Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 11–17

ISBN 978-83-60810-22-4

Improving the predictability of ICU illness severity scales

M. Alqarni,*, Y. Arabi,† T. Kakiashvili,‡ M. Khedr, † W. W Koczkodaj,*
J. Leszek, § A. Przelaskowski, ¶ K. Rutkowski‖
* Laurentian Universit, Sudbury, Ontario, Canada
† King Saud Bin Abdulaziz University for Health Sciences, Saudi Arabia
‡ Sudbury Therapy, Sudbury, Ontario, Canada
§ Medical University, Wroclaw, Poland
¶ Warsaw University of Technology, Warsaw, Poland
‖ Jagellonian University, Cracow, Poland

***Abstract*—This study demonstrates how to improve the predictability of one of the commonly used ICUs severity of illness scales, namely APACHE II, by using the consistency-driven pairwise comparisons (CDPC) method. From a conceptual view, there is little doubt that all items have exactly equal importance or contribution to predicting mortality risk of patients admitted to ICUs. Computing new weights for all individual items is a considerable step forward since it is based on reasonable to assume that not all individual items have equal contribution in predicting mortality risk. The received predictability improvement is 1.6% (from 70.9% to 72.5%) and the standard error decreased from 0.046 to 0.045. This must be taken as an indication of the right way to go.**

***Index Terms*—medical scales, illness severity, expert system, consistency-driven pairwise comparisons, inconsistency analysis.**

I. Introduction

EVER growing medical care costs motivate us to conduct more research toward the improvement of the severity of illness scales. The challenge is to leave the severity of illness scales unchanged as they are based on well established medical knowledge. Under this assumption, we are left with only one solution: weight for individual scale items must be computed instead of being arbitrarily set.

Severity of illness scales have wide application in medicine but psychiatry and intensive care settings are two top specializations where the use of severity of illness scales seems to be of great use, although for different reasons. In psychiatry, the use of tests, such as blood or X-rays, is limited and psychiatrists often rely on asking questions or observations. On the other hand, patients in the intensive care medicine must be rapidly evaluated based on many factors upon arrival and it can be at least in part done by nurses. Systems for predicting hospital mortality, such as the Acute Physiology and Chronic Health Evaluation (APACHE) II, are attractive options for this purpose because they rely on data collected within 24 hours after admission to the intensive care unit (ICU). It is mainly used to predict hospital mortality and reflect the severity of illness.

The corresponding author: wkoczkodaj@cs.laurentian.ca
Alphabetical order implies the equal contribution.

Acute Physiology and Chronic Health Evaluation, APACHE II, was introduced in [7] by Knaus in 1985 for predicting the hospital mortality in ICU patients. It has been designed based on data collected in 5,815 intensive care admissions from 13 hospitals. APACHE II measures severity of illness by a numeric score which can be converted into predicted mortality by using a logistic regression formula developed and validated on populations of ICU patients (for details, see [2], [3]).

II. ICU scales

Medical scales (sometimes called medical measures) are scales used to describe or assess medical conditions. Amongst them, ICU scales are of considerable importance. An intensive care unit (ICU) also called intensive therapy unit, critical care unit (CCU), or intensive treatment unit (ITU) is a specialized department in a hospital for providing intensive-care medicine. Some hospitals also have designated intensive care areas for certain specialties of medicine, depending on the needs and resources of the hospital. For example, stroke is usually treated this way. Glasgow Coma Scale (GCS) was first introduced in 1974 by Teasdale G, Jennett B. in [17]. It aims to provide an understandable and clear way of observing change in the level of consciousness of patients having head injures. In essence, the GCS was developed to standardize the reporting of neurologic findings and to provide an objective measure of the level of function of comatose patients [6]. Currently, GCS is one of the most used scales to assist the conditions of Trauma patients.It has only three items (elements):

1) Best eye response (E)
2) Best verbal response (V)
3) Best motor response (M)

There are four grades eye responses (E) starting with the most severe: 1 = ”No eye opening” to 4 = Eyes opening spontaneously. For verbal response (V), the grads range from 1 = ”Makes no sounds” to 5 = ”Oriented, converses normally”. Motor response (M) starts with 1 = ”Makes no movements” and ends with 6 = ”Obeys commands”. Generally, brain injury is classified as:

- Severe, with GCS ≤ 8

TABLE I
REVISED TRAUMA SCORE

Glasgow Coma Scale (GCS)	Systolic Blood Pressure (SBP)	Respiratory Rate (RR)	Coded Value
13-15	>89	10-29	4
9-12	76-89	>29	3
6-8	50-75	6-9	2
4-5	1-49	1-5	1
3	0	0	0

- Moderate, GCS 9 - 12
- Minor, GCS $\geq$ 13.

GCS is a part of several ICU scales, including APACHE II.

The Revised Trauma Score is a concatenation of: Glasgow Coma Scale, systolic blood pressure, and respiratory rate. Based on [4], TABLE II demonstrate the Revised Trauma Scale. The RTS ranges from 0 to 12. A patient with an RTS = 12 is categorized as DELAYED (e.g., walking wounded), 11 is URGENT (intervention is required but the patient can wait a short time), and 10-3 is IMMEDIATE (immediate intervention is necessary). The last possible category is MORGUE, which is given to mortally injured people having RTS score from o to 3.

Needless to say that with the method presented in this study, the predictability can be improved for all these scales. However, the deep throat for our research is data gathering. It is not only costly and time consuming but it is easy to envision that the data collection may often interfere with the rescuing efforts so the emergency physician intuition may surpass it.

III. DATA GATHERING AND ANALYSIS

Raw data, received from the Intensive Care Unit (ICU) of King Abdulaziz Medical City in Riyadh, Saudi Arabia in paper form, were entered into MS Excel to ease processing by other systems (e.g., SPSS). Excel provides a good tool for building forms for such a task using Visual Basic for Application (VBA) environment. Several forms were designed and then used to enter raw data. During the data entry process, 22 records from the received 165 records had to be removed. That was because three patients were re-admitted to the ICU, three patients had length of stay less than 24 hours, one patient had incomplete data, and all others had one or more missing value either it was not measured or was not available. According to [7], the first 3 group (re-admitted patients, patients with less than 24 hours stay, patients with incomplete data) are excluded from APACHE scoring. Moreover, in statical analysis, all patients with one or more missing values were removed.

While considerable efforts have been taken to ensure high standards throughout all stages of collection and processing, the resulting data may not be sterile as they are clinical. Despite this effort, it should be clear that accidental errors are inevitable. Data was looked at from the medical point of view for mistakes and errors. We stress the use of the clinical data in our analysis as opposed to trial data.

Trial data tend to be more sterile than clinical data but as such, less valuable. They have the tendency of generating better results since experiments are designed to prove a certain hypothesis.

TABLE II
APACHE II SIMPLIFIED SPECIFICATION WITH GIVEN WEIGHTS

No	Item	Description	Range	weight
1	Temp	Temperature	0-4	1.67%
2	MAP	Mean Arterial Pressure	0-4	1.67%
3	HR	Heart Rate	0-4	1.67%
4	RR	Respiratory Rate	0-4	1.67%
5	OXY	Oxygenation	0-4	1.67%
6	ρH	ρH (Arterial)	0-4	1.67%
7	SS	Sodium (Serum)	0-4	2%
8	PS	Potassium (Serum)	0-4	2%
9	CS	Creatinine (Serum)	0-4	2%
10	He	Hematocrit	0-4	2%
11	WBC	White Blood Count	0-4	2%
12	GCS	15-GCS	0-15	20%
13	Age	Age	0-6	20%
14	CHP	Chronic Health Points	0, 2, 5	40%

IV. PAIRWISE COMPARISONS LOGICAL MODEL

Based on existing data, we hypothesize that assuming identical weights for all scale items is not realistic for computing the predictability. Therefore, we grouped scale items according to their cohesiveness trying to have as loosely coupled groups as possible. It is done by the domain (medical) experts as following: First, a conceptual model must be designed by grouping criteria together. A rule of thumb proposed by Saaty in [14] is that no group should have more than seven criteria. To solve the problem of one group having a large number of criteria, split the group into subgroups with an acceptable number of criteria. Evidently, GCS, age and chronic conditions of patients have been kept in separation in the spirit of what was previously mentioned in the *Data collection and analysis* section. The remaining items have been grouped as shown in Fig. 1.

In a simplified model for APACHE II, we compared only the upper level observing that 11 items in the physiological group account for maximum 44 points while Chronic Health Point (CHP) is 2 or 5. It has created an immediate problem for relating groups at this level since our system allows us to define the ratio from 1 to 5 (and the inverse values). It is not a deficiency since the introduction of more groups is a solution. However, it requires more time so we decided to "cut the corners" and entered the compromised relative pairwise comparisons in Fig. 2.

The challenge posed to the pairwise comparisons method comes from the lack of consistency in assessments which arise in the real world [11]. With the development of the software, The Concluder, the consistency analysis has become relatively easy despite its complicated look. During the analysis process, the most inconsistent combinations of criteria are highlighted in the pairwise comparisons matrix in Fig. 3.

On Fig. 3, the maximum inconsistency is shown 0.67 and evidently bigger than the assumed heuristic $\frac{1}{3}$ (as explained in the Appendix) hence the relative importance had to be

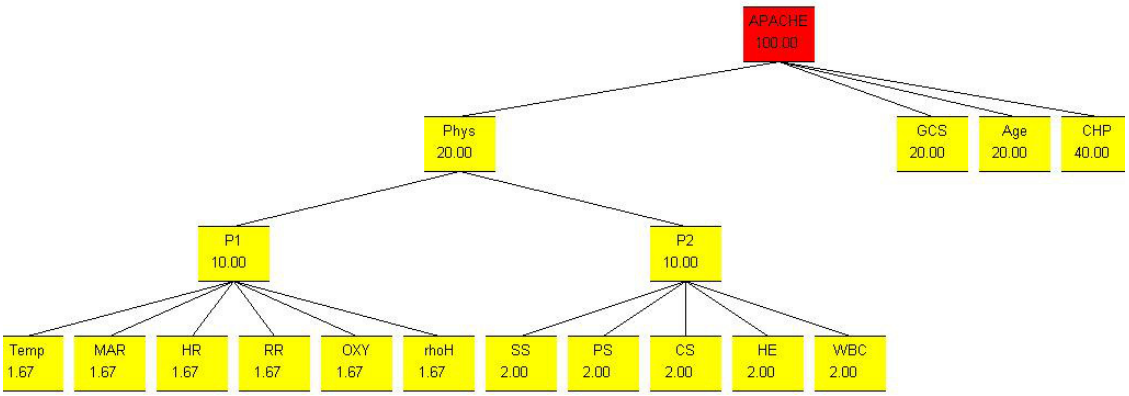

Fig. 1. The conceptual model of Apache II

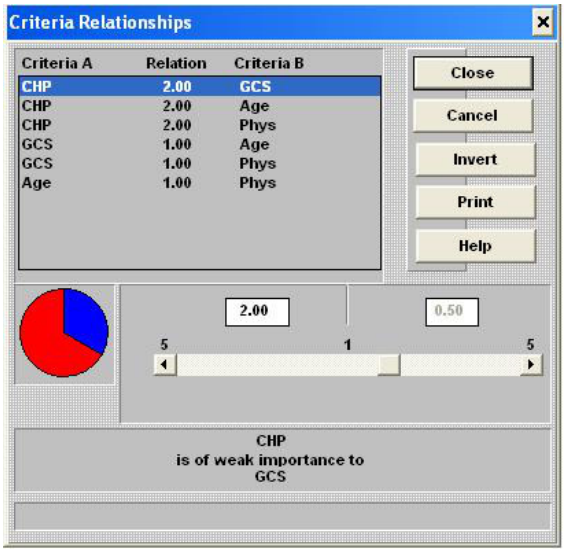

Fig. 2. The relative importance of the APACHE II scale items for level 1

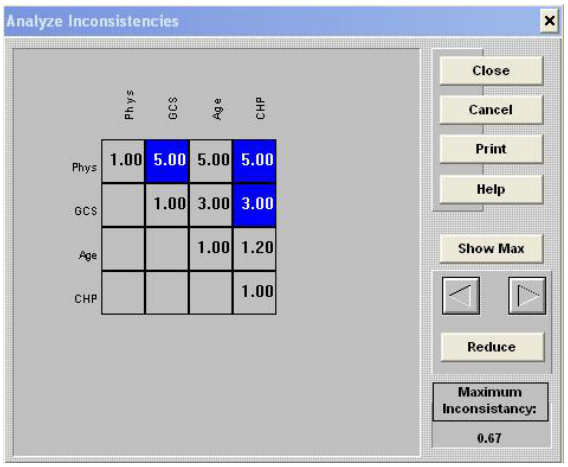

Fig. 3. The initial inconsistency analysis

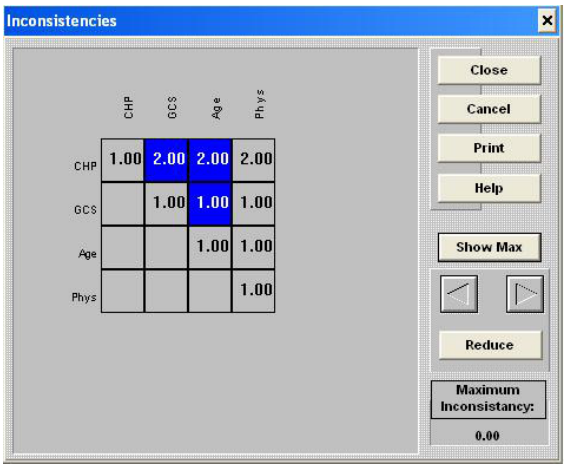

Fig. 4. The inconsistency analysis after the improvement

reconsidered by experts based on their professional judgement and medical knowledge.

When an acceptable consistency level is reached (in our case, it has happened to be 0.00) as shown in fig 4, the weights $[40\%, 20\%, 20\%20\%]$ are computed as normalized geometric means of rows and illustrated by Fig. 5. It needs to be stressed that for inconsistent pairwise matrix only approximated solution, in terms of weights, exists but it is sufficient since the reconstructed matrix from computed weights does not vary dramatically from the inconsistent pairwise comparisons matrix.

By the method of pairwise comparisons, the weight of 2.5 for CHP to Physio, GCS, and Age was optimal for the collected data. By diving the original ratios of CHP to Physio, CHP to GCS, and CHP to Age over 2.5, we get the values in the upper row in Fig. 4. By the inconsistency analysis, we reduced values in the second row of matrix in Fig. 4 from 3 to 1.

The comorbidity component of APACHE II is represented by Chronic Health Points (CHP). CHP are added for patients with a history of severe organ system deficiency or for patients who have immuno-compromised as follows:

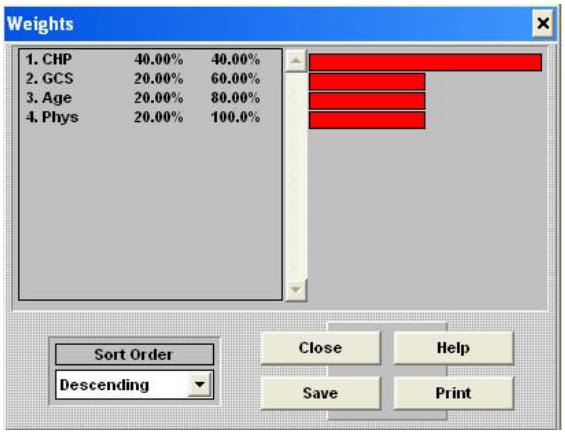

Fig. 5. The final weights computed for the APACHE II scale

- for nonoperative or emergency postoperative patients (5 points),
- for selective postoperative patients (2 points).

Immuno-compromised state must have been evident prior to the hospital admission and conform to the set medical criteria of liver, respiratory or renal system which are beyond the scope of this presentation. It was pointed out in [13] that: "The Chronic Health Points component of APACHE II had no significant discriminating ability (ROC area = 0.57, SE = 0.05)." However, we have provided evidence that the importance of CHP is greater than it has been originally assumed in [7] and should be changed from 2 and 5 to 5 and 12.5 respectively. To verify that these values are giving a better prediction of mortality, we applied ROC analysis by using SPSS. For CHP set to be 2.5, and 5, AUC (Area Under the Curve) was computed for the original data as 70.9% with the standard error of 0.046. While for the weight of 5 and 7.5 for CHP, AUC has increased to 72.5% with a better standard error of 0.045. The received predictability improvement is 1.6% and the standard error by 0.1%. Two ROCs are illustrated by Fig. 6.

In our humble opinion, the higher contribution of chronic conditions (CHP) can be explained by the simple observation that patients with chronic conditions are already receiving more medical attention than the rest of the population. When they are brought to ICU, usually it is for a serious enough reason that increases chance for their mortality.

V. Conclusions

In this data analytic study, we tested the impact of applying the pairwise comparisons method to intensive care scales such as APACHE II. As far as we are aware, this is the first study to examine APACHE II's effectiveness while improving the predictability of a clinical assessment using a well-established method. Our results, although seemingly modest, have been consistent with improved psychometric properties of the questionnaire examined, as evidenced by the superior AUC classifier percentage after weights were added.

It is a progress report and a part of the MSc degree thesis of the first author. The presented hypothesis of changing the weight for just one APACHE II scale item (Chronic Health Points) has been statistically proven and published in [9], [10], [1].

However, this is the first time of validation on the presented clinical data and every attempt will be made to use other clinical data in the future. We have proven the hypothesis that the proposed values in [7] in 1985 for Chronic Health Points should be changed from 2 and 5 to 5 and 7.5 respectively for elective post-operative patients and non-operative or emergency post-operative patients. This hypothesis is of considerable importance for health care planners. We hope that a more labor intensive analysis for the second level of the model would further improve the accuracy of the predictions of the presented scale. Similarly, adding the 50 principal diagnosis categories leading to ICU admission is a bit time consuming but will be incorporated in our approach for approximation of the mortality.

Acknowledgment

We would like to acknowledge the endeavors of the sponsor, the Kingdom of Saudi Arabia, Ministry of Higher Education.

References

[1] Adamic, P. , Babiy, V. , Janicki, R. Kakiashvili, T. , Koczkodaj, W. W., Tadeusiewicz, R. *Pairwise comparisons and visual perceptions of equal area polygons.* Perceptual and Motor Skills,108(1):37-42, 2009.

[2] Arabi, Y. , Abbasi, A., Goraj, R., Al-Abdulkareem,A., AL-Shimemeri,A., Kalayoglu,M., Wood, .K *External validation of a modified model of Acute Physiology and Chronic Health Evaluation (APACHE) II for orthotopic liver transplant patients.* Critical Care Medicine,6(3),245250,2002.

[3] Arabi, Y. , Haddad, S., Goraj, R. Al-Shimemeri, A., Al-Malik, S. *Assessment of performance of four mortality prediction systems in a Saudi Arabian intensive care unit.* Critical Care Medicen.6(2):166-74, 2006.

[4] http://www.trauma.org/archive/scores/rts.html (accessed on 2011-06-24)

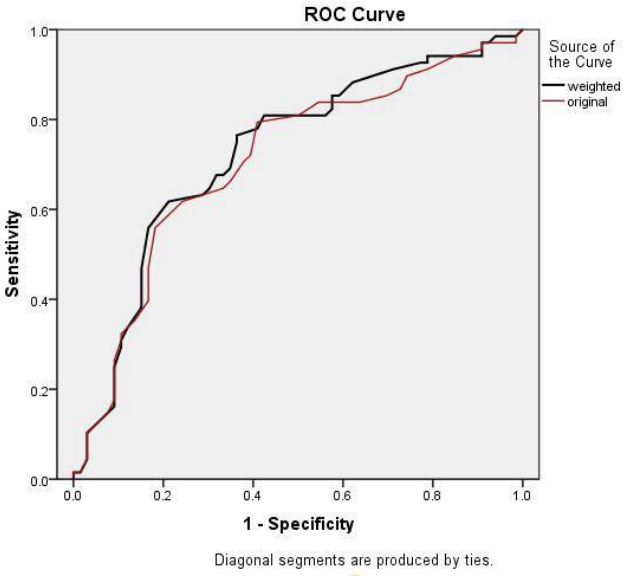

Fig. 6. Comparison between ROC results on both set of data

[5] Fawcett, T. *An introduction to ROC analysis*. Pattern Recognition Letters.27(8),861-874, 2006.
[6] Fischer J, Mathieson C. *The history of the Glasgow Coma Scale: Implications for practice* Critical Care Nurse,23(4):52-8, 2001
[7] Knaus, W., Draper, E., Wagner, D., Zimmerman, J. *APACHE II: a severity of disease classification system.*, Critical Care Medicine.13(10):818-29, 1985.
[8] Koczkodaj, W.W. *A new definition of consistency of pairwise comparisons*, Mathematical and Computer Modelling, (18)7,79-84,1993.
[9] Koczkodaj, W.W. *Statistically Accurate Evidence of Improved Error Rate by Pairwise Comparisons*, Perceptual and Motor Skills, 82,43-48, 1996.
[10] Koczkodaj, W.W., *Testing the Accuracy Enhancement of Pairwise Comparisons by a Monte Carlo Experiment*, Journal of Statistical Planning and Inference, 69(1), 21-32, 1998.
[11] Koczkodaj, W.W, LeBrasseur, R., Wassilew, A. ,Tadeusziewicz, R. *About Business Decision Making by a Consistency-Driven Pairwise Comparisons Method.*, Journal of Applied Computer Science.2009
[12] Koczkodaj, W.W., Szarek,S.J. *On distance-based inconsistency reduction algorithms for pairwise comparisons.*, Logic Journal of the IGPL 18(6):859-869, 2010.
[13] Poses, R., McClish, D., Smith, W., Bekes, C. , Scott, W. *Prediction of survival of critically ill patients by admission comorbidity.*, J Clin Epidemiol, 49(7):743-7, 1996.
[14] Saaty, L.T. *A Scaling Method for Priorities in HierarchicalStructures*, Journal of Mathematical Psychology 15(3),234-281,1977.
[15] Statsistics Canada offcial web page, http://www40.statcan.gc.ca/l01/cst01/demo02a-eng.htm?sdi=population, data retrieved on 2011-05-27
[16] Statsistics Canada offcial web page, http://www40.statcan.ca/l01/cst01/health30a-eng.htm, data retrieved on 2011-05-27
[17] Teasdale G, Jennett B. *Assessment of coma and impaired consciousness. A practical scale.* Lancet,2(7872):81-4, 1974.
[18] Zhai, Y., Janicki, R. *On Consistency in Pairwise Comparisons Based Numerical and Non-Numerical Ranking.*, Proceedings of the International Conference on Foundations of Computer Science, FCS 2010: 183-186, 2010.

APPENDIX

Using pairwise comparisons is a powerful method for synthesizing measurements and subjective assessments. From the mathematical point of view, the pairwise comparisons method generates a matrix (say A) of ratio values (a_{ij}) of the ith entity compared with the jth entity according to a given criterion. Entities/criteria can be both quantitative or qualitative allowing this method to deal with complex decisions. Comparing two entities in pairs to assess which of them is preferred, or has a greater amount of some property is irreducible since having one entity compared with itself has very little or practical meaning. However, subjective assessments often involve inconsistency, which is usually undesirable. The assessment can be refined via analysis of inconsistency, leading to reduction of the latter.

Making one comparison at a time is simpler than simultaneously assessing *all* items of a scale according to their contribution to the overall score. However, we need a method of synthesizing these partial assessments. The pairwise comparisons method, used since 1785, serves exactly this purpose, with the inconsistency analysis allowing us to localize the most questionable partial assessments and revise them if necessary.

From the mathematical point of view, the pairwise comparisons method creates a matrix (say A) of values (a_{ij}) of the ith entity compared with the jth entity:

$$A = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ \frac{1}{a_{12}} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{a_{1n}} & \frac{1}{a_{2n}} & \cdots & 1 \end{bmatrix}$$

A scale $[\frac{1}{c}, c]$ is used for 'i to j' comparisons where $c > 1$ is a not-too-large real number (5 to 9 is used in most practical applications). It is usually assumed that all the values a_{ii} on the main diagonal are 1 (the case of 'i compared with i', that is with itself) and that matrix A is *reciprocal* : $a_{ij} = \frac{1}{a_{ji}}$ since 'i to j' is (or at least, is expected to be) the reciprocal of 'j to i'. (In other words, for $x, y \neq 0$, $\frac{x}{y} = \frac{1}{\frac{y}{x}}$.) However, in practice even the reciprocity condition is not always guaranteed. For example, in blind wine testing we may conclude that *wine* i is better than *wine* i if it is served in unmarked glasses.

Since 1996, a *distance-based* adjective has been used by other researchers for the new inconsistency defined in 1993 in [8]. The distance-based adjective reflects the nature of the *inconsistency indicator*, which is defined, in essence, as a function of a distance from the nearest consistent *triad* in matrix A. Unlike the eigenvalue-based inconsistency, introduced in [14]), which is of a *global* indicator, and as such a non-idetifying, the distance-based inconsistency identifies the most inconsistent triad (or triads). It is the maximum over all triads $\{a_{ik}, a_{kj}, a_{ij}\}$ of elements of A (say, with all indices i, j, k distinct) of their inconsistency indicators, which in turn are defined as $ii := \min(|1 - \frac{a_{ij}}{a_{ik}a_{kj}}|, |1 - \frac{a_{ik}a_{kj}}{a_{ij}}|)$.

The inconsistency indicator of A equals zero if and only if A is fully consistent as it was (in all likeliness shown for the rist time in [14]. Consistent matrices correspond to the ideal situation in which we know all exact values of all properties (or at least it seems to be a reasonable assumption to make). However, a realistic situation which is complex enough, nearly always involves inconsistency and we need to deal with it. In fact, when we are able to locate it, our comparisons can be reconsidered to reduce the inconsistency in the next round.

Certainly, inconsistency is undesirable in a system. On the other hand, although this may sound strange, it is not easy (we suspect, impossible) to construct a non-trivial *fully* inconsistent system: an "ideal" system where everything contradicts everything else. This question (or a family of questions, which we suggest only vaguely here) seems quite important as such impossibility would imply that *every* scenario of answers to pairwise comparison queries (even deliberately false) would necessarily create "apparent" consistencies.

In practical applications, a high value of the inconsistency indicator is a "red flag," or a sign of potential problems. A distance-based inconsistency reduction algorithm focuses, at each step, on an inconsistent triad and "corrects" it by replacing it with a consistent (or, more generally, less inconsistent) triad. It resembles "whac-a-mole," a popular arcade game. One difference is that instead of one mole, we have three array elements as explained above. After "hitting the mole" (which generally results in some other "moles" coming out), the next triad is selected according to some rule (which may be for example the greedy algorithm), and

the process is repeated. Numerous practical implementations (e.g., a hazard rating system for abandoned mines in Northern Ontario) have shown that the inconsistency converges relatively fast. However, the need for rigorously *proving* the convergence (that is, showing that whacked moles *always* have the tendency of coming out less and less eagerly) was evident.

The distance-based inconsistency locates the most inconsistent triad or triads. This allows the user to reconsider the assessments included in the most inconsistent triad.

	A	**B**	**C**	**D**
A	1	1	5	4
B	1	1	2	$2\frac{1}{2}$
C	$\frac{1}{5}$	$\frac{1}{2}$	1	$\frac{1}{2}$
D	$\frac{1}{4}$	$\frac{2}{5}$	2	1

(1)

Changing the value 1 in the above triad to 2.5 makes this triad fully consistent since $2.5 \cdot 2 = 5$. Unfortunately, this is not the end of our problems since there is another triad $[2, 2\frac{1}{2}, \frac{1}{2}]$ that is inconsistent and "boxed" below:

	A	**B**	**C**	**D**
A	1	$2\frac{1}{2}$	5	4
B	$\frac{2}{5}$	1	2	$2\frac{1}{2}$
C	$\frac{1}{5}$	$\frac{1}{2}$	1	$\frac{1}{2}$
D	$\frac{1}{4}$	$\frac{2}{5}$	2	1

(2)

Assume that we have good reason (coming from the knowledge domain; not from mathematics), to change the value of $2\frac{1}{2}$ to 1 It is an arbitrary decision since 2 could have been changed to 5 or $\frac{1}{2}$ to $1\frac{1}{4}$ also making this triad consistent. Only the domain knowledge can determine the change of the value (or values) in a triad. However, changing 2 may not be wise since it belongs to a consistent triad altered in the previous step. In our case, the only reason why we have chosen to change $2\frac{1}{2}$ to 1 was to illustrate how the inconsistency procedure works and the reader may be disappointed to find that there is yet another triad "boxed" below which is inconsistent:

	A	**B**	**C**	**D**
A	1	$2\frac{1}{2}$	5	4
B	$\frac{2}{5}$	1	2	1
C	$\frac{1}{5}$	$\frac{1}{2}$	1	$\frac{1}{2}$
D	$\frac{1}{4}$	1	2	1

(3)

Finally, we change 4 to $2\frac{1}{2}$ making the entire table fully consistent.

	A	**B**	**C**	**D**
A	1	$2\frac{1}{2}$	5	2.5
B	$\frac{2}{5}$	1	2	1
C	$\frac{1}{5}$	$\frac{1}{2}$	1	$\frac{1}{2}$
D	$\frac{2}{5}$	1	2	1

(4)

In practice, inconsistent assessments are unavoidable when at least three factors are independently compared against each other. The corrections for real data are done on the basis of professional experience, the case-based knowledge, and by the careful examination of all criteria involved (not necessarily in the current triad).

An acceptable threshold of inconsistency, for most practical applications, turns out to be $\frac{1}{3}$. This is so because one value in a triad is not more than two grades off the scale from the remaining two values. This heuristic was introduced in [8] and it seems more mathematically sound than 10% proposed in [14].

There is no need to continue decreasing the inconsistency indefinitely to zero, as only a high value of it is harmful. In fact, a zero or a small inconsistency value may indicate that artificial data were entered hastily without reconsideration of former assessments, which is an unacceptable practice.
For the improved matrix, the normalized vector of weights is:

$$w = [0.5, 0.2, 0.1, 0.2]$$

It is identical for both the geometric means method, and the eigenvector method, since the eigenvector of a consistent pairwise comparisons matrix is always equal to the geometric means. For the original input matrix, which is inconsistent, the solutions are,
for the eigenvector method:

$$w = [0.441, 0.317, 0.101, 0.140]$$

and for geometric means method (computed as $\sqrt[4]{\prod_{j=1}^{N} a_{ij}}$):

$$w = [0.445, 0.315, 0.100, 0.141]$$

The difference between both solutions is negligible. However, both solutions for the inconsistent matrix vary drastically from the solution for the consistent matrix.

It is important to note the difference between inaccuracy and inconsistency. For example, in a triad $[2, 5, 3]$, a rash approach may lead us to believe that A/C should indeed be 6 since it is $2 \cdot 3$, but we do not have any reason to reject the estimation of B/C as 2.5 or A/B as $5/3$. This is what inconsistency is about. It is not inaccuracy, but when used wisely, it may help to decrease inaccuracy.

The reader will notice that while the three-step inconsistency-reduction procedure performed above does not offend the common sense, it is rather *ad hoc*, hence not fully satisfactory. This remark applies both to the choices of triads to be corrected, and to the choices of the particular members of each such triad that is being modified. The algorithm analyzed in [12] (and, by extension, the present note) is more canonical with respect to the second point. In general, it replaces the triad $\{a_{ik}, a_{kj}, a_{ij}\}$ by $\{a_{ik}/r, a_{kj}/r, ra_{ij}\}$, where $r := \sqrt[3]{a_{ik}, a_{kj}/a_{ij}}$. This corresponds to subtracting from the matrix $(\log a_{uv})$ its orthogonal projection onto the direction of the skew-symmetric matrix $B = (b_{uv})$ defined by the requirement that $a_{ik} = 1 = a_{kj}, a_{ij} = -1$ and that all other super-diagonal entries are 0; the corresponding subspace in the context of Theorem is $U = \{X : X$ is an $n \times n$ skew-symmetric matrix such that $\mathrm{tr}BX = 0\}$. In particular, for the first triad $[1, 2, 5]$

considered above, we have $r = 2/5$ and the corrected triad is $[\sqrt[3]{5/2}, 2\sqrt[3]{5/2}], 5\sqrt[3]{2/5}] \approx [1.36, 2.71, 3.68]$.

Monte Carlo studies have shown that approximations of highly inconsistent pairwise comparisons matrices yields high errors. Finding consistent approximations of such matrices makes little practical sense. From mathematical logic, we know that *only* falsehood can generate both truth or falsehood. However, the old adage that *one bad apple spoils the barrel* seems to be more applicable here: even a little bit of falsehood may contribute to massive errors and misjudgments. An approximation of a pairwise comparisons matrix is meaningful only if the initial inconsistency is acceptable (that is, located, brought under control and/or reduced to a certain predefined minimum; in our analogy, *always remove overripe fruit promptly if it is possible to find it*).

The new results and applications of pairwise comparisons show the importance of the consistency-driven approach. The inconsistency concept still remains enigmatic and more research needs to be done. In particular, inconsistency in a general system needs to be defined and this study is a step forward. The idea of improving inaccuracy by controlling inconsistency cannot be wrong and a new approach to it is presented in [18]. Knowing what we do not know is essential to managing the knowledge and improving it. On the other hand, it is hard to change our knowledge if we choose not to know what we know or even should know.

The method of pairwise comparisons was used by a research team, lead by W.W. Koczkodaj, to develop AMIS (Abandoned Mines Hazard Rating System) for the government of Ontario (The Ministry of Northern Ontario and Mines). The system ranked an abandoned mine, located in Northern Ontario, as one of the most dangerous from a public safety point of view. Its eventual collapse convinced the government that its research founding was well spent.