Proceedings of the International Multiconference on
Computer Science and Information Technology pp. 151–154

ISBN 978-83-60810-14-9
ISSN 1896-7094

Assessing the Properties of the World Health Organization's Quality of Life Index

Tamar Kakiashvili, Waldemar W. Koczkodaj*,
Phyllis Montgomery, Kalpdrum Passi
Laurentian University, Sudbury, Ontario, Canada
*the corresponding author: wkoczkodaj@cs.laurentian.ca

Ryszard Tadeusiewicz
AGH University of Science and Technology, Poland

***Abstract*—This methodological study demonstrates how to strengthen the commonly used World Health Organization's Quality of Life Index (WHOQOL) by using the consistency-driven pairwise comparisons (CDPC) method. From a conceptual view, there is little doubt that all 26 items have exactly equal importance or contribution to assessing quality of life. Computing new weights for all individual items, however, would be a step forward since it seems reasonable to assume that all individual questions have equal contribution to the measure of quality of life. The findings indicate that incorporating differences of importance of individual questions into the model is essential enhancement of the instrument.**

***Index Terms*—Quality of life, inconsistency analysis, consistency-driven pairwise comparisons**

I. Introduction

DURING the past three decades, many different quality of life measures have been developed for use as an indicator of patient-focused health outcomes. Commonly used quality of life measures had reported psychometric properties. One measure in particular, the World Health Organization Quality of Life (WHOQOL) measure has been widely field tested since its inception in 1991. This measure was developed to an international cross-culturally comparable quality of life assessment for clinical populations. Its purpose is to assess subjects' perceptions of their quality of life in the context of their personal values and beliefs as well as their social culture. There are three versions of this measure. The first contains 100-items and is commonly used in large clinical trails. The brief version, WHOQOL-BREF, consists of 26-items.

A third version available on the web (www.who.int/mental_health/media/en/76.pdf) seems to be a modification of what was previously published in [9].

To supplement to the WHOQOL User's manual published in 1998, this research will further examine the psychometric properties of the measure by using consistency-driven pairwise comparison. The assumption of this approach is that not all instrument items are of equal importance and including the relative importance in the model contribute to the enhancement of the measure.

II. The Pairwise Comparison Preliminaries and a Quality of Life Model

From a mathematical point of view, the pairwise comparisons method [2],[3],[4], Appendix A creates a matrix (for example, A) of values (aij) of the i-th candidate (or alternative) compared head-to-head (one-on-one) with the j-th candidate. A scale $[1/c, c]$ is used for i to j comparisons where $c > 1$ is a not-too-large real number (5 to 9 in most practical applications).

It is usually assumed that all the values (aij) on the main diagonal are 1 (the case of i compared with i and that A is reciprocal: $(aij) = 1/(aij)$ since i to j is (or at least, is expected to be) the reciprocal of j to i. (As explained below, the reciprocity condition is not automatic in certain scenarios of comparisons.) It is fair to assume that we are powerless, or almost powerless, as far as inconsistency is concerned. All we can do is to locate it and reconsider our own comparisons to reduce the inconsistency in the next round.

An pioneer of this method is Condorcet [5]. He used the pairwise comparisons in 1785 in the context of counting political ballots. In 1860, however, Fechner provided further yet limited psychometric information about this method. By way of refining the method, Thurstone [7] described pairwise comparisons method as a statistical analysis and proposed a solution. In 1977, Saaty [8] introduced a hierarchy instrumental for practical applications.

The WHOQOL addresses four domains: physical health (PH), psychological health (PSYCH), social relationships (SR), and environment (ENV). Using the Saaty's heuristic approach of having no more than seven items in one group, the ENV domain was mechanically subdivide into ENV1 and ENV2 since it has eight objects. Implementing the Concluder system, the results are shown in Fig. 1. This figure is an illustration of the full model due to screen limitation and scrolling. The items listed on the left-hand side of Fig. 1 should be attached to the first level (groups) to create a hierarchical structure.

Using a scale 1 to 5, the relative importance of each of the five groups were entered and compared objects in the smallest subgroup. For example, SR (social relationships) was compared against each other. For example, "personal relationship" and "sex" are compared to each other in the subgroup SR (social relationship) and given 4 out of 5 (which can be changed for every clinical case to which this instrument is applied). The results are presented in Table 1.

Clearly, the above matrix was not consistent since $a_{13} = 4$. It did not equal to $a_{12} * a_{23}$. To address the inconsistency, ii, the

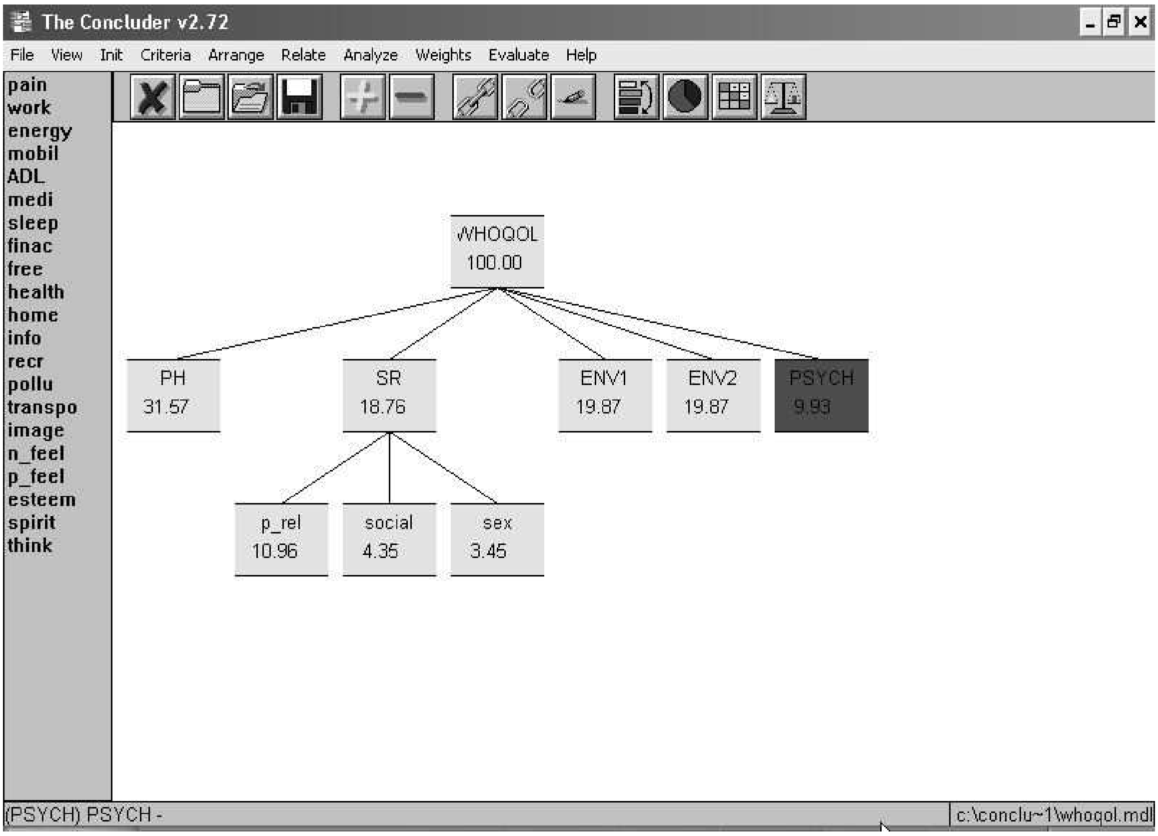

Fig. 1. Tree strusture

TABLE I
COMPARISONS FOR THE FIRST GROUP LEVEL

1	2	4
$\frac{1}{2}$	**1**	1
$\frac{1}{4}$	1	**1**

TABLE II
WEIGHTS FOR SUBGROUP SR

Group	Weight
A	0.1096
B	0.0435
C	0.0335

following formula where: ii =min$(|1 - a_{ij}/(a_{ik} * a_{kj})|, |1 - a_{ik} * a_{kj}/a_{ij}|)$ for $i = 1, j = 2$, and $k = 3$ (as introduced in [3]). was 0.5 and it was higher than the assumed threshold 1/3. The computed weights (as normalized geometric means of rows) are presented in Table 2. By changing, 4 to 3 for the illustration (rather than clinical), the new inconsistency index ii is computed as 1/3. Alternatively, another approach could have been changing a_{33} to 2 resulting in $ii = 0$. purposes We stress that the changes here have been done for the illustration of the method but in real life, there must be a reason coming from the refined clinical knowledge.

As explained Koczkodaj [[3]], the above values were computed as normalized geometric means of the matrix rows. Figure 1 shows one highlighted subgroup, SR (social relationships), because attaching more subgroups creates a structure that requires additional space. Alternatively, we could show one group in one figure but again, five screen images, which are nearly identical would only make this presentation excessively long so the reader is asked to use his/her imagination as we enter "behind the scene" all objects (listed as unchained on the left hand side margin in Fig. 1) and compare them against each other assigning relative importance and paying attention to inconsistency as it was demonstrated for SR subgroup. Again, the comparisons have been done to illustrate the method, not the real instrument and the overall results for all criteria is presented in Table 3.

CONCLUSION

Although the method of pairwise comparisons was originally used over 200 years ago, it has not been used to refine the properties of quality of life instruments. The method has strengthened the WHOQOL instrument by adding weights to individual items. Evidently, not all objects on the WHOQOL instrument are of equal importance. Appreciation of their relative differences, adds to the measure's precision. The inconsistency analysis further strengthens the measure by bringing the most problematic but often crucial comparisons of the instrument items. A challenge to the multiple experts in this tool's development can be "averaging" their individual assessments in the assumed model. Clinical trials and statistical analysis

TABLE III
THE FINAL WEIGHTS

Object	Weight
p_rel	0.1096
pain	0.1037
finac	0.0497
free	0.0497
health	0.0497
home	0.0497
info	0.0497
recr	0.0497
pollu	0.0497
transpo	0.0497
social	0.0435
mobil	0.0395
work	0.0369
ADL	0.0352
medi	0.0352
sex	0.0345
energy	0.0335
sleep	0.0318
image	0.0166
n_feel	0.0166
p_feel	0.0166
esteem	0.0166
spirit	0.0166
think	0.0166

need to follow the model enhancement. The enhancement to the WHOQOL project may be a challenging undertaking for years to come. Refinement of the WHOQOL may improve organizations' (such as WHO, the United Nations Educational, Scientific and Cultural Organization (UNESCO), United Nations) understanding of as well as health care professional practices in their efforts to assess quality of life.

ACKNOWLEDGMENT

This work has been supported by the Ministry of Science and Higher Education, Republic of Poland, under project number N516 007 32/0978 decision No. 0978/T02/2007/32 in Poland, the NSERC grant and LURF in Canada.

REFERENCES

[1] WHOQOL User's Manual, www.who.int/mental_health/media/en/76.pdf

[2] Kakiashvili, T., Kielan, K., Koczkodaj, W.W., Passi, K., Tadeusiewicz, R., Supporting the Asperger Syndrome Diagnostic Process by Selected AI Methods, proceedings of artificial intelligence studies. Vol. 4, pp. 21–27, 2007.

[3] Koczkodaj, W. W., A new definition of consistency of pairwise comparisons, Mathematical and computer modelling, (18)7, pp. 79–84, 1993.

[4] Koczkodaj, W. W., Herman, M. W., Orlowski, M. Using Consistency-driven Pairwise Comparisons in Knowledge-based Systems, International Conference on Information and Knowledge Management, 1997.

[5] Condorcet, M., The Essay on the Application of Analysis to the Probability of Majority Decisions, Paris: Imprimerie Royale, 1785.

[6] Fechner, G., Elemente der Psychophysik (1860, 2nd ed., 1889)

[7] Thurstone, L. L. (1927). A law of comparative judgement. Psychological Review, 34, 278–286.

[8] Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. Journal of Mathematical Psychology, 15, 234–281.

[9] McDowell. I., Measuring health : a guide to rating scales and questionnaires, 3rd ed., New York : Oxford University Press, 2006, 748 p.

APPENDIX A BASIC CONCEPTS OF PAIRWISE COMPARISONS

An n by n pairwise comparisons matrix is defined as a square matrix $A = [a_{ij}]$ such that $a_{ij} > 0$ for every $i, j = 1, ..., n$. Each a_{ij} expresses a relative preference of criterion (or stimulus)s_i over criterion s_j for $i, j = 1, ..., n$ represented by numerical weights (positive real numbers) and w_i and w_j respectively. The quotients $a_{ij} = w_i/w_j$ form a pairwise comparisons matrix:

$$A = \left| \begin{array}{cccc} 1 & a_{13} & ... & a_1 \\ & & & n \\ & & & \\ 1/a_1 & 1 & ... & a2 \\ 3 & & & n \\ ... & ... & ... & ... \\ 1/a_1 & 1/a_2 & ... & 1 \\ n & n & & \end{array} \right|$$

A pairwise comparisons matrix A is called *reciprocal* if $a_{ij} = 1/a_{ji}$ for every $i, j = 1, ..., n$ (then automatically $a_{ii} = 1$ for every $i = 1, ..., n$ because they represent the relative ratio of a criterion against itself). A pairwise comparisons matrix A is called *consistent* if $a_{ij} \cdot a_{jk} = a_{ik}$ holds for every $i, j, k = 1, ..., n$ since $w_i/w_j \cdot w_j/w_k$ is expected to be equal to w_i/w_k. Although every consistent matrix is reciprocal, the converse is not generally true. In practice, comparing of s_i to s_j, s_j to s_k, and s_i to s_k often results in inconsistency amongst the assessments in addition to their inaccuracy; however, the inconsistency may be computed and used to improve the accuracy.

The first step in pairwise comparisons is to establish the relative preference of each combination of two criteria. A scale from 1 to 5 can be used to compare all criteria in pairs. Values from the interval $[1/5, 1]$ reflect inverse relationships between criteria since $s_i/s_j = 1/(s_j/s_i)$. The consistency driven approach is based on the reasonable assumption that by finding the most inconsistent judgments, one can then reconsider one's own assessments. This in turn contributes to the improvement of judgmental accuracy. Consistency analysis is a dynamic process which is assisted by the software.

The central point of the inference theory of the pairwise comparisons is Saaty's Theorem, [8], which states that for every n by n consistent matrix $A = [a_{ij}]$ there exist positive real numbers $w_1, ..., w_n$ (weights corresponding to criteria $s_1, .n.., s)$ such that $a_{ij} = w_i/w_j$ for every $i, j = 1, ..., n$. The weights w_i are unique up to a multiplicative constant. Saaty (1977) also discovered that the eigenvector corresponding to the largest eigenvalue of A provides weights w_i which we wish to obtain from the set of preferences a_{ij}. This is not the only possible solution to the weight problem. In the past, a least squares solution was known, but it was far more computationally demanding than finding an eigenvector of a matrix with positive elements. Later, a method of row geometric means was proposed (Jensen, 1984), which is the simplest and most effective method of finding weights. A statistical experiment demonstrated that the accuracy, that is, the distance from the original matrix A and the matrix AN

reconstructed from weights with elements $[a_{ij}] = [w_i/w_j]$, does not strongly depend on the method. There is, however, a strong relationship between the accuracy and consistency. Consistency analysis is the main focus of the consistency driven approach.

An important problem is how to begin the analysis. Assigning weights to all criteria (e.g., $A = 18, B = 27, C = 20, D = 35$) seems more natural than the above process. In fact it is a recommended practice to start with some initial values. The above values yield the ratios: $A/B = 0.67$, $A/C = 0.9$, $A/D = 0.51$, $B/C = 1.35$, $B/D = 0.77$, $C/D = 0.57$. Upon analysis, these may look somewhat suspicious because all of them round to 1, which is of equal or unknown importance. This effect frequently arises in practice, and experts are tempted to change the ratios by increasing some of them and decreasing others (depending on knowledge of the case). The changes usually cause an increase of inconsistency which, in turn, can be handled by the analysis because it contributes to establishing more accurate and realistic weights. The pairwise comparisons method requires evaluation of all combinations of pairs of criteria, and can be more time consuming because the number of comparisons depends on n^2 (the square of the number of criteria). The complexity problem has been addressed and partly solved by the introduction of hierarchical structures [8]. Dividing criteria into smaller groups is a practical solution in cases in which the number of criteria is large.

Appendix B Consistency Analysis

Consistency analysis is critical to the approach presented here because the solution accuracy of *not-so-inconsistent* matrices strongly depends on the inconsistency. The consistency driven approach is, in brief, the next step in the development of pairwise comparisons.

The challenge to the pairwise comparisons method comes from a lack of consistency in the pairwise comparisons matrices which arises in practice. Given an n by n matrix A that is not consistent, the theory attempts to provide a consistent n by n matrix AN that differs from matrix A "as little as possible". In particular, the geometric means method produces results similar to the eigenvector method (to high accuracy) for the ten million cases tested. There is, however, a strong relationship between accuracy and consistency.

Unlike the old eigenvalue based inconsistency, introduced in [8], the triad based inconsistency locates the most inconsistent triads [3]. This allows the user to reconsider the assessments included in the most inconsistent triad.

Readers might be curious, if not suspicious, about how one could arrive at values such as 1.30 or 1.50 as relative ratio judgments. In fact the values were initially different, but have been refined and the final weights have been calculated by the consistency analysis. It is fair to say that making comparative judgments of rather intangible criteria (e.g., overall alteration and/or mineralization) results not only in imprecise knowledge, but also in inconsistency in our own judgments. The improvement of knowledge by controlling inconsistencies in the judgments of experts, that is, the consistency driven approach, is not only desirable but is essential.

In practice, inconsistent judgments are unavoidable when at least three factors are independently compared against each other. For example, let us look closely at the ratios of the four criteria A, B, C, and D in Figure $C1$. Suppose we estimate ratios A/B as 2, B/C as 3, and A/C as 5. Evidently something does not add up as $(A/B)@\ (B/C) = 2 \cdot 3 = 6$ is not equal to 5 (that is A/C). With an inconsistency index of 0.17, the above triad (with highlighted values of 2, 5, and 3) is the most inconsistent in the entire matrix (reciprocal values below the main diagonal are not shown in Figure $C1$). A rash judgment may lead us to believe that A/C should indeed be 6, but we do not have any reason to reject the estimation of B/C as 2.5 or A/B as 5/3. After correcting B/C from 3 to 2.5, which is an arbitrary decision usually based on additional knowledge gathering, the next most inconsistent triad is (5,4,0.7) with an inconsistency index of 0.13. An adjustment of 0.7 to 0.8 makes this triad fully consistent ($5 \cdot 0.8$ is 4), but another triad (2.5,1.9,0.8) has an inconsistency of 0.05. By changing 1.9 to 2 the entire table becomes fully consistent. The corrections for real data are done on the basis of professional experience and case knowledge by examining all three criteria involved.

An acceptable threshold of inconsistency is 0.33 because it means that one judgment is not more than two grades of the scale 1 to 5 away (an off-by-two error) from the remaining two judgments. There was no need to continue decreasing the inconsistency, as only its high value is harmful; a very small value may indicate that the artificial data were entered hastily without reconsideration of former assessments.