Aspects of Certainty in Patient Classification using A Health-Related Quality-of-Life Instrument in Inflammatory Bowel Disease

Urban Bergquist¹, Ankica Babic¹,², Ph.D.
1 Medical Informatics, Department of Biomedical Engineering, Linkoping University, Sweden
2 Faculty of Electrical Engineering, University of Ljubljana, Slovenia

The study has focused on deriving a certainty measure for the classification of disease activity in patients suffering from inflammatory bowel disease (IBD). The aim is to build an Internet-based health-related quality-of-life (HRQoL) questionnaire to continuously monitor a patient’s condition. Data from 109 patients was collected four times in intervals of three months, using a standardized disease-specific quality-of-life questionnaire, the Rating Form of IBD Patient Concern (RFIPC), extended with 11 additional questions. Correlation analysis showed that the RFIPC items along with “general wellbeing” were highly correlated (significance < 0.001). Factor analysis confirmed this high correlation and only one factor was identified among those variables. Multivariate discriminant analysis was successful to 78.1% in classifying between cases of remission and relapse. Implementation of a smooth threshold function decreased the classification error. However, discrimination regarding change in disease activity over time has to be further improved.

INTRODUCTION

The Rating Form of IBD Patient Concern (RFIPC) is a standardized quality-of-life questionnaire developed by Drossman et al. [1]. It is a disease specific questionnaire and contains 25 items of concern related to IBD. Questions are formulated as: “Because of your condition, how concerned are you with …?” The answer is given using a visual analog scale ranging from 0 to 100 (0 = Not at all, 100 = A great deal). A total score is calculated as the average of all concerns. The RFIPC has in previous studies proved very successful in classifying patients with inflammatory bowel disease [1,2]. The composite questionnaire used here has previously been found to have a good discriminating ability between patients in remission and in relapse, but not for changes in disease activity over time for an individual patient [3,4]. However, tracing the disease activity over time for an individual patient is a key issue for our purposes.

A Web-based questionnaire will be implemented to monitor continuously the condition of patients suffering from inflammatory bowel disease. The aim of this study has been to investigate the possibility of implementing an artificial intelligence to handle the incoming data. Using this approach the patient could instantly get feedback from the questionnaire. If necessary, an alert could simultaneously be sent to a treating physician. We believe that feedback and other forms of interactivity are important for motivating patients to take time to fill out the questionnaire. To get a response from the questionnaire would in addition provide a sense of security for the patients.

When presenting the classification result to the patient and the treating physician we would like to be more informative than only providing the results sick or healthy, in this case relapse or remission. For the patient, statements like “It might be good to contact your clinic” might be appropriate, while for a physician, a classification presenting an approximate certainty of the given classification might provide useful additional information. This study will focus on deriving this certainty measure, with the intention to implement later the patient’s recommendation based on it.

MATERIAL

The material was collected in Linkoping, Sweden, in 1993 – 1994 among IBD patients, diagnosed for ulcerative colitis, frequenting Linkoping University Hospital. The 109 participating patients answered the questionnaire at four different occasions, each separated by three months. The patients in relapse are few compared to those in remission (Table 1), which gives an unbalanced learning set. This distribution probably does not reflect the overall prevalence of relapse cases in a normal population of IBD patients [4]. The invalid cases are due to the occurrence of missing data.
Table 1. The data material.

The RFIPC with 11 additional items of concern (Table 2) was used. Those additional items were included to assess the effect of medication on quality of life. Their discriminative value was evaluated by comparing them with the standardized RFIPC items. The data consists of answers to the 36 questions in the questionnaire, no additional information, such as age, gender, psychosocial function etc., has been available. The RFIPC consists of 25 items regarding bowel and systemic symptoms as well as social and emotional functioning.

Table 2. The 11 items of concern added to the RFIPC.

<table>
<thead>
<tr>
<th>Wellbeing</th>
<th>Information</th>
<th>Complications due to medication</th>
<th>Experience of taking medicine</th>
</tr>
</thead>
<tbody>
<tr>
<td>General well-being</td>
<td>Regarding the disease</td>
<td>Stomach complications</td>
<td>Life restriction</td>
</tr>
<tr>
<td>Feeling healthy and fine</td>
<td>Regarding the medication</td>
<td>Rectum complications</td>
<td>Freedom restriction</td>
</tr>
<tr>
<td>Feeling completely disabled</td>
<td>Complications</td>
<td>Experience of taking medicine</td>
<td>Difficulties remembering to take the medicine</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Feeling tied up</td>
</tr>
</tbody>
</table>

**METHODS**

**Correlation analysis**

Correlation analysis is a basic statistical method that investigates the relationship between the values of different items.

**Factor analysis**

In previous studies factor analysis has been applied, often with great success, and therefore it appeared to be a very suitable method for this study. Factor analysis identifies principal components in the material by solving either the correlation or covariance matrix for eigenvectors [5,6]. The variables are then grouped after their accordance with the found eigenvectors. The eigenvectors are orthogonal and can therefore be seen as the different dimensions of the data. When evaluating the material, scores are computed for each of these dimensions and these so-called factors are later used for classification.

**Multivariate linear discriminant analysis**

Multivariate linear discriminant analysis is a widely used method of classification. It searches for optimal linear combinations of the independent variables that maximize the difference between the groups given by the discrete classification variable (the dependent variable) [5,6]. The classification is performed by setting a threshold for the score that is computed from the linear combination of the independent variables.

**Capturing changes over time**

Variations over time can be studied for different purposes such as reproducibility, validity and reliability [7]. Here variations are studied to hopefully improve the classification result. For this study we derive the mean value and standard deviation for each patient’s discriminant score. Two new scores are then computed, one is the discriminant score with the patient’s mean value subtracted and the other is the standardized Z-score [8].

The basic assumption is that at times of relapse the score is likely to be significantly different from the score obtained at during remission. A confounding factor is a patient’s ability to cope and to exaggerate sometimes or, conversely, to underrate concerns. The patient’s mean condition over follow-up time could be found as the mean value of the measured scores. In a similar way, the standard deviation illustrates the patient’s attitude when using the scale.

**Considering the aspect of uncertainty**

Classification using the above-mentioned linear discriminant analysis results in a prediction of disease activity. With a traditional discrete threshold the answer given is either remission or relapse, without any indication of the certainty of the prediction. A smoother threshold function can, with an appropriate adjustment of parameters, be interpreted as an approximate indication of the certainty of the prediction. This reasoning is based on the assumption that a discriminant score further away from the threshold limit signifies a more certain classification.

Therefore we introduce a sigmoid function (Figure 1), whose output is a value between 0 and 1, where 0 means remission and 1 means relapse. Any value in between signifies a certain degree of uncertainty. A value of 0.5, for instance, means equal chances of remission and relapse.
the smoothness should be adapted same as the threshold limit has been shown especially to low correlation with the first 26 items, which indicates that they capture additional aspects of HRQoL.

Factor analysis
Most studies that have evaluated the Rating Form of IBD Patient Concern (RFIPC) have concluded that the questionnaire possesses a number of different dimensions or indices, like impact of disease, sexual intimacy, complications of disease and body stigma [1,2]. However, in this material, factor analysis was unable to distinguish any factors in the RFIPC questions (one single factor) and no interpretation of answers into different aspects of patient concerns is therefore possible. In the last 10 questions however, four factors were found, but when investigating further these groups proved to have a poor discriminating ability for disease activity, apart from stomach complications due to the medication. We therefore assume that they capture other aspects of the patient’s condition.

Multivariate discriminant analysis
The discriminant function consisted of 6 variables, financial difficulties, pain of suffering, loss of bowel control, passing the disease onto others, having access to quality medical care, the uncertain nature of your disease and difficulties remembering to take my medication, and classified the disease activity with as much as 92.9% of cross-validated grouped cases correctly classified. This was however the result of a very unequal a priori probability (91.5% to 8.5%) between the two groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>Predicted Group</th>
<th>Membership</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis</td>
<td>Remission</td>
<td>Relapse</td>
</tr>
<tr>
<td>Remission</td>
<td>332</td>
<td>7</td>
</tr>
<tr>
<td>Relapse</td>
<td>19</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 2. Classification result from multivariate linear discriminant analysis using all cases.
To have a more balanced data set, remission cases were removed at random to get a material with 35 remission and 29 relapse cases. The variables of the discriminant function now changed somewhat and included “loss of bowel control”, “passing the disease onto others” and “stomach complications regarding the medication”. The latter two very negatively correlated to disease activity. The classification was performed. The 25 questions from the RFIPC along with general wellbeing showed very high correlation (significance < 0.001). Three items (have received enough information regarding the disease, have received enough information regarding my medication and difficulties remembering to take my medication) showed especially low correlation with the first 26 items, which indicates that they capture additional aspects of HRQoL.

RESULTS

Correlation analysis
In order to investigate the relationship between certain variables, a correlation analysis was

\[ f(x) = \frac{1}{1 + e^{-A(x-B)}} \text{ where } A \text{ affects the smoothness, B is the threshold limit.} \]

The threshold limit, “0.5-limit”, should essentially be the same as for a discrete threshold function, while the smoothness should be adapted to the overlap between the groups. Well separated groups can be classified using a conventional step function (a special case of the sigmoid function) while overlapping groups (which are common in clinical practice) are preferably classified with a smoother sigmoid function.

![Figure 2. The appropriate smoothness depends on the distribution of the data.](image)

The choice of parameters for the sigmoid function here was done ad hoc and requires some fine-tuning to reach an optimum. Both the smoothness and the threshold limit has to be adjusted to the actual distribution in the material. The output should reflect the Bayesian probability of the classification, but further theoretical analysis is a subject for future research.

The interesting aspect of the approach is that it results in a subtle output that highlights the presence of a great deal of uncertainty in the classification. Understanding the degree of uncertainty is obviously an important factor for correctly interpreting the classification result and therefore making a correct decision.
successful with 78.1% of cross-validated grouped cases correctly classified.

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Predicted Group Membership</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remission</td>
<td>Remission</td>
</tr>
<tr>
<td>Relapse</td>
<td>Remission</td>
</tr>
</tbody>
</table>

Table 3. Classification result from multivariate linear discriminant analysis using equally sized patient groups.

The conclusion is that the classification was successful in the well-balanced patient material.

We have also taken a mean RFIPC score, as suggested by Drossman et al. [1], but found it less discriminative. The classification result was 68.1% of cross-validated grouped cases correctly classified for the balanced material.

Capturing changes over time
Z-score was calculated by subtracting the patient’s mean value from the discriminant score and then normalizing with the standard deviation. The other score was computed by simply subtracting the patient’s mean value. We expected better results when using these measures, but the result was in fact the opposite. Despite the fact that the HRQoL instrument that we used was disease-specific it showed a lack of sensitivity to changes of disease activity over time [9].

Considering the aspect of uncertainty
Multivariate discrimination in material of 367 complete and valid cases gave poor results and the sigmoid function can not change this fact (Figure 3).

![Figure 3. Classification of the unbalanced data set.](image)

The classification of the balanced data sets (64 valid cases) was successful. The erroneous predictions tend to be located near the threshold limit (Figure 4).

![Figure 4. Classification using the balanced data set.](image)

So does this approach decrease the prediction error? A common definition of the prediction error is the mean squared error:

\[
e = \frac{1}{N} \sum_{i=1}^{N} (d_i - p_i)^2
\]

where \( d \) is the diagnosis, \( p \) is the prediction.

The error when using the conventional discriminant analysis was 0.27 compared to 0.17 for the sigmoid function. This indicates that the sigmoid function reflects well the certainty of the classification. Classification close to the threshold is likely not to be reliable, therefore it is better to accept decisions with some uncertainty.

**DISCUSSION**

Classification of patients with respect to disease activity was more successful in a smaller amount of material with data subsets of balanced sizes. This falls into line well with other experiences of learning from clinical data [10]. Considering more patient aspects could perhaps result in an even better classification. It has been suggested that this HRQoL instrument captures more than the disease activity [11]. For instance, psychosocial function proved to be an important factor when evaluating the relationship between HRQoL and disease activity [12].

A smooth non-linear threshold has a strong advantage since it opens a decision space for the physician. Instead of being presented a definite decision, the physician can interpret the output. If the expert is aware of the actual uncertainty in a classification an appropriate action can be taken. Presentation of uncertainty in terms of percentages or ratios is easily understood. We believe that patient recommendations can also be based on this measure. A notion of urgency can be developed where different levels of probabilities correspond to different levels of urgency. This, as well as how to formulate patient recommendations, has to be evaluated further, preferably in close cooperation with experts. Patient recommendations is in reality a demanding task,
especially when it comes to personal issues or medical complaints [13]. An example of Web-based patient recommendations can be found on [14].

This study has given us valuable knowledge on how to proceed with the implementation of the Web application. The Web questionnaire will continuously collect information and store it in a database, which will provide valuable research material for future studies. The Web implementation will be a cost-effective way to collect the material since the administrative requirements are minimal. The intention is to collect this complementary material while running the system and then conduct a more thorough evaluation of all components. This will be followed by necessary adjustments of the software.

Continuous monitoring with feedback will be beneficial for the patients. It might ensure a feeling that they are continuously cared for. This feeling of increased safety is likely to be appreciated. By identifying patients' concerns, physicians can address them through counseling and education to improve the quality of life of these patients [2]. It was also proven that RFIPC scores are correlated with medical resource utilization, which indicates that addressing patients' concerns may decrease health care costs.

We believe that a Web implementation is the optimal solution for our purpose. It is affordable, quick and there are possibilities of implementing interactive and flexible communication with patients. The response from the questionnaire is instantaneous and alerts can immediately be sent to the treating physician. A Web questionnaire might dissolve some patient discomfort of disclosing psychosocial concerns, dissatisfaction and other problems that can occur in a face-to-face doctor-patient interaction. In addition an Internet connection is common nowadays in Sweden as well as affordable for most people. Evaluation of the acceptance of the Internet among IBD patient groups is the next step.

The aspect of giving wrong recommendations has to be evaluated further, but the implementation of the sigmoid threshold function does not increase the risk of error. In fact it was proven that giving uncertain answers for uncertain cases using the sigmoid threshold function can somewhat reduce the error.

Acknowledgements
Authors would like to acknowledge Drs Associate Professor Sven Almer, Ph.D., Magnus Ström, Ph.D., Henrik Hjortswang, and Associate Professor Göran Bodemar, Ph.D., for the research initiative, the proposed HRQoL instrument and patient material collected during their research project.

References