Original Paper

Classification and Regression Tree and Computer Adaptive Testing in Cardiac Rehabilitation: Instrument Validation Study

Linda Peute[1], BSc, MSc, PhD; Thom Scheeve[2], BSc, MSc; Monique Jaspers[1], BSc, MSc, PhD

[1]Center of Human Factors Engineering of Health Information Technology, Department of Medical Informatics, Amsterdam Institute of Public Health, Amsterdam University Medical Centers, Amsterdam, Netherlands

[2]Signal Processing Systems, Video Coding and Architectures, Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands

Corresponding Author:
Monique Jaspers, BSc, MSc, PhD
Center of Human Factors Engineering of Health Information Technology, Department of Medical Informatics
Amsterdam Institute of Public Health
Amsterdam University Medical Centers
J1b-116
Meibergdreef 9
Amsterdam, 1105 AZ
Netherlands
Phone: 31 640120422
Email: m.w.jaspers@amsterdamumc.nl

Abstract

Background: There is a need for shorter-length assessments that capture patient questionnaire data while attaining high data quality without an undue response burden on patients. Computerized adaptive testing (CAT) and classification and regression tree (CART) methods have the potential to meet these needs and can offer attractive options to shorten questionnaire lengths.

Objective: The objective of this study was to test whether CAT or CART was best suited to reduce the number of questionnaire items in multiple domains (eg, anxiety, depression, quality of life, and social support) used for a needs assessment procedure (NAP) within the field of cardiac rehabilitation (CR) without the loss of data quality.

Methods: NAP data of 2837 CR patients from a multicenter Cardiac Rehabilitation Decision Support System (CARDSS) Web-based program was used. Patients used a Web-based portal, MyCARDSS, to provide their data. CAT and CART were assessed based on their performances in shortening the NAP procedure and in terms of sensitivity and specificity.

Results: With CAT and CART, an overall reduction of 36% and 72% of NAP questionnaire length, respectively, was achieved, with a mean sensitivity and specificity of 0.765 and 0.817 for CAT, 0.777 and 0.877 for classification trees, and 0.743 and 0.40 for regression trees, respectively.

Conclusions: Both CAT and CART can be used to shorten the questionnaires of the NAP used within the field of CR. CART, however, showed the best performance, with a twice as large overall decrease in the number of questionnaire items of the NAP compared to CAT and the highest sensitivity and specificity. To our knowledge, our study is the first to assess the differences in performance between CAT and CART for shortening questionnaire lengths. Future research should consider administering varied assessments of patients over time to monitor their progress in multiple domains. For CR professionals, CART integrated with MyCARDSS would provide a feedback loop that informs the rehabilitation progress of their patients by providing real-time patient measurements.

XSL•FO
RenderX

Introduction

Background

Given the prominent role of the internet in many patients' lives nowadays, patient portals are increasingly deployed to involve patients in their care process. These portals, for example, allow fewer time-consuming consultations between patients and health care professionals by integrating batteries of questionnaires needed for diagnosis or as part of a patient's needs assessment. A precondition of this type of use of patient portals is the high quality of data to be exchanged. This is especially required when the patient portal is linked to a computerized decision support system (CDSS), used by health care providers for advice on therapy planning during the decision-making process. Extensive assessment procedures, however, may result in increased response burden on the patient, possibly resulting in low quality of response data [1-4].

There is a need for assessment procedures of shorter lengths that capture patient questionnaire data while attaining high data quality without undue response burden on patients. Computerized adaptive testing (CAT) methods, using Item Response Theory (IRT), have the potential to meet these needs and offer an attractive option to shorten questionnaires. A CAT algorithm uses information from questions already answered, to select the most appropriate question to be administered next. Therefore, a patient is offered only the fewest possible items. Chien et al [1] verified the effectiveness and efficacy of saving time and reducing the burden on patients through CAT applied on the Activities of Daily Living Scale. They found that mobile nursing services placed at the bedsides of patients could, through a CAT module, reduce the burden on patients and save time more than the traditional paper-and-pencil testing appraisals. Similarly, CAT-based administration of surveys of patient perception substantially reduced patient burden without compromising the precision of measuring patients' perceptions of hospitalization [2]. Another promising method is classification and regression tree (CART) analysis, originating from clinical decision rules research, which is mostly used to classify patients into clinically important categories [5,6]. It can be used to shorten questionnaires by selecting predictor variables (questionnaire items) that allow different questions to be identified for patients with different levels of complaints of a disease. As an illustration, Lu et al [7] successfully applied CART methods in the development of brief screening tools based on questions from existing psychiatric diagnostic instruments. Potential advantages of both CAT and CART for clinical practice are efficient testing and a reduction in the test burden in patients and, consequently, less measurement error during testing. However, as far as we know, CAT and CART performances in terms of their yield in shortening questionnaire lengths and sensitivity and specificity levels have not been compared.

In this study, we examined if CAT or CART analysis could be used to shorten the questionnaires included in the needs assessment procedure (NAP) of cardiac rehabilitation (CR) patients. CR is a therapy to support patients with cardiac issues in recovering from a cardiac incident in order to improve their physical and physiological condition [8]. To offer a patient a tailored rehabilitation plan, every patient has to complete an NAP including 80-130 questionnaire items of which answer data are sent to a CDSS. We aimed to test which method, CAT or CART, was best suited to reduce the number of questionnaire items in multiple domains (eg, anxiety, depression, quality of life, and social support) used for the NAP, without the loss of data quality.

Cardiac Rehabilitation in the Netherlands: Case Study

We used the data collected in the Dutch multicenter CARDSS program [8]. Within this program, CR clinics use a CARDSS electronic patient record (EPR) with computerized decision support (CDS) based on the most recent version of the Dutch CR guidelines [9]. The CDS provides CR professionals with advice on a patient-tailored rehabilitation program based on an NAP. The Dutch guideline requires gathering of 80-130 data items regarding a patient's quality of life, work resumption, psychological and social functioning, and lifestyle. The patient-tailored rehabilitation program can comprise four possible group-based therapies: disease-specific education; exercise training; lifestyle modification; and relaxation and stress management training, supplemented by individual counseling (eg, by a psychologist, dietician, or social worker) and, if needed, different forms of individual therapies. During the NAP, a CR professional can immediately discuss the CDS advice with the patient to set the final patient-tailored rehabilitation plan. To improve the efficiency of the NAP data gathering process, we developed an electronic portal for patients, called MyCARDSS, that patients can use to enter their data, either at home or at the CR clinic. Some patients fill in the NAP in MyCARDSS just before their consultation at the CR clinic, as they need the help of a nurse or do not have an internet connection at home [10]. MyCARDSS is linked to the EPR system.

Data of 2837 CR patients of this multicenter CARDSS Web-based program was included in this study. The CARDSS Web-based dataset comprised patient identification data and CR needs assessment data. We used the database of CARDSS Web-based program to obtain the scores on the individual questions of the patients' NAP data and the following demographic patient information: age, gender, diagnosis, and cardiac intervention.

Questionnaires Data

We included data from seven questionnaires used in the NAP for CR, which allow the classification of patients based on the outcome: no (low), mild (moderate), and serious (serious or high) symptoms. The following questionnaires are used in the multicenter CARDSS Web-based program:

The Dutch version of the Quality of Life after Myocardial Infarction (QLMI) is a 27-item questionnaire, scored on a 7-point Likert scale, to measure health-related quality of life for patients after myocardial infarction. It comprises 10 physical dimension items (QLMI-P), subscores ranging from 1 to 7, where subscale scores between 1.0 and 3.39 classify a patient as having a low and >4.0 as having a high exercise capacity; 7 social dimension items (QLMI-S), subscores ranging from 1 to

7, where subscale scores between 1.0 and 4.4 indicate a high, between 4.5 and 5.9 indicate a moderate, and >6.0 indicate a low risk on social dysfunctioning; and 10 emotional dimension items (QLMI-E) [11]. The QLMI-E is not used in the NAP for CR.

The Hospital Anxiety and Depression Scale (HADS) is a 14-item questionnaire, scored on a 4-point Likert scale (0-3), to detect the presence of anxiety and depression. It comprises seven anxiety items (HADS-A) and seven depression items (HADS-D). A total score is not calculated. For both subscales, scores range from 1 to 21. For both HADS-A and HADS-D, subscale scores between 0 and 4 points indicate a low, between 5 and 7 points indicate a moderate, and >8 points indicate a serious risk on anxiety [12].

The Patient Health Questionnaire–9 (PHQ-9) is a 9-item questionnaire on a 4-point Likert scale to detect the presence and severity of mental health disorders with scores ranging from 0 to 27: Scores between 0 and 4 points indicate a low, between 5 and 9 points indicate a moderate, and >10 points indicate a serious risk on depression [13].

The Generalized Anxiety Disorder-7 (GAD-7) is a 7-item questionnaire on a 4-point Likert scale to detect the presence of generalized anxiety disorders. Scores range from 0 to 21: Scores between 0 and 4 points indicate a low, between 5 and 9 points indicate a moderate, and >10 points indicate a serious risk on anxiety [14].

The Multidimensional Perceived Social Support Scale (MPSSS) is a 12-item questionnaire, scored on a 7-point Likert scale, to measure social support and specific availability and satisfaction with support from family, friends, or a special person. Scores range from 12 to 84: Scores between 12 and 64 points indicate a low, between 65 and 78 points indicate a moderate, and >79 points indicate a high level of social support [15].

The Dutch guidelines state that all patients have to fill in the QLMI-P/QLMI-S questionnaire, but CR clinics can choose between a combination of HADS-A/HADS-D or GAD-7/PHQ-9 to assess anxiety and depression levels in their patients. This means that an individual patient fills in the QLMI-P/QLMI-S and HADS-A/HADS-D or QLMI-P/QLMI-S and GAD-7/PHQ-9 questionnaires. Administration of the MPSSS is not mandatory; this questionnaire is used by merely one clinic in the CARDSS Web-based program to assess social support provided to a patient.

Methods

Classification and Regression Tree

We first used CART as a method to reduce the number of items of each questionnaire included in the NAP. CART represents a hierarchical model structured as trees for predicting continuous (regression trees) and categorical (classification trees) variables [16,17].

CART models predict a response variable (ie, outcome) based on the values of ≥1 predictors (ie, items within the questionnaire or demographic data) using a technique called recursive partitioning groups [18]. This algorithm looks for subgroups in the dataset in which the response variable is relatively homogeneous. These subgroups become the leaves of the tree [16,18]. At each node in the tree, the recursive partitioning algorithm identifies a predictor variable and a *split* by which cases may be subclassified. This predictor variable and split combination are chosen to have the greatest predictive power among all predictor split combinations at the tree node. Once the cases at the node have been partitioned by the split, the algorithm is applied to both resulting subclassifications. The bottom node of the tree reports a classification for the patient. In most cases, the constructed model is asymmetric, which means that it depends on previous answers on items administered to a patient. For qualitative outcome variables, the resulting tree is a classification tree, and for quantitative outcome variables, it is a regression tree. To shorten the NAP overall, all individual questions of each questionnaire were entered into a CART analysis to develop a CART per questionnaire. The CART analysis selects independent items that differentiate the outcome variable but allows different combinations of the predictor variables in different subgroups, creating flexible questionnaires [16,18]. CART allows the set of questionnaire items presented to a patient to be adapted to the responses already provided by him or her; going left at a node may result in a very different set of questionnaire items being presented as compared with going right. Thus, CART has the potential to shorten the length of the NAP. To determine if CART can shorten the NAP, we determined the mean, maximum, and minimum length of each CART per questionnaire included in the NAP.

We also tested if demographic or clinical variables—patient age, gender, cardiac diagnosis, and intervention—as predictors influenced the length of the CART.

Computer Adaptive Testing

The second method we used to shorten the number of questionnaire items of each questionnaire included in the NAP is CAT. The net result of a CAT is a small, optimal number of items to be administered to the patient without loss of measurement precision. CAT is based on IRT. IRT models are statistical models of the relationship between a person's score on the construct being measured and the probability of choosing each response on each item measuring that construct. IRT models can be used to evaluate how informative an item is for a specific range of scores and estimate a person's IRT score [19]. An IRT model expresses a probability (vertical axis) of the selection of each item response category as a function of the score (horizontal axis) on the underlying latent trait (the measured construct, ie, anxiety or depression in this study). To estimate the latent trait, a great number of different IRT models can be used [19]. For questions with ordered response categories, the Graded Response Model (GRM) has been proposed [19]. As the included NAP questionnaires have ordered response options, we fitted a GRM to our data to measure the item parameters of all questionnaire items. A CAT begins with an initial global question; all patients answer the same first item. On the basis of the response to the first item, the score, CI, and latent trait are estimated using maximum likelihood information [19]. The algorithm selects further items based on the highest possible information for the current latent trait score. The latent trait is estimated after each item administration based on the

accumulated information combined with the information on the new response [19]. The adaptive testing stops when a stopping rule is met. In this study, CAT stops as soon as a patient can be classified in a low, moderate, or high class with a CI of 95%. A patient had to have answered at least one item before the stopping rule was checked. With this stopping rule, the complete test cycle has a variable length, depending on the patient's individual responses and the point at which the stopping rule is applied. We did not set a minimum for item administration. Thus, theoretically, the algorithm could stop after the administration of only two or three items, given that the item is informative enough to classify a patient in the low, moderate, or high class. To determine if CAT can shorten the NAP, we determined the mean, maximum, and minimum length of each CAT per questionnaire included in the NAP.

Performance Testing

To determine the performance for both CAT and CART, the sensitivity and specificity were computed. For the regression trees, the root mean squared error (RMSE) and the normalized RMSE (NRMSE) were additionally measured. The RMSE is a measure of the differences (ie, the prediction errors) between values predicted by a regression/classification tree and those actually observed [18]. The RMSE serves to aggregate the prediction errors into a single measure of predictive power. The lower the RMSE or NRMSE, the better the model predicts.

Using the same data for calibration and evaluation of the model results in overly optimistic estimates of performance. Cross-validation is a method for validation of a procedure for model building that avoids the requirement for a new or independent validation set [20]. We, therefore, randomly split the data into two sets: a training set (2127/2837, 74.97%) for calibration of the models and a validation set (710/2837, 25.02%) for evaluating the performance of the CAT and CART. Besides, 10-fold cross-validation was divided into 10 subsets, each subset, in turn, being used to test the performance of the CAT/CART created with the other 9 subsets.

Software

All analyses were performed in R (R Foundation of Statistical Computing, Vienna, Austria), a programming language for statistical computing. Different packages were used to perform the analyses. R packages for simulating Item Response Theory based on Computerized Adaptive Tests and Latent Trait Modes packages were used for the CAT simulation. R packages for Classification and Regression Training and Recursive Partitioning And Regression Trees were used for CART. The training and test datasets were created with the *caret* package [21,22].

Results

Patient Characteristics

Demographic and clinical characteristics of the patients from the clinics participating in the CARDSS Web-based program are shown in Table 1.

Table 1. Demographic and clinical characteristics of 2837 cardiac rehabilitation patients.

Characteristics	Value
Gender, n (%)	
Male	1984 (69.99)
Female	826 (29.11)
Missing	27 (0.95)
Age (years), mean (SD)	
Men	65.8 (10.9)
Women	68.6 (11.3)
Mean	66.6 (11.1)
Diagnosis and intervention, n (%)	
ACS[a] (myocardial infarction or unstable angina pectoris) with intervention (CABG[b], PCI[c], CABGVALVE[d], or VALVESUR[e])	831 (29.29)
Chronic diagnosis (heart failure or stable angina pectoris)	648 (22.84)
Elective PCI (PCI without ACS)	604 (21.29)
Elective CABG (CABG, CABGVALVE, or VALVESUR without ACS)	404 (14.24)
ACS without intervention	350 (12.33)

[a]ACS: acute coronary syndrome.

[b]CABG: coronary artery bypass grafting.

[c]PCI: percutaneous coronary intervention.

[d]CABGVALVE: coronary artery bypass grafting in combination with heart valve surgery.

[e]VALVESUR: heart valve surgery.

Of the 2837 patients who participated in the program. 69.99% (1984/2837) were male and 29.29% (831/2837) had an acute coronary syndrome; men were younger (mean age 65.8, SD 10.9 years) than women (mean age 68.6, SD 11.3 years).

Inclusion and Exclusion of Questionnaires

Table 2 provides an overview of the total number of questionnaires fully filled out, missing questionnaires, and insufficiently filled out questionnaires. As explained previously, the Dutch guidelines state that CR clinics can choose between a combination of HADS-A/HADS-D or GAD-7/PHQ-9 to assess anxiety and depression levels in their patients. Administration of the MPSSS is not mandatory; this questionnaire is used by merely one clinic in the CARDSS Web-based program. Patient data on questionnaires were excluded from the CAT and CART analysis if (1) a questionnaire was not filled out by the patient (missing) or (2) a provided questionnaire was insufficiently filled out by a patient to calculate a total score (≥1 item responses missing).

Table 2. Number of patients who filled out, did not fill out, or insufficiently filled out the questionnaires (N=2837).

Questionnaire	Fully filled out, n (%)	Missing, n (%)	Insufficiently filled out, n (%)
Quality of Life after Myocardial Infarction - Physical dimension	2633 (92.81)	189 (6.66)	15 (0.52)
Quality of Life after Myocardial Infarction - Social dimension	2633 (92.81)	189 (6.66)	15 (0.52)
Patient Health Questionnaire - 9	1156 (40.75)	213 (7.50)	0 (0.00)
Generalized Anxiety Disorder - 7	1223 (43.11)	244 (8.60)	1 (0.04)
Hospital Anxiety and Depression Scale - Anxiety	1266 (44.62)	152 (5.35)	2 (0.07)
Hospital Anxiety and Depression Scale - Depression	1264 (44.55)	149 (5.25)	4 (0.14)
Multidimensional Perceived Social Support Scale	716 (25.23)	57 (2.00)	5 (0.17)

Classification and Regression Tree

CART models were built for every questionnaire in the NAP using the training set, resulting in a total of 28 CART models: (1) seven classification trees without the additional clinical/demographic data (ie, age, gender, cardiac diagnosis, and intervention) as features, (2) seven classification trees with the additional clinical/demographic data as features, (3) seven regression trees without the additional clinical/demographic data as features, and (4) seven regression trees with the additional clinical/demographic data as features.

The seven classification and seven regression trees with data on the clinical/demographic variables—age, gender, diagnosis, and intervention—showed no inclusion of these variables in the classification or regression trees.

The developed classification trees comprise four to six levels, with five to eight terminal nodes. The developed regression trees comprise four levels, with five terminal nodes.

Performances of Computerized Adaptive Testing and Classification and Regression Tree

To evaluate the performances of CAT and CART, patient data in the validation dataset were used. Figure 1 shows the maximum, minimum, and mean number of items administered by CART and CAT. Figure 2 displays the percentage decrease per questionnaire for CART and CAT. Table 3 lists the performances of both CAT and CART in terms of sensitivity and specificity per questionnaire. These performance measures are provided for each of the categories—low, moderate, and high—except for the QLMI-P with only low and high classes.

Figure 1. Maximum, minimum, and mean number of items administered in each CAT (computerized adaptive testing) and classification and regression tree. GAD-7: Generalized Anxiety Disorder - 7; HADS-A: Hospital Anxiety and Depression Scale - Anxiety; HADS-D: Hospital Anxiety and Depression Scale - Depression; MPSSS: Mul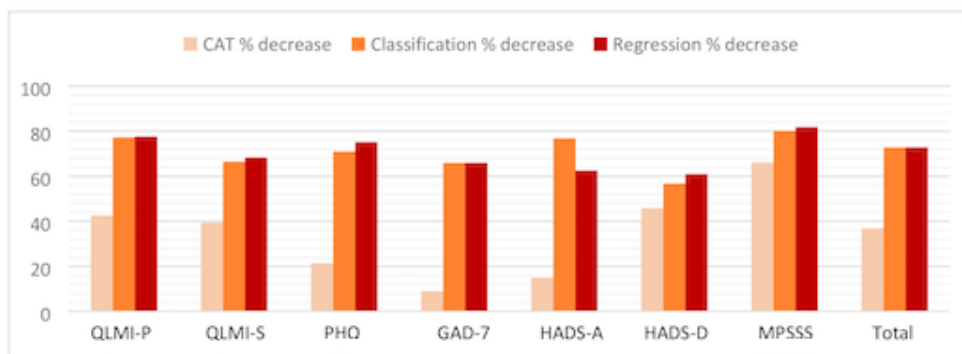tidimensional Perceived Social Support Scale; PHQ-9: Patient Health Questionnaire - 9; QLMI-P: Quality of Life after Myocardial Infarction - Physical dimension; and QLMI-S: Quality of Life after Myocardial Infarction - Social dimension.
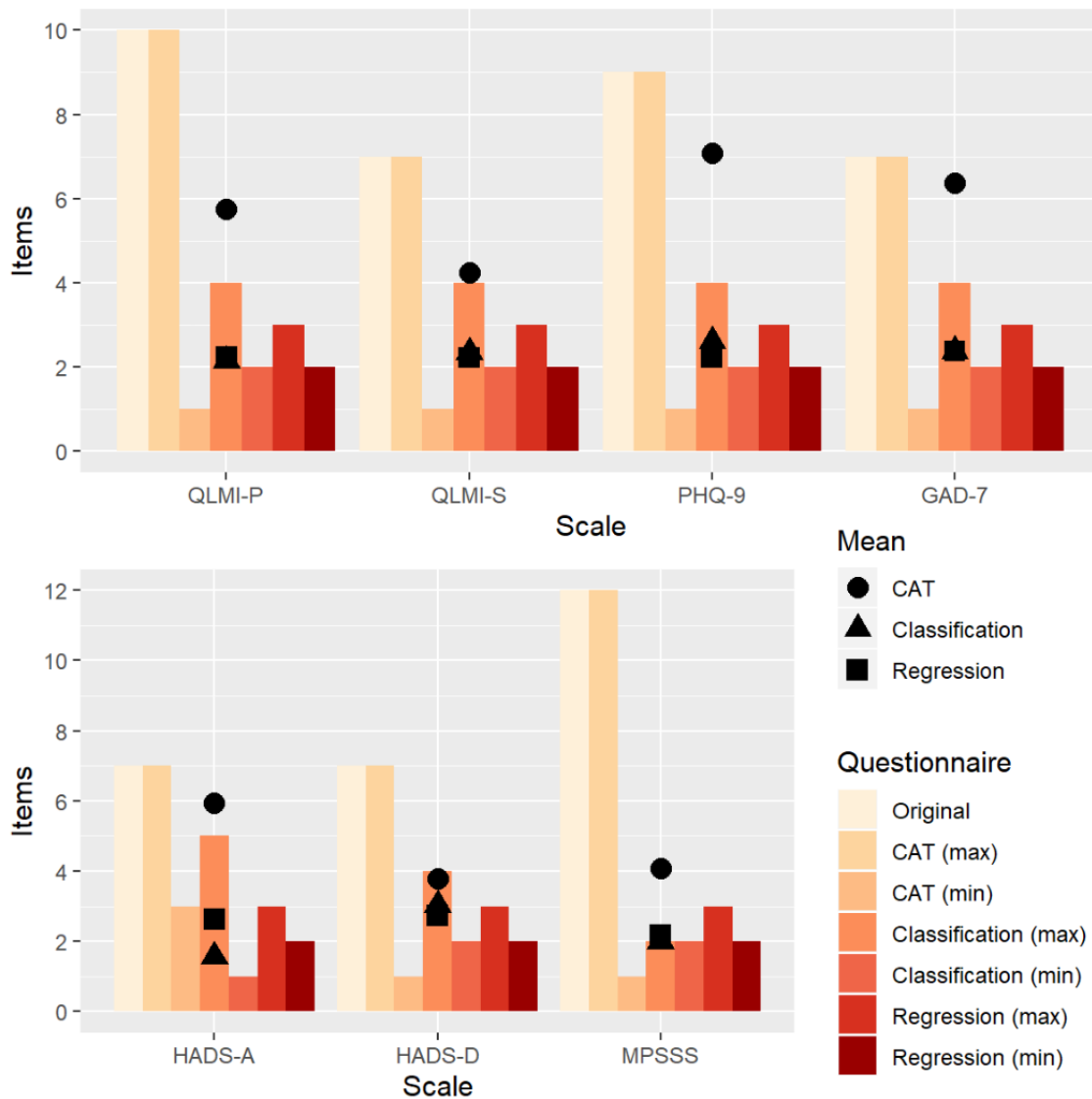
Figure 2. Percentage decrease per questionnaire for computerized adaptive testing and classification and regression tree. QLMI-P: Quality of Life after Myocardial Infarction - Physical dimension; QLMI-S: Quality of Life after Myocardial Infarction - Social dimension; PHQ-9: Patient Health Questionnaire - 9; GAD-7: Generalized Anxiety Disorder - 7; HADS-A: Hospital Anxiety and Depression Scale - Anxiety; HADS-D: Hospital Anxiety and Depression Scale -Depression; MPSSS: Multidimensional Perceived Social Support Scale.

Table 3. Sensitivity and specificity of computerized adaptive testing, classification trees, and regression trees by questionnaire (sensitivity and specificity of a class are computed as the class versus the other classes, ie, low versus moderate and high/serious, moderate versus low and high/serious, and high/serious versus low and moderate). High scores on Patient Health Questionnaire - 9, Generalized Anxiety Disorder - 7, Hospital Anxiety and Depression Scales but conversely low scores on Multidimensional Perceived Social Support Scale and Quality of Life after Myocardial Infarction - Physical and Social dimensions instrument indicate that patients report surveyed symptoms.

Questionnaire	Sensitivity			Specificity			Root mean squared error	Normalized root mean squared error
	Low	Moderate	High	Low	Moderate	High		
Computerized adaptive testing								
QLMI-P[a]	0.609[b]	N/A[c]	N/A	1[b]	N/A	N/A	N/A	N/A
QLMI-S[d]	1	0.008	0.201	0.339	0.691	1	N/A	N/A
PHQ-9[e]	0.773	0.000	0.988[f]	0.887	0.997	.0581[f]	N/A	N/A
GAD-7[g]	0.935	0.222	1[f]	0.766	0.996	0.852[f]	N/A	N/A
HADS-A[h]	0.765	0.333	0.884[f]	0.961	0.887	0.810[f]	N/A	N/A
HADS-D[i]	0.234	0	1[f]	1	1	0.126[f]	N/A	N/A
MPSS[j]	1	0	0.515	0.256	1	1	N/A	N/A
Classification trees								
QLMI-P	0.954[b]	N/A	N/A	0.763[b]	N/A	N/A	N/A	N/A
QLMI-S	0.734	0.735	0.807	0.947	0.786	0.892	N/A	N/A
PHQ-9	0.907	0.554	0.549[f]	0.852	0.843	0.922[f]	N/A	N/A
GAD-7	0.931	0.681	0.800[f]	0.869	0.921	0.964[f]	N/A	N/A
HADS-A	0.941	0.417	0.970[f]	0.969	0.970	0.879[f]	N/A	N/A
HADS-D	0.979	0.275	0.936[f]	0.938	0.960	0.793[f]	N/A	N/A
MPSSS	0.644	0.836	0.758	0.970	0.712	0.938	N/A	N/A
Regression trees								
QLMI-P	0.946[b]	N/A	N/A	0.640[b]	N/A	N/A	0.661	0.247
QLMI-S	0.849	0.648	0.755	0.919	0.787	0.871	0.608	0.184
PHQ-9	0.861	0.551	0.832[f]	0.833	0.857	0.945[f]	2.696	0.222
GAD-7	0.957	0.408	0.694[f]	0.645	0.933	0.977[f]	2.231	0.145
HADS-A	0.750	0.667	0.922[f]	0.966	0.907	0.966[f]	1.833	0.153
HADS-D	0.640	0.190	0.985[f]	0.996	0.917	0.667[f]	1.582	0.146
MPSSS	0.489	0.647	0.923	0.977	0.755	0.805	7.962	0.162

[a]QLMI-P: Quality of Life after Myocardial Infarction - Physical dimension.

[b]The quality of Life after Myocardial Infarction - Physical dimension has only a low and high class, so the sensitivity and specificity are computed as low versus high.

[c]Not applicable.

[d]QLMI-S: Quality of Life after Myocardial Infarction - Social dimension.

[e]PHQ-9: Patient Health Questionnaire - 9.

[f]Actual class is serious.

[g]GAD-7: Generalized Anxiety Disorder - 7.

[h]HADS-A: Hospital Anxiety and Depression Scale - Anxiety.

[i]HADS-D: Hospital Anxiety and Depression Scale - Depression.

[j]MPSSS: Multidimensional Perceived Social Support Scale.

Classification and Regression Tree

The minimum and maximum number of questionnaire items to be selected for the different questionnaires, based on the training set, are shown in Figure 1. For the classification trees, an average of 2.4 items per questionnaire should be administered (minimum=2 and maximum=3) to classify a patient (Figure 1), with a questionnaire reduction between 60.9% (HADS-D) and 77.6% (QLMI-P), compared with the original questionnaire (Figure 2). For the regression trees, an average of 2.4 items per questionnaire should be administered (minimum=1 and maximum=5) to classify a patient (Figure 1), with a questionnaire reduction between 56.8% (HADS-D) and 77.4% (QLMI-P), compared with the original questionnaire (Figure 2).

For all questionnaires except the HADS-A, the minimum number of questionnaire items (two items) in the classification trees equals those of the regression trees. The maximum number of questionnaire items is higher in the classification (four or five items) than in the regression (three items for all questionnaires) trees, except for the MPSSS (two items) and QLMI-P (three items). The mean number of items and the mean percentage decrease over all questionnaires in the classification trees equal those of the regression trees (2.38). A mean 72% reduction over all questionnaires in the NAP can be realized by CART.

Per questionnaire, the sensitivity and specificity levels of CARTs, classification, and regression trees are displayed in Table 3. The mean sensitivity and specificity of CART over all questionnaires and classes are 0.777 and 0.877, respectively. The mean sensitivity and specificity of regression trees over all questionnaires and classes are 0.743 and 0.840, respectively. The NRMSE of the regression trees ranges from 0.145 (GAD-7) to 0.247 (QLMI-P).

Computerized Adaptive Testing

The minimum and maximum number of questionnaire items used by CAT to be selected for the different questionnaires, based on the training set, are shown in Figure 1. With CAT, an average of 5.3 items (minimum=1 and maximum=original length) were needed per questionnaire to classify a patient (our goal was not to classify but to shorten the questionnaire), with a questionnaire reduction between 9.0% (GAD-7) and 45.8% (HADS-D), compared with the original questionnaire (Figure 2). CAT shows a smaller percentage decrease in questionnaire items per questionnaire and a smaller overall decrease, compared with CART (Figure 2). A mean 36% reduction over all questionnaires in the NAP can be realized by CAT. The sensitivity and specificity levels of CATs per questionnaire are displayed in Table 3. The mean sensitivity and specificity of CAT over all questionnaires and classes are 0.765 and 0.817, respectively.

Computerized Adaptive Testing and Classification and Regression Tree: Questionnaire Comparison

The differences in percentage decrease in questionnaire items per questionnaire between CAT and CART are highest for the PHQ-9 (56.7%), GAD-7 (56.5%), and HADS-A (54.6%).

HADS-D (13.0%) and MPSSS (14.9%) show the lowest differences in percentage decrease per questionnaire (Figure 2).

Discussion

Principal Findings

This study shows that both CAT and CART can be used to shorten the questionnaires of the NAP used within the field of CR. CART, however, showed the best performance with an overall about twice as large decrease in questionnaire items of the NAP and the highest sensitivity and specificity. Demographic/clinical variables—patient age, gender, cardiac diagnosis, and intervention—as predictors did not influence the length of the CARTs, meaning that these variables do not determine the classification of patients in the trees.

Relation to Other Studies

CAT has nearly four decades of research behind it but has only been applied more recently to health care. CAT has been used to shorten or develop questionnaires for assessment of fatigue [23], depression [24-26], suicide ideation [4], other mental health disorders [27,28], physical [29] and upper extremity functioning [30], health status in patients with knee osteoarthritis [31], activities of daily living in outpatients with stroke [32], and exposure of nurses to workplace bullying [33,34] and in patient-reported outcome measurement studies [24,29]. Overall, its application has proven to be successful in shortening questionnaires, while patient measurements remained valid and reliable. Some studies even demonstrated that by applying CAT, existing instruments for patient-reported outcomes could be improved. These new instruments reduced the questionnaire burden on patients while increasing measurement precision [29], possibly leading to reduced sample size requirements.

Similarly, half a century has passed since the publication of the first CART algorithms, but again, their application in health care is of a far more recent date.

CART has, for the most part, been used to classify (new) patients into clinically important (risk) categories, such as diabetic nephropathy [35] and colorectal adenocarcinoma [36]. CART has also been applied to define factors associated with delayed treatment of acute myocardial infarction [37] and quality of life [38].

As far as we know, CART has not been used, at least not in the health care domain, with the aim to shorten questionnaires.

To our knowledge, our study is the first to assess the differences in performance between CAT and CART for shortening questionnaire lengths. Overall, CART outperformed CAT, with a larger reduction in the length of the questionnaire for the NAP procedure and in sensitivity and specificity.

Further, we did not observe an influence of the predictors such as age, gender, diagnosis, and intervention in the construction of CARTs. CART would have probably captured interactions across many NAP questionnaire scores and these clinical/demographic variables. Our findings are in concordance with the findings of the study by Miscio et al [39] wherein the inclusion of clinical/demographic data such as age and gender

did not have an effect on the construction of CARTs for patient measurement tools.

Meaning of This Study

Ideally, an NAP procedure for CR including several questionnaires should be highly sensitive and specific, so that few patients with depression/anxiety/social complaints are missed and few without depression/anxiety/social complaints are identified as having complaints. In this context, CART is, overall, more sensitive and specific than CAT, while it shortens the NAP procedure more than CAT. CART analysis has the statistical advantage of being a nonparametric method, with no assumptions about the functional form of the data. CART might further be a good alternative to CAT, as this method not only has the ability to efficiently shorten questionnaires by segmenting patient groups into meaningful subgroups, but it also presents knowledge on these subgroups in a graphical way. These graphs provide a good understanding of how this segmentation was attained. CART, as an algorithmic rather than statistical method, further offers good insight into interactions between variables that are not revealed by linear quantitative research. But CART does not provide distributions, likelihood ratios, or CIs to quantify or support the validity of the findings. For the CATs, we, for example, made use of CIs; we stopped a CAT as soon as a patient could be classified in a low, moderate, or high class with a CI of 95%. For evaluating the performance of both the CAT and CART, we made use of cross-validation by splitting the dataset in a training set for calibration of the models and a validation set. We further applied 10-fold cross-validation on the training and validation sets to validate the generality of the CAT and CARTs.

Recommendations for Care Practice and Future Research

We demonstrated that an NAP with shortened questionnaires can be used for screening cardiac patients on rehabilitation needs without compromising its measurement accuracy. Both CAT and CART can be used for this purpose, but a CART approach with many questionnaires is rare. Obtaining the information during the joint administration of the shortened questionnaires would take about one-third of the time that it would take with the traditional fixed-length questionnaires, that is, 30 min instead of 90 min per patient. We also plan to extend the MyCARDSS portal with adaptive tests that will be administered to our CR patients over time, at their home or just before their consultation at the clinic. The results of these assessments will be interfaced with the CARDSS EPR, which is easily accessible by the CR professionals from any device. For CR professionals, CART integrated with MyCARDSS and CARDSS would then provide a feedback loop that informs the rehabilitation progress of their patients by providing these real-time patient measurements. An example for future research is to monitor our CR patients' physical and mental health progress by these varied assessments. Although dimension reduction strategies have been employed for numerous data problems, they are scantly discussed in the context of analyzing survey data. Another example for future research, thus, is to examine the performance of other methods such as dimensionality reduction techniques, for example, principal component analyses (PCA) for reducing questionnaire

lengths. Dimension reduction techniques can be an effective approach for reducing dimensionality in more complex survey data sources than the data source used in this study. Methods such as CART, CAT, and PCA could, for example, be applied on federal and other publicly available datasets to further improve the validity and generalizability of the findings of this case study and conduct more efficient and cost-effective surveys in the future.

Strengths and Limitations

A strength of this study pertains to its large sample size; 2837 patients from various CR clinics completed the various NAP questionnaires, confirming the external validity of the findings. The number of fully filled out questionnaires per type of questionnaire varied from 716 to 2633, with MPSSS being used in one CR clinic and having the lowest response rate. The original version of the MPSSS, furthermore, has a low number of questionnaire items in comparison with QLMI, for example. This might have impacted the CAT and CART analysis. It, hence, remains uncertain whether the results, particularly those of MPSSS, can be generalized to other CR clinics. An even larger sample size could have led to a more precise estimation of NAP questionnaire lengths needed to determine patients' specific CR needs.

Another strength of this study is the application, exploration, and comparison of the two techniques for shortening questionnaires. The use of CART analysis as one of these techniques has been suboptimal at the least; we did not find any study using CART aiming at reducing questionnaire lengths. Finally, the Dutch guideline for CR prescribes that CR clinics can choose between a combination of HADS-A/HADS-D or GAD-7/PHQ-9 to assess anxiety and depression levels in patients. We do not know if these combinations of questionnaires are equally valid in measuring the constructs of anxiety and depression. With CART, a combination of HADS-A/HADS-D would result in a reduction in questionnaire lengths of six and five items, respectively, whereas the combined GAD-7/PHQ-9 would lead to a reduction of seven and five items, respectively.

Finally, we did not examine if the questionnaires of the NAP produced similar results when patients filled them in on the Web through MyCARDSS at their home or in the CR clinic.

Conclusions

CART and CAT both have shown to be accurate methods for reducing the length of the NAP used in the field of CR. Of both methods, CART overall showed the largest decrease in the number of questionnaire items and the best performance in terms of sensitivity and specificity. This study is the first to apply and compare the performances of CAT and CART for shortening questionnaires, and it demonstrated that the use of CART analysis would be a step forward in the development of a shorter NAP questionnaire for CR patients and possibly other questionnaires.

Instead of using long, fixed-length questionnaires on paper or on the Web for the NAP, a much smaller set of questionnaire items will suffice to identify patients' varying needs for CR,

without the loss of information and data quality and an excessive   burden on the patient.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Chien T, Wu H, Wang W, Castillo R, Chou W. Reduction in patient burdens with graphical computerized adaptive testing on the ADL scale: tool development and simulation. Health Qual Life Outcomes 2009 May 5;7:39 [FREE Full text] [doi: 10.1186/1477-7525-7-39] [Medline: 19416521]

2. Chien T, Wang W, Huang S, Lai W, Chow JC. A web-based computerized adaptive testing (CAT) to assess patient perception in hospitalization. J Med Internet Res 2011 Aug 15;13(3):e61 [FREE Full text] [doi: 10.2196/jmir.1785] [Medline: 21844001]

3. Ma S, Chien T, Wang H, Li Y, Yui M. Applying computerized adaptive testing to the Negative Acts Questionnaire-Revised: Rasch analysis of workplace bullying. J Med Internet Res 2014 Feb 17;16(2):e50 [FREE Full text] [doi: 10.2196/jmir.2819] [Medline: 24534113]

4. de Beurs DP, de Vries AL, de Groot MH, de Keijser J, Kerkhof A. Applying computer adaptive testing to optimize online assessment of suicidal behavior: a simulation study. J Med Internet Res 2014 Sep 11;16(9):e207 [FREE Full text] [doi: 10.2196/jmir.3511] [Medline: 25213259]

5. Zimmerman R, Balasubramani G, Nowalk M, Eng H, Urbanski L, Jackson M, et al. Classification and Regression Tree (CART) analysis to predict influenza in primary care patients. BMC Infect Dis 2016 Sep 22;16(1):503 [FREE Full text] [doi: 10.1186/s12879-016-1839-x] [Medline: 27659721]

6. Valera V, Walter B, Yokoyama N, Koyama Y, Iiai T, Okamoto H, et al. Prognostic groups in colorectal carcinoma patients based on tumor cell proliferation and classification and regression tree (CART) survival analysis. Ann Surg Oncol 2007 Jan;14(1):34-40. [doi: 10.1245/s10434-006-9145-2] [Medline: 17024555]

7. Lu F, Petkova E. A comparative study of variable selection methods in the context of developing psychiatric screening instruments. Stat Med 2014 Feb 10;33(3):401-421 [FREE Full text] [doi: 10.1002/sim.5937] [Medline: 23934941]

8. van Engen-Verheul MM, de Keizer NF, van der Veer SN, Kemps HM, Reimer WJ, Jaspers MW, et al. Evaluating the effect of a web-based quality improvement system with feedback and outreach visits on guideline concordance in the field of cardiac rehabilitation: rationale and study protocol. Implement Sci 2014 Dec 31;9:780 [FREE Full text] [doi: 10.1186/s13012-014-0131-y] [Medline: 25551791]

9. Rehabilitation CNSFC(HF(GPAAHR. Multidisciplinary Guidelines for Cardiac Rehabilitation (in Dutch). Utrecht: Netherlands Society for Cardiology (NVVC) 2012:1-16.

10. Wiggers A, Vosbergen S, Kraaijenhagen R, Jaspers M, Peek N. Changes in the cardiac rehabilitation workflow process needed for the implementation of a self-management system. Stud Health Technol Inform 2013;192:1140. [doi: 10.3233/978-1-61499-289-9-1140] [Medline: 23920914]

11. Dempster M, Donnelly M, O'Loughlin C. The validity of the MacNew Quality of Life in heart disease questionnaire. Health Qual Life Outcomes 2004 Jan 22;2:6 [FREE Full text] [doi: 10.1186/1477-7525-2-6] [Medline: 14738566]

12. Herrmann C. International experiences with the Hospital Anxiety and Depression Scale--a review of validation data and clinical results. J Psychosom Res 1997 Jan;42(1):17-41. [doi: 10.1016/s0022-3999(96)00216-4] [Medline: 9055211]

13. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

14. Spitzer R, Kroenke K, Williams J, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

15. Dahlem N, Zimet G, Walker R. The Multidimensional Scale of Perceived Social Support: a confirmation study. J Clin Psychol 1991 Nov;47(6):756-761. [doi: 10.1002/1097-4679(199111)47:6<756::aid-jclp2270470605>3.0.co;2-l] [Medline: 1757578]

16. Alpaydin E. Introduction to Machine Learning. Second Edition. London: MIT Press; 2010.

17. Song Y, Lu Y. Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry 2015 Apr 25;27(2):130-135 [FREE Full text] [doi: 10.11919/j.issn.1002-0829.215044] [Medline: 26120265]

18. Breiman L, Friedman J, Stone C, Olshen R. Classification And Regression Trees. London: Chapman And Hall; 2017.

19. van der Linden WJ, Glas CA. Elements of Adaptive Testing. An Advanced and Comprehensive Introduction to the Foundation of CAT. Amsterdam: Springer; 2010.

20. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Statist Surv 2010;4:40-79. [doi: 10.1214/09-SS054]

21. Kuhn M. Building Predictive Models in R Using the caret Package. J Stat Soft 2008;28(5):1-26. [doi: 10.18637/jss.v028.i05]

22.   Therneau TM, Atkinson EJ. CRAN - R Project. 2018. An Introduction to Recursive Partitioning Using the RPART Routines URL: https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf [accessed 2019-09-15]

23.   Nikolaus S, Bode C, Taal E, Vonkeman HE, Glas CA, van de Laar MA. Construct validation of a multidimensional computerized adaptive test for fatigue in rheumatoid arthritis. PLoS One 2015;10(12):e0145008 [FREE Full text] [doi: 10.1371/journal.pone.0145008] [Medline: 26710104]

24.   Flens G, Smits N, Terwee C, Dekker J, Huijbrechts I, de Beurs E. Development of a computer adaptive test for depression based on the Dutch-Flemish version of the PROMIS item bank. Eval Health Prof 2017 Mar;40(1):79-105. [doi: 10.1177/0163278716684168] [Medline: 28705028]

25.   Gibbons R, Weiss D, Pilkonis P, Frank E, Moore T, Kim J, et al. Development of a computerized adaptive test for depression. Arch Gen Psychiatry 2012 Nov;69(11):1104-1112 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2012.14] [Medline: 23117634]

26.   Gibbons R, Hooker G, Finkelman M, Weiss D, Pilkonis P, Frank E, et al. The computerized adaptive diagnostic test for major depressive disorder (CAD-MDD): a screening tool for depression. J Clin Psychiatry 2013 Jul;74(7):669-674 [FREE Full text] [doi: 10.4088/JCP.12m08338] [Medline: 23945443]

27.   Choi SW, van der Linden WJ. Ensuring content validity of patient-reported outcomes: a shadow-test approach to their adaptive measurement. Qual Life Res 2018 Jul;27(7):1683-1693. [doi: 10.1007/s11136-017-1650-1] [Medline: 28710673]

28.   Gibbons R, Weiss D, Frank E, Kupfer D. Computerized adaptive diagnosis and testing of mental health disorders. Annu Rev Clin Psychol 2016;12:83-104. [doi: 10.1146/annurev-clinpsy-021815-093634] [Medline: 26651865]

29.   Fries J, Witter J, Rose M, Cella D, Khanna D, Morgan-DeWitt E. Item response theory, computerized adaptive testing, and PROMIS: assessment of physical function. J Rheumatol 2014 Jan;41(1):153-158. [doi: 10.3899/jrheum.130813] [Medline: 24241485]

30.   Lehman L, Woodbury M, Shechtman O, Wang Y, Pomeranz J, Gray D, et al. Development of an item bank for a computerised adaptive test of upper-extremity function. Disabil Rehabil 2011;33(21-22):2092-2104. [doi: 10.3109/09638288.2011.560336] [Medline: 21401332]

31.   Öztuna D, Elhan A, Küçükdeveci AA, Kutlay S, Tennant A. An application of computerised adaptive testing for measuring health status in patients with knee osteoarthritis. Disabil Rehabil 2010;32(23):1928-1938. [doi: 10.3109/09638281003777572] [Medline: 20384449]

32.   Hsueh I, Chen J, Wang C, Hou WH, Hsieh CL. Development of a computerized adaptive test for assessing activities of daily living in outpatients with stroke. Phys Ther 2013 May;93(5):681-693. [doi: 10.2522/ptj.20120173] [Medline: 23329557]

33.   Ma S, Chien T, Wang H, Li Y, Yui M. Applying computerized adaptive testing to the Negative Acts Questionnaire-Revised: Rasch analysis of workplace bullying. J Med Internet Res 2014 Feb 17;16(2):e50 [FREE Full text] [doi: 10.2196/jmir.2819] [Medline: 24534113]

34.   Ma S, Wang H, Chien T. A new technique to measure online bullying: online computerized adaptive testing. Ann Gen Psychiatry 2017;16:26 [FREE Full text] [doi: 10.1186/s12991-017-0149-z] [Medline: 28680455]

35.   Sayyad M, Gopal G, Shahani A. Classification and regression trees: a possible method for creating risk groups for progression to diabetic nephropathy. J Appl Sci 2011;11(12):2076-2083. [doi: 10.3923/jas.2011.2076.2083]

36.   Chang YJ, Chen LJ, Chang YJ, Chung KP, Lai MS. Risk groups defined by Recursive Partitioning Analysis of patients with colorectal adenocarcinoma treated with colorectal resection. BMC Med Res Methodol 2012 Jan 3;12:2 [FREE Full text] [doi: 10.1186/1471-2288-12-2] [Medline: 22214198]

37.   Kuhn L, Worrall-Carter L, Ward J, Page K. Factors associated with delayed treatment onset for acute myocardial infarction in Victorian emergency departments: a regression tree analysis. Australas Emerg Nurs J 2013 Nov;16(4):160-169. [doi: 10.1016/j.aenj.2013.08.002] [Medline: 24199901]

38.   Su X, Azuero A, Cho J, Kvale E, Meneses K, McNees M. An introduction to tree-structured modeling with application to quality of life data. Nurs Res 2011;60(4):247-255 [FREE Full text] [doi: 10.1097/NNR.0b013e318221f9bc] [Medline: 21720217]

39.   D'Alisa MS, Miscio G, Baudo S, Simone A, Tesio L, Mauro A. Depression is the main determinant of quality of life in multiple sclerosis: a classification-regression (CART) study. Disabil Rehabil 2006 Mar 15;28(5):307-314. [doi: 10.1080/09638280500191753] [Medline: 16492625]

Abbreviations

CART: classification and regression tree
CAT: computerized adaptive testing
CDS: computerized decision support
CDSS: computerized decision support system
CR: cardiac rehabilitation
EPR: electronic patient record
GAD-7: Generalized Anxiety Disorder - 7
GRM: Graded Response Model

HADS-A: Hospital Anxiety and Depression Scale - Anxiety
HADS-D: Hospital Anxiety and Depression Scale - Depression
IRT: Item Response Theory
MPSSS: Multidimensional Perceived Social Support
NAP: needs assessment procedure
NRMSE: normalized root mean squared error
PCA: principal component analyses
PHQ-9: Patient Health Questionnaire - 9
QLMI-E: Quality of Life after Myocardial Infarction - Emotional dimension
QLMI-P: Quality of Life after Myocardial Infarction - Physical dimension
QLMI-S: Quality of Life after Myocardial Infarction - Social dimension
RMSE: root mean squared error

XSL•FO
RenderX