31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Identifying relative Cut-Off Scores with Neural Networks for Interpretation of the Minnesota Living with Heart Failure Questionnaire

Hassan Behlouli PhD[8], Deborah E. Feldman PhD[1,2], Anique Ducharme MD MSc[3], Marc Frenette MD[4], Nadia Giannetti MD[5], François Grondin MD[7], Caroline Michel MD[6], Richard Sheppard MD[7], Louise Pilote MD PhD[8]

ABSTRACT BACKGROUND: Quality of life (QoL) is an important end point in heart failure (HF) studies. The Minnesota Living with Heart Failure questionnaire (MLHFQ) is the instrument most widely used to evaluate QoL in Heart Failure (HF) patients. It is a questionnaire containing 21 questions with scores ranging from 0 to 105. A best cut-off value for MLHFQ scores to identify those patients with good, moderate or poor QoL has not been determined. **OBJECTIVE:** To determine a cut-off score for the MLHFQ based on the neural network (NN) approach. These cut-off scores will help discriminate between HF patients having good, moderate or poor QoL. **METHODS:** This research was carried out in the context of a longitudinal cohort study of new patients attending specialized HF clinics in six participating centers in Quebec, Canada. Patients completed a questionnaire that included the MLHFQ. In addition to this scale, self-perceived health status and clinical information related to the severity of HF were obtained including: the New York Heart Association (NYHA) functional class, 6 minute walk test and survival status. We analyzed the database using NN and conventional statistical tools. The NN is a statistical program that recognizes clusters of MLHFQ and relates similar QoL measures to one another. Among the 531 eligible patients, 447 patients with complete questionnaires were used to build randomly two sets for training (learning set) and for testing (validation set) the NN. **RESULTS**: Participants had a mean age of 65 years and 24% were women. The median MLHFQ score was 45 (inter-quartile range: 27 to 64). NN identified 3 distinct clusters of MLHFQ that represent the full spectrum of possible scores on the MLHFQ. We estimated that a score of < 24 on the MLHFQ represents a good QoL, a score between 24 and 45 represents a moderate QoL, and a score > 45 represents a poor QoL. Validation with the different severity measures confirmed these categories. These cut-offs allowed us to reach a good total accuracy (91%). These cut-offs were strongly correlated with survival status (p= 0.004), self-perceived health status (p=0.0032), NYHA functional class (p<0.0001) and standardized 6 minutes walk test (p=0.05) **CONCLUSION:** The identification of three levels of MLHFQ should be useful in clinical decision making.

Manuscript received March 19, 2009
1 University of Montreal, School of Rehabilitation, 2 Public Health Department of Montreal, 3 Montreal Heart Institute, University of Montreal, 4 Hôpital du Sacré-Cœur de Montréal, 5 Royal Victoria Hospital, McGill University Health Centre, 6 Hôpital Hôtel Dieu de Lévis, 7 SMBD Jewish General Hospital , 8 Divisions of Internal Medicine and Clinical Epidemiology, McGill University Health Centre

I. INTRODUCTION

In many areas, modern medicine has gone beyond the point of saving lives to that of improving life. This has led to the development of tools capable of measuring quality of life (QoL), mainly through self-assessed questionnaires. These questionnaires examine the impact of disease and treatment on an individual's emotional, social, and physical well-being [Fletcher, 1988]. Two broad categories of QoL questionnaires exist: generic and disease-specific. Generic measures have been developed for use with a wide range of clinical populations whereas disease-specific measures include items directly related to a medical condition. These questionnaires are useful in reflecting change in a specific patient, but the problem remains: How do we define levels of QoL? Are tertile or quartile cut-off scores useful and efficient for QoL categorization? Numerical values allow a precise evaluation of patient change. However, the use of clinically relevant categories may facilitate interpretation of a QoL measurement for a given patient. Categorization of QoL scores may also facilitate the use of QoL scores as decisional factors in the implementation of treatment.

The MHLF questionnaire is a disease-specific Qol questionnaire developed to measure the effects of heart failure and treatments for heart failure on an individual's quality of life [Rector 2005]. MLHF scores are associated with both Six-minute walk test (6MWT), or NYHA functional class [Rector 2005]. However there are no gold standards to determine when an individual's quality of life has truly changed to help define cutoff values for improvements and deterioration. A single best cutoff value for MLHF scores to identify those who definitely felt better or worse has not been determined [Rector 2005]. We propose to identify cutoff scores by using NN to categorize the data. In this manner, we hope to present a clinically meaningful definition of whether a score on the MLHF represents better or worse QoL. NN are an important class of pattern classifiers [Rumelhart, 1986] with growing acceptance in medical and biological research. There are several circumstances where neural computing techniques provide attractive alternatives to conventional software, particularly when there is a degree of uncertainty or when explicit knowledge of the domain is not available [Behlouli et al., 1996]. In recent years, NN have been widely applied in various medical fields [Snow et al., 1994; Dvorchie et al. 1996; However, to the best of our knowledge, they have been applied in the data analysis of QoL in only a few studies [Krongrad et al., 1997; Corcos et al., 2002].

II. Materials and Methods

This research was carried out as a sub-analysis of a large longitudinal cohort study (Access-Clinic) of patients newly referred to specialized CHF (congestive heart failure) clinics [Feldman to be published]. Patients were followed for 12 months at 6-month intervals: time 0 or date of patient recruitment, time 1 or 6 months from time 0, and time 2 or one year from time 0. The aim of the study was to determine whether there were gender differences in access, management, and outcomes for patients newly enrolled in specialized CHF clinics. Patients were recruited by research nurses for the longitudinal cohort study in 6 different centers in Quebec, Canada, between 2004 2007.

Clinical characteristics of 531 patients were extracted from the common software program/patient database entitled (''Vison Cardiologie''), and patients responded to a questionnaire at each of the three times. Clinical information extracted at the time 0 from the database for the purposes of this study included sex patient's age, NYHA Functional Class, left ventricular ejection fraction (LVEF), survival status, the 6MWT, and standardized to 6MWT (Percent Predicted value Troosters : PPVT) [Balashov 2008]. The questionnaire included socio-demographic information, disease history (including initial CHF diagnosis), referral history and health services utilization, self-reported health status, and the MLHFQ.

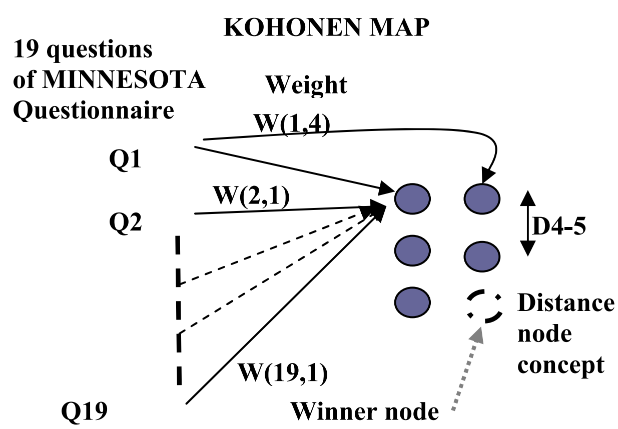
The MLHFQ is a 21-item questionnaire that includes 8 items on physical aspects, 6 on emotional aspects, and 7 other items. It measures the patient's perceptions regarding how CHF symptoms impact on their life during the preceding month. Each item is graded on a scale of 0 to 5, with the resultant global summed score ranging between 0 and 105 [RECTOR 2005]. Higher scores indicate a lower health related quality of life.

Statistics

We used descriptive statistics to compare the continuous scores of MLHF questionnaires with patient characteristics such as age, sex and different measures of severity of HF. In order to learn to recognize clusters of MLHFQ data and to relate similar QoL to each other, we employed the NN approach as per Kohonen's method also called a Self-Organizing Map (SOM). An unsupervised learning paradigm is used in the algorithm. In Fig. 1 the structure of the Kohonen neural network applied in the present study is presented. It is composed of input nodes (the 19 questions). Every input node is associated with all 6 nodes of the processing layer. A weight (Wij) value is given to the connection from an input node to a processing node.

The weight values are adaptively changed during a learning process, in which the network learns properties of the input data from the learning set of cases according to the principle of self-organizing. The nodes of the processing layer are in the grid. Each has close neighbors as well as those that are further away. This neighborhood affects the adaptation of weight values so that when the weight value associated with a node is adapted; its close neighbors are also adapted. During the process, the size of the neighborhood or the number of close neighbors decreases and finally reaches a minimum of one, i.e. the node itself [Kohonen et al., 1996]. In general we have two steps in the NN modeling: the learning process and the testing process. In the learning process, the nodes have learned by means of the algorithm to detect different level of QOL. When we get the best neural network, we use it to classify previously unknown cases of the test set (testing step).

Figure 1: Kohonen Neural network MAP



Among the 531 eligible patients, 447 patients with completed questionnaires were divided randomly in two sets: training step (learning set N=247) and testing step (validation set N=200).

Due to missing values, only nineteen of the 21 questions of the MLHFQ were used as inputs to the NN. Since we used most of questions of the MLHFQ, there is no impact on cut-off score determination.

We attempted, by interpreting the NN output, to find three levels of QoL that would correspond to good, moderate, and poor QoL. To establish the cutoff value we used a simple rule based on descriptive results of the identified clusters. We chose the best cutoffs according to the higher total accuracy of the confusion matrix. Finally, to validate the three levels of QoL, we examined the relationships between the three MLHF categories and other measures of severity of HF that are more familiar to clinicians. Six measures of severity were used: NYHA, 6MWT, LVEF, PPVT, self-reported health status and survival status.

As we do not have a gold-standard, we did a simple descriptive comparison of the neural network's cut-off scores with the cut-offs obtained by simple tertile categorisation. To be more general in terms of tertile categorisation fixing cut-offs, we use the ''original'' range of MLHF score (0 to 105) rather than the range present in our distribution data. We used survival status for this comparison.

We used Matlab version 7.01 to do NN modeling.

III.RESULTS

The study sample was composed of 447 participants with a mean age of 65 years. Twenty-four percent were women, and 24% were considered to have more severe disease (NYHA functional class of 3). Fifteen percent of patient died during the follow up. The median of the MLHFQ score was 45 (inter-quartile range: 27 to 64) (Table 1). Figure 2 shows the correlations between MLHFQ and other severity measures. MLHFQ was moderately correlated with PPVT (r=-0.3, p-value <.0001), 6MWT (r=-0.3, p-value <.0001) and Self-perceived health status (r=-0.4, p-value <.0001). Kohonen analysis of the 247 MLHFQ questionnaires in L resulted in three main clusters: N3, N2-N6, and N1-N4-N5. Table 2 shows the number of questionnaires (L+V) that were assigned to each cluster, the average score, standard deviations and ranges within each cluster. Neurons N2 and N6 were combined to form one cluster because ad hoc statistics showed that these two sub-clusters were not statistically different, and the clusters overlapped. The same overlapping happened with N1, N4 and N5. Once the clusters were established by the NN, an analysis of variance was performed and each cluster (N3, N2-N6 and N1-N4-N5) was found to be statistically different (P< 0.0001). Analysis of the validation set (V) with the model derived from L resulted in similar data aggregation (data not shown).

Table 1.a Descriptive results

Binary variables		N	percent
Sex	female	106	24
NYHA	Class		
	1	79	18
	2	256	58
	3	111	24
Survival status	alive	381	85

Table 1.b Descriptive results

Continuous variables	*N*	*Mean*	*Std Dev*	*Median*	*lower Quartile*	*Upper Quartile*
Age	447	65	11	66	57	75
Minnesota score	447	46	24	45	27	64
PPVT	380	47	14	48	38	58
LVEF	442	29	13	25	20	35
SPHS*	430	57	20	30	45	70

***SPHS: self-perceived health status**

Table 2: NN cutoff finding and descriptive statistics of the three clusters identified

L+V / Neurons clusters	*N3*	*N2-N6*	*N1-N4-N5*
Number of patients	114	116	217
Average Score	15	38	67
Std Dev	8	7	14
Max	30	51	100
Min	0	21	40

Based on the simple rule: " (average score + standard deviations) on cluster's neurons", we estimated that a score less than 24 on the MLHFQ would be representative of good QoL, between 24 and 45 would be representative of a moderate QoL, and a score greater than 45 would be indicative of poor QoL.

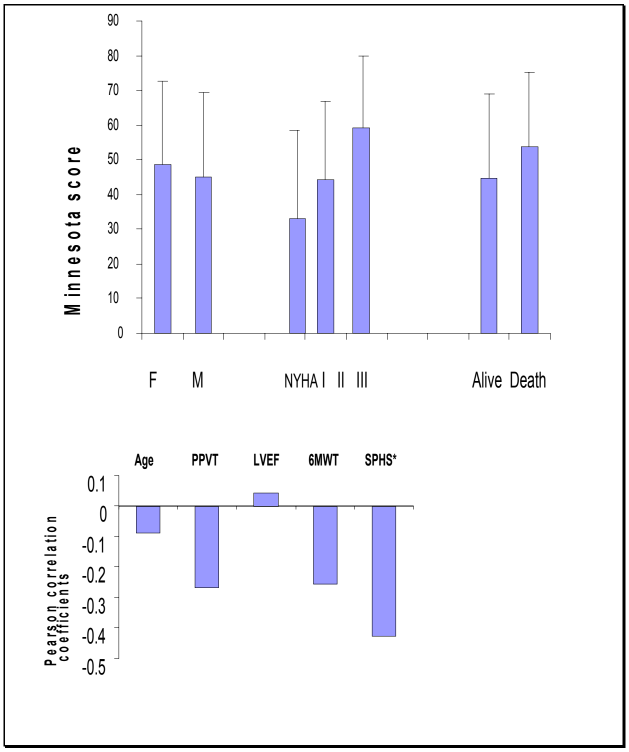
Using these definitions, a confusion matrix (Table 3) shows the number of questionnaires assigned to each class. This provides an indication of how well the Kohonen model can discriminate between different levels of QoL.

Table 3: Confusion matrix for the whole database

Cutoff score / NN class	*N3*	*N2-N6*	*N1-N4-N5*
Score<24	**94**	1	0
24<=Score<46	20	**104**	9
Score>=46	0	11	**208**
			Total accuracy=91%

We compared the Kohonen model, the continuous MLHFQ score and a tertile classification of MLHFQ with respect to survival status. Both a NN and continuous MLHFQ score were associated to survival status (p-value <0.05), however tertile cutoffs were not (Table 4). These cut-offs were strongly correlated with self-perceived health status (p=0.0032), NYHA functional class (p<0.0001), standardized 6 minutes walk test (p=0.05) (data not shown).

Figure2. Minnesota score vs. diverse clinical measures



***SPHS: self-perceived health status**

Table 4: Comparison NN cutoff and tertile cutoff versus continuous score: Survival status severity measure

Continuous score	*Survival status*	*Number of patient*	*Mean Minnesota score*	*Std Dev*
Minnesota score	*Alive*	*381*	*44.6*	*24.5*
Minnesota score	*Dead*	66	54.0	21.4

Pvalue< 0.01

NN class	*Survival Alive*	*Status Dead*
POOR QOL	174	43
MODERATE QOL	100	16
GOOD QOL	107	7
		Pvalue<0.01

Tertile class	*Survival Alive*	*Status Dead*
POOR QOL	68	13
MODERATE QOL	168	37
GOOD QOL	145	16
		pvalue=0.1

In this analysis only 19 of the 21 questions of Minnesota questionnaire were used as input of NN, due to missing data on two questions. However we did a sensitivity analysis using all the 21 questions. Missing answers were replaced in one experience by 0 and in the second experience by 5. The NN gave us similar cut-offs.

IV. DISCUSSION

The advantage of using Kohonen networks is their ability to highlight similarities between classes, in this case, levels of QoL. Because they are not based on linear correlations, they are capable of identifying similarities that are not evident with other statistical methods. In this study input to the NN consisted of patient answers to 19 questions on the MLHFQ. The results of the Kohonen based analysis show that the MLHFQ score can be categorized by three main classes. The developed Kohonen network can now be used to find the most suitable class for new investigation. It is important to note that the range of scores of the questionnaires in N3 did overlap with those in any other cluster. N2-N6 and N1-N4-N5 were the ‘‘neurons’’ sub-clusters for which the range of scores overlapped, and therefore, they were combined to form one cluster, respectively referred to as the N2-N6 cluster, and N1-N4-N5 cluster. This overlapping obeyed the distance node concept. This overlapping led us to perform a post-processing analysis based on the average, standard deviation of each cluster identified by the NN to define good, moderate, and poor QoL and allowed us to reach a good total accuracy (91%). Besides being statistically different from each other, the three clusters represent the full spectrum of possible scores on the MLHFQ.

After the identification of clusters, validation analysis was performed to verify the capability of the model to assign a questionnaire to the appropriate category. From this analysis, and by using together the learning and validation set, 101 of 119 questionnaires were assigned to the ‘‘good’’ category, 93 of 112 were assigned to the ‘‘moderate’’ category, and 204 of 216 were classified as having ‘‘poor’’ QoL. We suggest that a patient admitted to specialized CHF clinics whose score is less than 24 on the MLHFQ has a good QoL. Only 4 patients were misclassified. Similar observations were noted with the worst QoL category, a cut-off score of >45 that had only 7% misclassification. The CHF patient with moderate QoL (cutoff between 24 and 45) presented a higher rate of misclassification (24%), consequently we advise caution when interpreting scores that fall into the ‘‘moderate’’ QoL category.

It is hard to evaluate a cut off score in the absence of gold–standard data; however when we compared NN cutoff to tertile cutoff categorization, the survival status (pvalue=0.1) was not well correlated. The NN cutoff correlations were quite satisfactory for this measure of severity. However, the NN use a more restricted range than tertile values. The relationship between NN cutoff scores and NYHA versus tertile cutoff scores and NYHA are also debatable. In fact, a study [Kubo 2004] that used a new standardized questionnaire to classify patients into NYHA classes reported class I, class II and class III patients had mean MLHFQ scores of 14, 34 and 57 respectively. Others studies [Rose 2001] established mean MLHF scores to 20, 37 and 57 in class I, II and III patients respectively. We believe that these values are more related to NN cutoff (<24, 24 to 45 and >45) than to tertile cutoff values(<=35, 36 to70 and >70) mainly for the classes I and II.

To further validate the model, MHLFQ score categories could be compared with another questionnaire such as generic questionnaire MOS 36-item Short-form health Survey [Ware JE et all, 1993] [Ni H et all, 2000] or other specific CHF questionnaire.

V. CONCLUSIONS

As far as we know, our study is the first to report in a cohort of patients with CHF a fixed cutoff value for MLHFQ scores to identify those who had better, moderate or worse quality of life. The MLHFQ can be completed easily and rapidly by most patients.

Through the ‘‘clustering’’ of data in the learning set (L), competing NN identified three levels of QoL. Then, through the validation (V) phase, an individual patient’s score was compared with these profiles and associated with a particular level of QoL. This allowed us to identify that an MLHFQ score less than 26 was indicative of good QoL, between 26 and 45 was considered moderate QoL, and greater than 45 reflected poor QoL. Treatment decisions taking into consideration QoL may be important. Refinement of these models may allow us to eliminate or decrease the gray zone of moderate QoL and determine a more precise cut-off in QoL for patients admitted to specialized CHF clinics.

REFERENCES

[1] H. Behlouli, M. Miquel, J. Fayn, P. Rubel, Towards self-improving NN based ECG classifiers. Proc 18th Ann Int Con IEEE Engineering in Medicine and Biology Society, Amsterdam [CD-ROM], 1996.

[2] J. Corcos, H. Behlouli, S. Beaulieu, Identifying cut-off scores with neural networks for interpretation of the incontinence impact questionnaire. Neurourol. Urodyn., vol. 21, pp. 198-203, 2002.

[3] I. Dvorchic, M. Subotin, Performance of multi-layer feedforward neural networks to predict liver transplantation outcome. Met. Inform. Med., vol, 35, pp. 12-18, 1996.

[4] D.E. Feldman, C. Thivierge, L. Guerard, V. Dery, C. Kapetanakis, G. Lavoie, et al. Changing trends in mortality and admissions to hospital for elderly patients with congestive heart failure in Montreal. CMAJ, vol. 165, no. 8, pp. 1033-1036, 2001.

[5] D. Ehrmann Feldman, A. Ducharme, M. Frenette, N. Giannetti, R. Sheppard, F. Grondin, et al. Factors related to time to admission to specialized multidisciplinary clinics in patients with congestive heart failure. Can J Cardiol., to be published.

[6] A.E. Fletcher, Measurement of quality of life in clinical trials of therapy. Recent Results Cancer Research, vol. 111, pp. 216-30, 1998.

[7] T. Kohonen, E. Oja, O. Simula, A. Visa, J. Kangas, Engineering applications of the self-organizing map. Proc IEEE, vol. 84, pp. 1358-83, 1996.

[8] K. Balashov, D. Ehrmann Feldman, S. Savard, S. Houde, M. Frenette, A. Ducharme, et al., Percent predicted value for the 6-minute walk test: using norm-referenced equations to characterize severity in persons with CHF. J. Card. Fail., vol. 14, no. 1, pp. 75-81, Feb. 2008.

[9] A. Krongrad, L.J. Granville, M.A. Burke, R.M. Golden, S. Lai, L. Cho, et al. Predictors of general quality of life in patients with benign prostate hyperplasia or prostate cancer. J. Urol., vol. 157, pp 534-538, 1997.

[10] S.H. Kubo, S. Schulman, R.C. Starling, et all. Development and validation of a patient questionnaire to determine New York Heart Association classification. J. Card. Fail., vol. 10, pp. 228-235, 2004.

[11] H. Ni, W. Toy, D. Burgess, K. Wise, D. Nuaman, K. Crispell, et al., Comparative responsiveness of short-form 12 and Minnesota Living with Heart Failure Questionnaire in patients with heart failure. J. Card. Fail., vol. 6, no. 2, pp. 83-91, 2000.

[12] T. Rectors, Overview of the Minnesota Living with Heart Failure Questionnaire, 2005. http://www.mlhfq.org/_dnld/mlhfq_overview.pdf

[13] E.A. Rose, A.C. Gelijns, A.J. Moskowitz, et all. Long-term use of a left ventricular assist device for end-stage heart failure. N. Eng. J. Med., vol. 345, pp. 1435-1443, 2001.

[14] D.E. Rumelhart, Parallel distributed processing. Cambridge: MIT Press, vol. 1, pp. 77-109, 1986.

[15] P.B. Snow, S.S. Deborah, J.C. William, Artificial neural networks in the diagnosis and prognosis cancer: a pilot study. J. Urol., vol. 152, pp. 1923-1926, 1994.

[16] J.E. Ware, K.K. Snow, M. Kosinski, B. Gandek, SF-36 health survey manual and interpretation guide. Boston, MA: The Health Institute, New England Medical Center, 1993.