MMSE-scores and age-classifications by the letters A, B, C and D are also defined in this table.

age	Subject age (A: ≤ 65 years, B: 66–75 years, C: 76–85 years, D: > 85 years)
ast_alt	De Ritis ratio
bpdia	Diastolic blood pressure
bpsys	Systolic blood pressure
lb_alb	Serum albumine
lb_alp	Serum alkaline phosphatase
lb_alt	Serum alanine aminotransferase
lb_ast	Serum aspartate aminotransferase
lb_baso	Basophils, particle concentration
lb_bili	Serum indirect bilirubin
lb_bun	Blood Urea Nitrogen
lb_ca	Serum calcium
lb_chol	Serum cholesterol
lb_ck	Serum creatine kinase
lb_cl	Serum chlorine
lb_creat	Serum creatinine
lb_eos	Eosinophils, particle concentration
lb_gluc	Serum glucose
lb_hba1c	Hemoglobin A1C
lb_hbsag	Hepatitis B virus surface antigen
lb_hct	Hematocrit
lb_hgb_blood	Blood hemoglobin
lb_k	Serum potassium
lb_ketones	Ketones
lb_ldh	Lactate dehydrogenase
lb_lym	Lymphocytes, particle concentration
lb_lymle	Lymphocytes/leukocytes ratio
lb_mch	Mean corpuscular hemoglobin
lb_mchc	Mean corpuscular hemoglobin concentration
lb_mcv	Mean corpuscular volume
lb_mono	Monocytes, particle concentration
lb_monole	Monocytes/leukocytes ratio
lb_neut	Neutrophils, particle concentration
lb_neutle	Neutrophils/leukocytes ratio
lb_ph	pH
lb_phos	Phosphate
lb_plat	Platelets
lb_prot	Total protein
lb_rbc_blood	Red blood count
lb_sodium	Serum sodium
lb_tsh	Thyrotropin
lb_vitb12	Serum B12 vitamin
lb_wbc_blood	White blood count
mm_attcal	MMSE attention and calculation subscore (B: 0–1, C: 2, D: 3, E: 4–5)
mm_lang	MMSE language subscore (B: 0–2, C: 3–4, D: 5–6, E: 7–9)
mm_ori	MMSE orientation subscore (B: 0–2, C: 3–4, D: 5–7, E: 8–10)
mm_recall	MMSE recall subscore (B: 0, C: 1, D: 2, E: 3)
mm_total	MMSE total score (A: < 10, B: 10–14, C: 15–23, D: ≥ 24)
pulse	Heart rate
resp	Respiratory rate
sex	Subject sex (F: female, M: male)
temper	Temperature

Table 5
Elementary clauses with the greatest positive effect on normal cognition. The "score" values refer to the MMSE scores. For the definition of the "elementary clause" we refer to Section 1.1.

lb_vitb12 = h	score: 67
lb_mch = h	score: 25
lb_mchc = l	score: 22
lb_k = h	score: 17
sex = M	score: 10
pulse = l	score: 9
lb_bun = l	score: 8
age = AB	score: 4
lb_mono = nh	score: 3
resp = ln	score: 3
lb_plat = ln	score: 2
lb_eos = nh	score: 2
lb_prot = nh	score: 2

Table 6
Elementary clauses with the greatest negative effect on normal cognition. The "score" values refer to the MMSE scores. For the definition of the "elementary clause" we refer to Section 1.

temper = nh	score: −10
lb_wbc_blood = h	score: −10
age = BCD	score: −10
lb_prot = h	score: −12
lb_gluc = h	score: −12
pulse = h	score: −12
lb_ck = h	score: −12
lb_hct = nh	score: −12
lb_k = ln	score: −12
lb_alp = h	score: −12
lb_chol = ln	score: −13
lb_ph = h	score: −13
lb_hct = l	score: −13
lb_alt = h	score: −13
age = A	score: −14
bpsys = ln	score: −14
lb_creat = ln	score: −14
lb_creat = h	score: −16
temper = l	score: −17
lb_alp = ln	score: −18
lb_bun = ln	score: −18
lb_alt = l	score: −19
lb_wbc_blood = l	score: −20
lb_chol = l	score: −21
pulse = nh	score: −21
lb_prot = ln	score: −22
lb_bun = h	score: −22
lb_plat = h	score: −26
lb_gluc = ln	score: −27
bpdia = ln	score: −28
age = CD	score: −32
lb_chol = h	score: −42
lb_ast = h	score: −43
lb_ca = l	score: −50
sex = F	score: −57
age = D	score: −99
lb_cl = h	score: −173
lb_sodium = h	score: −224

encephalopathy (HE), when severe liver damage resulting in acute liver insufficiency cannot detoxificate ammonia and other neurotoxins. On the other hand, the association of elevated AST/ALT ratio with impaired mental status proposes that another obscure element (e.g. chronic alcohol consumption) might be the factor responsible for both cognitive and metabolic damages.

Our results raise the possibility of a pathogenetic linkage between liver function and mental status in patients with AD. Such linkage has also been proposed by other studies (Astarita et al., 2010; Sutcliffe, Hedlund, Thomas, Bloom, & Hilbush, 2011). One study concludes that peripheral reduction of β-amyloid is sufficient to reduce brain β-amyloid and proposes that β-amyloids, which are of major pathogenic importance in AD may originate from the liver (Sutcliffe et al., 2011). Another research found that deficient liver production of docosahexaenoic acid (a neuroprotective fatty acid) correlates with impaired cognitive status in AD patients (Astarita et al., 2010).

To rule out the possibility when the elevated AST level is due to some medications taken, we compiled a detailed in Table S3 (in the supporting on-line material at http://uratim.com/CAMD) containing the number of subjects taking certain drugs, and the number of drug-takers with high AST. The data shows that, for example, 1929 subjects took Donepezil, while among the Donepezil-takers, only 415 have had high AST levels.

4.2. Serum sodium

A great number of rules (224) have high sodium on the left hand side, all of which have impaired cognition on the right hand side. Net

water loss is responsible for the majority of cases of hypernatremia (Adrogue & Madias, 2000). A recent publication, examining the causes and comorbidities in patients older than 65 years, has found that the most common cause of community-acquired hypernatremia is dehydration due to reduced oral intake (Turgutalp et al., 2012). More interestingly, they found that the most common comorbidity in this patient group was AD, present in 31.4% of patients with hypernatremia (Turgutalp et al., 2012). Hydration status has a significant impact on the volume of grey and white matter in the brain and on the quantity of cerebrospinal fluid as a hallmark of ventricular enlargement (Streitbuerger et al., 2012). The pattern of shrinkage in white matter volume and increase of the ventricular system due to dehydration is consistent with the structural brain changes observed during the progression of AD (Streitbuerger et al., 2012). In another study, patients with AD underwent bioelectrical impedance vector analysis to assess the body cell mass and hydration status related to AD (Buffa, Mereu, Putzu, Floris, & Marini, 2010). Results demonstrated a tendency towards dehydration in patients with AD (Buffa et al., 2010). Although the association of dehydration and AD is supported by these publications, the specific pathogenic nature of this association remains obscure (Buffa et al., 2010; Streitbuerger et al., 2012; Turgutalp et al., 2012).

4.3. Vitamin B12

Our results were able to present the beneficial impact of high levels of vitamin B12, also known as cobalamin, on cognition. Along with folate, vitamin B12 has an important role in the maintenance of genome integrity (Fenech, 2012). Although previous publications found association of low serum levels of vitamin B12 and AD (Malaguarnera et al., 2004; McCaddon et al., 2004), a recent systemic review on vitamin B12 status and cognitive impairment fails to declare a clear association between vitamin B12 status and dementia (O'Leary, Allman-Farinelli, & Samman, 2012). However, this review also found that studies using newer and more specific biomarkers of vitamin B12 status such as methylmalonic acid and holotranscobalamin were able to draw an association between mental decline and poor vitamin B12 status (O'Leary et al., 2012).

Although clinically vitamin B12 deficiency may result in macrocytic anaemia, in the case of AD patients the occurrence of macrocytic anaemia is rare and the neurological and hematological features are unrelated (McCaddon et al., 2004).

4.4. Hematological parameters

Additional interesting rules were detected regarding hematological parameters. In particular, independently from each other, high values of mean corpuscular hemoglobin (MCH), low values of mean corpuscular hemoglobin concentration (MCHC), and low values of mean corpuscular volume (MCV) were also associated with high MMSE scores. Although high values of MCH and low values of MCHC are present in the case of macrocyctic anaemia (with the addition of high levels of mean corpuscular volume, low levels of hemoglobin and hematocrit), such solely associations should not be discussed, as they may be coincidental.

Among the rules with lift values exceeding 2.00, other parameters of hematological status (such as level of hemoglobin, red blood cell number, white blood cell number) were also present. Monocyte and eosinophil levels also appear on the left hand side of many rules with high lift. These premises appear in combinations with various other (mostly non-hematological) premises.

4.5. Blood cholesterol and cognition

The positive or negative effects of high cholesterol values to Alzheimer's disease and cognition is a controversial issue. Some studies (e.g., Helzner et al., 2009; Whitmer, Sidney, Selby, Johnston, & Yaffe,

2005; Zambon et al., 2010) show negative effects of high cholesterol value for cognition, while other studies (Mielke et al., 2005; Reitz, Luchsinger, Tang, Manly, & Mayeux, 2005; Reitz, Tang, Luchsinger, & Mayeux, 2004) prove the positive effects for cognition.

Our data supports both conclusions in a sense. That is, low, low-normal and high cholesterol levels are all associated with impaired mental status, but with a different extent (scores −21, −13 and −42, respectively). See Table 3 for a selection of cholesterol-related rules from the larger Table S1 in the on-line supporting material at http://uratim.com/CAMD.

It is worth to note that, by Table 3, elevated, low or low-normal cholesterol levels do not necessarily mean a higher likelihood of impaired cognition by themselves, but only combined with high sodium.

A most recent study (Pierrot et al., 2013) shows that the neuronal expression of amyloid precursor protein APP controls the cholesterol 24-hydroxylase mRNA levels and decreases cholesterol turnover; therefore in certain setups, the presence of amyloid precursor proteins imply lowered cholesterol levels.

5. Conclusions

A 5821-patient, high-quality database was analyzed with original methods for combinatorial biomarkers of dementia. We have found some novel and also some already well established relations connected to cognition characteristics in a 5821 patient cohort. The already established findings prove the validity of our datamining approach, and the new findings, related to MCH, ALP and AST levels prove its power. Some more controversial biomarkers, including cholesterol level, were also re-discovered, and were placed into context of other attributes for negative and positive effects to cognition.

Conflicts of interest

The authors declare no conflicts of interest.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.archger.2017.08.006.

References

Adrogue, H. J., & Madias, N. E. (2000). Hypernatremia. *New England Journal of Medicine, 342*(May (20)), 1493–1499. http://dx.doi.org/10.1056/NEJM200005183422006.

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In B. Jorge, M. J. Bocca, & C. Zaniolo (Eds.). *Proc. of the 20th international conference on very large data bases (VLDB '94), vol. 1215* (pp. 487–499).

Agrawal, R., Imielinski, T., & Swami, A. N. (1993). Mining association rules between sets of items in large databases. In P. Buneman, & S. Jajodia (Eds.). *Proceedings of the 1993 ACM SIGMOD international conference on management of data* (pp. 207–216).

Astarita, G., Jung, K. M., Berchtold, N. C., Nguyen, V. Q., Gillen, D. L., Head, E., et al. (2010). Deficient liver biosynthesis of docosahexaenoic acid correlates with cognitive impairment in Alzheimer's disease. *PLoS ONE, 5*(9), e12538. http://dx.doi.org/10.1371/journal.pone.0012538.

Bermejo-Pareja, F., Benito-Leon, J., Vega, S., Medrano, M. J., Roman, G. C., & in Central Spain (NEDICES) Study Group ND (2008). Incidence and subtypes of dementia in three elderly populations of central Spain. *Journal of the Neurological Sciences, 264*(January (1–2)), 63–72.

Buffa, R., Mereu, R. M., Putzu, P. F., Floris, G., & Marini, E. (2010). Bioelectrical impedance vector analysis detects low body cell mass and dehydration in patients with Alzheimer's disease. *Journal of Nutrition Health and Aging, 14*(December (10)), 823–827.

Carlo, A. D., Baldereschi, M., Amaducci, L., Lepore, V., Bracco, L., Maggi, S., et al. (2002). Incidence of dementia, Alzheimer's disease, and vascular dementia in Italy. The ILSA Study. *Journal of the American Geriatrics Society, 50*(January (1)), 41–48.

Chaves, R., Gorriz, J. M., Ramirez, J., Illan, I. A., Salas-Gonzalez, D., & Gomez-Rio, M. (2011). Efficient mining of association rules for the early diagnosis of Alzheimer's disease. *Physics in Medicine and Biology, 56*(September (18)), 6047–6063. http://dx.doi.org/10.1088/0031-9155/56/18/017.

Dubois, B., Feldman, H. H., Jacova, C., Dekosky, S. T., Barberger-Gateau, P., Cummings, J., et al. (2007). Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *Lancet Neurology, 6*(August (8)), 734–746. http://dx.doi.org/10.1016/S1474-4422(07)70178-3.

Epstein, R. J. (2009). Unblocking blockbusters: Using Boolean text-mining to optimise clinical trial design and timeline for novel anticancer drugs. *Cancer Informatics, 7*, 231–238.

Fenech, M. (2012). Folate (vitamin B9) and vitamin B12 and their function in the maintenance of nuclear and mitochondrial genome integrity. *Mutation Research, 733*(May (1–2)), 21–33. http://dx.doi.org/10.1016/j.mrfmmm.2011.11.003.

Galustian, C., & Dalgleish, A. G. (2010). The power of the web in cancer drug discovery and clinical trial design: Research without a laboratory? *Cancer Informatics, 9*, 31–35.

Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining.* MIT Press.

Helzner, E. P., Luchsinger, J. A., Scarmeas, N., Cosentino, S., Brickman, A. M., Glymour, M. M., et al. (2009). Contribution of vascular risk factors to the progression in Alzheimer's disease. *Archives of Neurology, 66*(March (3)), 343–348. http://dx.doi.org/10.1001/archneur.66.3.343.

Ivan, G., Szabadka, Z., & Grolmusz, V. (2007). Being a binding site: Characterizing residue composition of binding sites on proteins. *Bioinformation, 2*(5), 216–221.

Jack, C. R., Lowe, V. J., Weigand, S. D., Wiste, H. J., Senjem, M. L., Knopman, D. S., et al. (2009). Serial PIB and MRI in normal, mild cognitive impairment and Alzheimer's disease: Implications for sequence of pathological events in Alzheimer's disease. *Brain, 132*(May (Pt 5)), 1355–1365. http://dx.doi.org/10.1093/brain/awp062.

Malaguarnera, M., Ferri, R., Bella, R., Alagona, G., Carnemolla, A., & Pennisi, G. (2004). Homocysteine, vitamin B12 and folate in vascular dementia and in Alzheimer's disease. *Clinical Chemistry and Laboratory Medicine, 42*(9), 1032–1035. http://dx.doi.org/10.1515/CCLM.2004.208.

McCaddon, A., Tandy, S., Hudson, P., Gray, R., Davies, G., Hill, D., et al. (2004). Absence of macrocytic anaemia in Alzheimer's disease. *Clinical and Laboratory Haematology, 26*(August (4)), 259–263. http://dx.doi.org/10.1111/j.1365-2257.2004.00618.x.

Mielke, M. M., Zandi, P. P., Sjogren, M., Gustafson, D., Ostling, S., Steen, B., et al. (2005). High total cholesterol levels in late life associated with a reduced risk of dementia. *Neurology, 64*(May (10)), 1689–1695. http://dx.doi.org/10.1212/01.WNL.0000161870.78572.A5.

Mossello, E., Ballini, E., Mello, A. M., Tarantini, F., Simoni, D., Baldasseroni, S., et al. (2010). Biomarkers of Alzheimer's disease: From central nervous system to periphery? *International Journal of Alzheimer's Disease, 2011*, 342980. http://dx.doi.org/10.4061/2011/342980.

Nanavati, A. A., Chitrapura, K. P., Joshi, S., & Krishnapuram, R. (2001). Mining generalised disjunctive association rules. In L. L. Henrique Paques, & D. Grossman (Eds.). *Proceedings of the tenth international conference on Information and knowledge management (CIKM '01)* (pp. 482–489). New York, NY, USA: ACM Available from http://dl.acm.org/citation.cfm?id=502666http://dl.acm.org/citation.cfm?id=502666http://dl.acm.org/citation.cfm?id=502666.

O'Bryant, S. E., Xiao, G., Barber, R., Reisch, J., Doody, R., Fairchild, T., et al. (2010). A serum protein-based algorithm for the detection of Alzheimer disease. *Archives of Neurology, 67*(September (9)), 1077–1081. http://dx.doi.org/10.1001/archneurol.2010.215.

O'Bryant, S. E., Xiao, G., Barber, R., Reisch, J., Hall, J., Cullum, C. M., et al. (2011). A blood-based algorithm for the detection of Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders, 32*(1), 55–62. http://dx.doi.org/10.1159/000330750.

O'Leary, F., Allman-Farinelli, M., & Samman, S. (2012). Vitamin B12 status, cognitive decline and dementia: A systematic review of prospective cohort studies. *British Journal of Nutrition, 108*(December (11)), 1948–1961. http://dx.doi.org/10.1017/S0007114512004175.

Pierrot, N., Tyteca, D., D'auria, L., Dewachter, I., Gailly, P., Hendrickx, A., et al. (2013). Amyloid precursor protein controls cholesterol turnover needed for neuronal activity. *EMBO Molecular Medicine, 5*(April (4)), 608–625. http://dx.doi.org/10.1002/emmm.201202215.

Prince, M., & Jackson, J. (2009). *World Alzheimer Report 2009* Alzheimer's Disease International Available from http://www.alz.co.uk/research/files/WorldAlzheimerReport.pdf.

Reitz, C., Tang, M. X., Luchsinger, J., & Mayeux, R. (2004). Relation of plasma lipids to Alzheimer disease and vascular dementia. *Archives of Neurology, 61*(May (5)), 705–714. http://dx.doi.org/10.1001/archneur.61.5.705.

Reitz, C., Luchsinger, J., Tang, M. X., Manly, J., & Mayeux, R. (2005). Impact of plasma lipids and time on memory performance in healthy elderly without dementia. *Neurology, 64*(April (8)), 1378–1383. http://dx.doi.org/10.1212/01.WNL.0000158274.31318.3C.

Rogers, J. A., Polhamus, D., Gillespie, W. R., Ito, K., Romero, K., Qiu, R., et al. (2012). Combining patient-level and summary-level data for Alzheimer's disease modeling and simulation: A beta regression meta-analysis. *Journal of Pharmacokinetics and Pharmacodynamics, 39*(October (5)), 479–498. http://dx.doi.org/10.1007/s10928-012-9263-3.

Romero, K., de Mars, M., Frank, D., Anthony, M., Neville, J., Kirby, L., et al. (2009). The Coalition Against Major Diseases: Developing tools for an integrated drug development process for Alzheimer's and Parkinson's diseases. *Clinical Pharmacology and Therapeutics, 86*(October (4)), 365–367. http://dx.doi.org/10.1038/clpt.2009.165.

Romero, K., Corrigan, B., Tornoe, C. W., Gobburu, J. V., Danhof, M., Gillespie, W. R., et al. (2010). Pharmacometrics as a discipline is entering the "industrialization" phase: Standards, automation, knowledge sharing, and training are critical for future success. *Journal of Clinical Pharmacology, 50*(September (9 Suppl.)), 9S–19S. http://dx.doi.org/10.1177/0091270010377788.

Ruczinski, I., Kooperberg, C., & LeBlanc, M. (2003). Logic regression. *Journal of Computational and Graphical Statistics, 12*(3), 475–511. http://dx.doi.org/10.1198/1061860032238.

Streitbuerger, D. P., Moller, H. E., Tittgemeyer, M., Hund-Georgiadis, M., Schroeter, M. L., & Mueller, K. (2012). Investigating structural brain changes of dehydration using voxel-based morphometry. *PLoS ONE, 7*(8), e44195.. http://dx.doi.org/10.1371/journal.pone.0044195.

Sutcliffe, J. G., Hedlund, P. B., Thomas, E. A., Bloom, F. E., & Hilbush, B. S. (2011). Peripheral reduction of beta-amyloid is sufficient to reduce brain beta-amyloid: Implications for Alzheimer's disease. *Journal of Neuroscience Research, 89*(June (6)), 808–814. http://dx.doi.org/10.1002/jnr.22603.

Turgutalp, K., Ozhan, O., Gok Oguz, E., Yilmaz, A., Horoz, M., Helvaci, I., et al. (2012). Community-acquired hypernatremia in elderly and very elderly patients admitted to the hospital: Clinical characteristics and outcomes. *Medical Science Monitor, 18*(December (12)), CR729–CR734.

Whitmer, R. A., Sidney, S., Selby, J., Johnston, S. C., & Yaffe, K. (2005). Midlife cardiovascular risk factors and risk of dementia in late life. *Neurology, 64*(January (2)), 277–281. http://dx.doi.org/10.1212/01.WNL.0000149519.47454.F2.

Wortmann, M. (2012). Dementia: A global health priority – Highlights from an ADI and World Health Organization report. *Alzheimer's Research and Therapy, 4*(September (5)), 40. http://dx.doi.org/10.1186/alzrt143.

Wu, W., Juan, W. C., Liang, C. R. M. Y., Yeoh, K. G., So, J., & Chung, M. C. M. (2012). S100A9, GIF and AAT as potential combinatorial biomarkers in gastric cancer diagnosis and prognosis. *Proteomics Clinical Applications, 6*(April (3–4)), 152–162. http://dx.doi.org/10.1002/prca.201100050.

Zambon, D., Quintana, M., Mata, P., Alonso, R., Benavent, J., Cruz-Sanchez, F., et al. (2010). Higher incidence of mild cognitive impairment in familial hypercholesterolemia. *American Journal of Medicine, 123*(January (3)), 267–274. http://dx.doi.org/10.1016/j.amjmed.2009.08.015.