Published paper
Biddiss , E., Brownsell, S. and Hawley, M. (2009) *Prediction of health status in individuals with congestive heart failure using a home-based telecare system.* Journal of Telemedicine and Telecare, 15 . pp. 226-231.
http://dx.doi.org/doi:10.1258/jtt.2009.081203

Predicting need for intervention in individuals with congestive heart failure using a home-based telecare system

Elaine Biddiss[1,2,3*], Simon Brownsell[1,4], and Mark S. Hawley[1,4]

[1]School of Health and Related Research, University of Sheffield, Regent Street, Sheffield, S1 4DA, UK

[2]Bloorview Research Institute, Bloorview Kids Rehab, Toronto, M4G 1R8, Canada

[3]Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, M5S 3G9, Canada

[4]Dept of Medical Physics and Clinical Engineering, Barnsley Hospital NHS Foundation Trust, Barnsley, S75 2EP, UK

*Corresponding author information:

Address: Bloorview Research Institute, 150 Kilgour Rd., Toronto Ontario, Canada, M4G 1R8

Phone number: +1 416-425-6220 x3505

Fax: +1 416-425-1634

E-mail: elaine.biddiss@utoronto.ca

Short Title: Predicting health needs using telecare

ABSTRACT

Telecare is increasingly used to remotely monitor long-term conditions such as congestive heart failure (CHF) and provide interventions based upon the data collected. In order to improve health care efficiency, there remains a need for decision support tools to automate this monitoring function and help guide interventions; this study sought to develop such a tool. Data was obtained from 45 elderly individuals with CHF who participated in a telecare trial for an average duration of 18 months. Physiological data along with subjective health perspectives and symptoms were reported. Clinicians responded to abnormalities in the data resulting in 154 key medical interventions/events. A multivariate logistic regression model was developed to predict these medical interventions/events. The developed model correctly predicted key medical events in 75% of cases with a specificity of 74% and an overall cross-validated accuracy of 74% [68-80%, 95% confidence interval]. Key predictors included: number of system alerts, self-rated mobility, self-rated health, and self-rated anxiety, strongly suggesting the utility of subjective measures in addition to physiological ones for prediction of health status. Overall this study demonstrates the potential of a multivariate decision-support model to enhance predictions of medical need in CHF patients using home-based telecare systems.

INTRODUCTION

In the United States alone, 4.8 million lives are affected by chronic heart failure (CHF), which accrues health care costs of $38.1 billion annually [1]. Regular monitoring is especially pertinent to the management of CHF wherein signs of diminishing health may be subtle and difficult to recognize by patients and their caregivers alone [1,2]. Nevertheless, it has been suggested that many of these symptoms of worsening health (i.e. edema, dyspnoea, weight gain) present 8 to 12 days before hospitalization [3]. Telecare systems present the opportunity to address this issue in a cost-effective and patient-acceptable way [3,4].

Despite growing interest and investment in this area, there remain numerous questions as to how to achieve the greatest increase in clinical and cost effectiveness [5]. For instance, questions remain as to the predictors that are most indicative for a particular cardiovascular population [6]. In this paper, we ask: how well can need for medical intervention be predicted by a telecare monitoring system based on self-rated health-related quality of life (HQOL), physical symptoms, lifestyle, and physiological measurements in individuals with CHF?

METHODOLOGY

Participants

A review of Barnsley Hospital records identified potential participants for this study. All participants had echocardiographic evidence of heart failure and conventional symptoms. Exclusion criteria included: (1) Ejection fraction >40%; (2) Unstable angina; (3) <60 years; (4) Debilitating dementia or psychiatric disorder; (5) Inability to comprehend words presented on an electronic screen; (6) Planned coronary revascularization procedures; (7) On a waiting list for heart transplantation; (8) Participation in another, conflicting heart failure research study within the past 6 months; (9) Lack of an operational home telephone line and electrical socket within close proximity; (10) Not living in the mainstream housing sector (e.g. residential or nursing care).

Data collection

Participants were provided with a Doc@Home (Docobo Inc, United Kingdom) health monitor through which they entered daily information pertaining to their symptoms and health status through a set of questions developed by the research and clinical team. Daily physiological measurements of blood pressure, pulse rate, and weight were also entered using A&D UA-767 Plus BT and Hanson HCV800 scales. Twice weekly, patients also completed a health-related quality of life measure (EQ-5D [7]) directly on the Docobo unit, giving data on self-rated health (visual analogue scale), mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Data were transmitted nightly through telephone lines and screened for abnormalities. If data fell outside user-specific parameters, clinicians were notified. This therefore provided two additional sets of data: (a) daily record of system alerts generated by the Doc@Home system in accordance with the parameters listed in Table 1, and (b) qualitative log of clinical interventions and medical events recorded by the monitoring healthcare practitioners.

Data analysis

Data analysis was conducted using SPSS 14.0 statistical software with a standard significance level of $p=0.05$. Logistic Regression (LR) [8-12] was used to predict the occurrence of *key medical events/interventions* as extracted from the healthcare practitioner logs. For each week of the study duration, the average/median values of predictor variables were calculated, the number of system alerts was enumerated, and the presence of a key medical event or intervention was noted from the logs of monitoring healthcare practitioners. The rationale motivating this approach was two-fold:

(1) Data was collected with varying frequency (either weekly, twice a week, or daily).

(2) A time lag was evident between user-inputted data, generated system alerts, and nurse response.

Stepwise, forward selection based on the log likelihood ratio was used to avoid over-fitting the model [8]. The importance of all potential interactions was evaluated via the likelihood ratio test. Lastly, the model's "goodness of fit" was assessed based on changes in deviance (i.e. > 4 indicates poor fit [12]).

Models were evaluated using K-fold cross validation with K=10. Due to the large number of non-events (i.e. a meaningful healthcare intervention was not required) in comparison to key events (i.e. a meaningful healthcare intervention was required), over-sampling [13,14] was used to obtain a balanced data set. The approximate proportion of key events to non-events was maintained in each test set. Model sensitivity (i.e. the proportion of key events that were correctly classified), specificity (i.e. the proportion of non-events that were correctly classified), and overall prediction accuracy were calculated and used to evaluate the performance of the model.

Ethics

This study received NHS research ethics approval and research governance approval from Barnsley Hospital.

RESULTS

Summary of participants

Of the 45 participants, six individuals passed away during the course of the study and eight returned their equipment. The average duration of data collection was 18 ± 5 months. Detailed description of patient characteristics is presented in Table 2.

Predicting need for medical intervention

8576 alerts were generated by the telecare monitoring system based on self-reported symptoms, lifestyle, and physiological measurements. In the majority of cases, response to system alerts did not require patient and service provider interaction. When system alerts were considered of greater severity, or if symptoms persisted, the patient was contacted. 171 key medical events (6 deaths; 28 hospital admissions; 59 changes in medication; 54 advice given; 24 instances where immediate medical attention was recommended) were recorded in the monitoring logs. As such, there were 154 weeks during which one or more key medical events occurred and 2779 weeks during which no key medical events were observed for the participants. Generation of a system alert and subsequent response by a healthcare practitioner was not considered a key medical event unless a specific action was actually required and taken (i.e. false alarms were not counted as key medical events). In order to obtain an approximately balanced data set, key medical events were over-sampled by a rate of 18 times. The average number of medical events experienced per patient per year was 3.5 ± 4 (with a median of 2 and an interquartile range of 1-4). The average number of non-key alerts generated per patient per year was 49 ± 4 (with a median of 49 and an interquartile range of 47-51). The average percentage of total alerts that were identified as key medical events was 6.4 ± 4% (with a median of 4% and an interquartile range of 1.4%-8%).

Table 3 summarizes the univariate significance of predictor variables. All variables with a p-value < 0.1 were examined for inclusion in the logistic regression model as per guidelines prescribed in [12]. Some of the variables listed in Table 3 were highly correlated. For example, self-rated mobility and self-rated pain were also shown to be highly correlated ($r = 0.742$, $p<0.01$). In these cases, the strongest predictor in the group of correlated variables was selected for inclusion in the model.

Table 4 presents the optimal multivariate logistic regression model for prediction of key medical interventions/events. The number of alerts generated by the system emerged as the primary predictor. As evident by the odds ratio ($e^{\beta} = 1.196$) listed in Table 4, for every additional system alert generated, the probability of a key medical event increased by 19.6%. Alone, this variable predicts 82% of non-events and 61% of key events with an unadjusted coefficient of $\beta = 0.183$. The addition of subjective factors (i.e. self-rated mobility, health, and anxiety) improved prediction significantly (log likelihood ratio, $p<0.001$). Figure 1(a) presents the ROC (receiver-operator curve) for this revised model. With a classification cut-off probability of 0.5, the overall cross-validated prediction accuracy was 74% [68-80%, 95% confidence interval]. Most importantly, the sensitivity (i.e. prediction of key events) was increased from 61% to 75% with a specificity (prediction of non-events) of 74%. With a maximum sensitivity of approximately 80%, the specificity drops to 67% (cut-off = 0.62). Figure 1(b) presents the sensitivity and specificity of the model for a range of classification cut-off values. Of the data, 100% of key medical events and 97% of non-events were well-fitted (i.e. change in deviance <4). Approximately 72% of poorly fitted data points were associated with four particular participants. A significantly higher number of daily system alerts (1.5 ± 0.5) were associated with these patients as compared to the average (0.5 ± 0.5), $p<0.001$.

Lastly, a breakdown of all the daily system alerts (N=8576) generated by the telecare monitoring system indicates that a large proportion of the total number of alerts was attributable to the physiological measurements, namely pulse rate, blood pressure and weight gain. In response to these alerts, monitoring practitioners typically flagged the patient for elevated observation. If the symptom persisted, a phone call was made to ascertain the health of the patient and possible reasons for the physiological change (e.g. over-eating, an unrelated cold, etc.). In approximately 86% of cases, alerts generated by these physiological measures were not accompanied by a key medical event/intervention. Alerts pertaining to physical and psychological symptoms, such as anxiety, swollen ankles, and need for extra pillows at night, were most often correct in predicting a key event (Figure 2).

DISCUSSION

Key Findings

The model predicted key medical events/interventions with an overall cross-validated accuracy of 74% [68-80%, 95% confidence interval]. With a classification cut-off probability of 0.5, the sensitivity of the model was 75% and the specificity was 74%. To minimize the risk of not identifying need for medical intervention, a higher classification cut-off probability could be used; this increase in sensitivity is of course accompanied by a decrease in specificity (i.e. more false alarms). Of note, the strongest predictor in this model was the cumulative number of system alerts generated in a given week. When considering Figure 2, it is the system alerts stemming from patients' subjective descriptions of their symptoms as opposed to the physiological metrics that are most indicative of need for medical intervention. This implies that patients are giving medically meaningful reports of their symptoms. With additional predictors based on subjective health perceptions (i.e. self-rated mobility, health, and anxiety), correct predictions of key medical events are improved from 61% to 75%. Consequently, in order to increase the effectiveness of telecare systems, it may be important to record more than just physiological parameters. Further research to objectively explore an even wider range of potential predictors is needed to identify those most effective for identifying need for health intervention.

The performance of the model's predictions was compared to the clinicians' responses. To estimate the positive predictivity (i.e. the ratio of true key events to the total number of key events predicted) of the clinicians, the assumption was made that contact with a patient that ensued from a medical concern, but did not result in a key intervention, was a "false alarm" or a "false positive". Within the context of this definition, the positive predictivity was estimated at 39% for clinicians as compared to 75% for the decision-support model. Of note, two thirds of all "false alarms" generated by the model (i.e. incorrectly predicted key events) were instances where the clinician involved also demonstrated a heightened concern for the patient and decided to increase monitoring of and/or contact the patient based on the information collected. This suggests that, although incorrect in its prediction of a key medical event in these cases, the decision-support model did identify instances of elevated risk in line with clinicians' assessments.

Clinical Significance

It is important to emphasize that predictive models should be regarded as a useful tool to *assist,* not supercede, clinical decision making and prioritizing. It is further noted that "non-key interventions" (i.e. contact with patients that did not result in a tangible medical intervention), may still have rendered a meaningful healthcare service by promoting patient satisfaction, increasing confidence in the quality of care provided, alleviating feelings of social isolation, increasing perceived social support, encouraging adherence to treatment recommendations, addressing a co-morbidity or other issue, and improving clinician-patient relationships. All of these social and healthcare perceptions have been implicated as factors in hospital re-admission rates, mortality, and/or quality of life in CHF patients [1,2,15-18]. Increasingly, emphasis is being placed on patients with CHF to self-care through initiatives such as the Expert Patient Programme [19]. It could be that telecare systems with predictive modeling could further complement such initiatives to ensure the best possible outcomes for patients and the funders of health care systems. After all, early identification of high-risk patients, improved home care, and education on heart failure and self-management, are fundamental strategies to decrease morbidity and mortality among patients, and to alleviate the economic burden accrued by frequent hospital re-admissions [20-22].

Study limitations and areas for future work

This study explored the development of a practical decision-support tool that incorporates a mixture of physiological measurements, physical symptoms and subjective perspectives on health and well-being. As such, we have identified a few important predictors of health status and have explored the factors framing self-rated HQOL. Larger data sets will enable the development of more accurate, robust, and generalized models that can predict not only the occurrence of a key medical event and/or need for intervention, but the level of severity of the event.

Logistic regression does not account for the longitudinal nature of the data. It is a “safe” estimate in that it is more likely to include potentially unimportant variables, as opposed to exclude important predictors [23]. Although the odds ratios are likely to be comparable, standard errors may be underestimated in comparison to methods that account for repeated measures such as generalized estimating equations (GEE) [23]. The latter approach, however, requires the assumption that missing data occur completely at random and independent of the outcome variable. In our study, missing data were commonly due to hospitalization and the patient’s inability to access their Doc@Home system. For comparison purposes, a GEE model was constructed. Systems alerts and self-rated mobility emerged as the primary predictors with odds ratios comparable to those of the simple logistic regression model. As expected, the standard errors (SE) calculated through GEE were significantly larger for both system alerts (SE= 0.017) and self-rated mobility (SE=.225). The goodness-of-fit measure, the Corrected Quasi Likelihood under Independence Model Criterion (QICC), for the GEE model incorporating system alerts and self-rated mobility was slightly higher (QICC=934) than the model which also included self-rated health and self-rated anxiety (QICC=889). This implies that the latter two variables, although not statistically significant predictors, may contribute to the goodness-of-fit of the GEE model.

The necessity of carrying out analyses on a weekly basis to account for time delays between user inputs, system response, and clinical action, should also be noted. Inconsistent adjustments of system parameters by monitoring practitioners may also have been an issue. For example, for some individuals whose physiological measurements had greater acceptable fluctuations than others, system parameters were changed to eliminate superfluous alerts, while for others, these alerts were simply ignored. This may have affected the model fit and issued a higher number of false alarms. Differences in each individual's ability to self-manage (e.g. medication, diet) are also not captured in this analysis. It is also possible that some individuals may be more in tune with their health needs than others and that the model could be refined on an individual-by-individual basis to reflect patient variations in sensitivity or anxiety regarding perceived symptoms. It remains to be seen how and if this decision-support model applies on a daily or continuous basis, as is ideal for rapid identification of high-risk individuals and prompt provision of medical interventions.

CONCLUSIONS

From this study, four important conclusions emerged with respect to the performance and development of telecare systems. Firstly, current systems for health monitoring are useful in indicating when medical interventions are needed. Secondly, the performance of these systems can be improved by including targeted questions relating to health outcomes. Thirdly, self-perceived symptoms and health status surfaced as valuable indicators contrary to current trends which focus on physiological measurements. Lastly, the potential of a multivariate decision support model to supplement practitioners and current telecare systems in identifying CHF patients in need of medical intervention is demonstrated. Inclusion of such systems in real time could enhance system effectiveness, enable preventative healthcare, and increase practitioner efficiency. These developments could result in a step change in performance of telecare systems and therefore provides an important insight into present and future developments in the delivery of community based services to people with CHF and possibly other long term conditions.

ACKNOWLEDGEMENTS

This research builds upon work carried out under EPSRC's EQUAL Programme (GR/S29058/01, EP/F001835/1). Thanks to Bloorview Kids Rehab for financial support of E. Biddiss' fellowship at the University of Sheffield. Dr. Michael Campbell (ScHARR, University of Sheffield) is gratefully acknowledged for his statistics guidance.

REFERENCES

[1] Zickmund S, Blasiole J, Brase V, Arnold R. Congestive Heart Failure Patients Report Conflict With Their Physicians. *J Cardiac Failure* 2006;12:546-53.

[2] Vinson JM, Rich MW, Sperry JC, Shah AS, McNamara T. Early readmission of elderly patients with congestive heart failure. *J Am Geriatr Soc* 1990;38:1290-5.

[3] Clark R, Inglis S, McAlister F, Cleland J, Stewart S. Telemonitoring or structured telephone support programmes for patients with chronic heart failure: systematic review and meta-analysis. *Brit Med J* 2007;334:942-51.

[4] Pare G, Jaana M, Sicotte C. Systematic Review of Home Telemonitoring for Chronic Diseases: The Evidence Base. *J Am Med Inform Assn* 2007;14:269-77.

[5] Brownsell S. Supporting Long Term Conditions and disease management through Telecare and Telehealth: evidence and challenges. Department of Health, 2008. [Available online: http://www.icn.csip.org.uk/nl/?l=202_1_1_3]. England. Accessed: August 27, 2008.

[6] Artinian N. Telehealth as a Tool for Enhancing Care for Patients With Cardiovascular Disease. *J Cardio Nursing* 2007;22:25-31.

[7] Lenzen M, Scholte op Reimer, W., Pedersen S, Boersma E, Maier W, Widimsky P, et al. The additional value of patient-reported health status in predicting 1-year mortality after invasive coronary procedures: a report from the Euro Heart Survey on Coronary Revascularisation. *Heart* 2007;93:339-44.

[8] Bagley S, White H, Golomb B. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *J Clin Epidemiol* 2001;54:979-85.

[9] Wielandt T, Mckenna K, Tooth L, Strong J. Factors that predict the post-discharge use of recommended assistive technology (AT). *Disabil Rehab: Assitive Tech* 2006;1:29-40.

[10] Johansson U,Bernspång B. Predicting return to work after brain injury using occupational therapy assessments. *Disabil Rehab* 2001;23:474-80.

[11] Biddiss E,Chau T. Multivariate Prediction of upper limb prosthesis acceptance or rejection. Disability and Rehabilitation: *Assistive Technology* [E-publication ahead of print available online: Feb, 2008].

[12] Hosmer D,Lemeshow S. Applied Logistic Regression. 1989;307.

[13] Eitrich T, Kless A, Druska C, Meyer W, Grotendorst J. Classification of Highly Unbalanced CYP450 Data of Drugs Using Cost Sensitive Machine Learning Techniques. *J Chem Inf Model* 2007;47:92-103.

[14] Kotsiantis S, Kanellopoulos D, Pintelas P. Handling imbalanced datasets: A review. *GESTS Int Trans Comp Sci & Eng* 2006;30:12.

[15] Yua D, Leea D, Woob J, Thompson D. Correlates of psychological distress in elderly patients with congestive heart failure. *J Psychosomatic Res* 2004;57:573-81.

[16] Gorkin L, Schron EB, Brooks MM, Wiklund I, Kellen J, Verter J, et al. Psychosocial predictors of mortality in the cardiac arrhythmia suppression Trial-1 (CAST-1). *Am J Cardiol* 1993;71:263-7.

[17] Williams R, Barefoot JC, Califf RM, Haney TL, Saunders WB, Pryor DB, et al. Prognostic importance of social and economic resources among medically treated patients with angiographically documented coronary artery disease. *Journal of the American Medical Association* 1992;267:520-4.

[18] Murberg T,Bru E. Social relationships and mortality in patients with congestive heart failure. *J Psychosomatic Res* 2001;51:521-7.

[19] Griffiths C, Foster G, Ramsay J, Eldridge S & Taylor S. How effective are expert patient (lay led) education programmes for chronic disease? *BMJ Journal* 2007;334:1254-1256.

[20] Lennie T, Moser D, Heo S, Chung M, Zambroski C. Factors Influencing Food Intake in Patients With Heart Failure A Comparison With Healthy Elders. *J Cardio Nursing* 2006;21:123-9.

[21] Proctor E, Morrow-Howell N, Li H, Dore P. Adequacy of home care and hospital readmission for elderly congestive heart failure patients. *Health & Social Work* 2000;25:87-96.

[22] Cline C, Israelsson B, Willenheimer R, Broms K, Erhardt L. Cost effective management programme for heart failure reduces hospitalisation. *Heart* 1998;80:442-6.

[23] Williamson D, Bangdiwala S, Marshall S, Waller A. Repeated measures analysis of binary outcomes: applications to injury research. *Accid Anal and Prev* 1996:571-9.

List of Tables

List of Figures

Table 1. Default criteria used by the monitoring system to generate system alerts. Specific parameters could be tailored to each individual as appropriate.

Health Factor Query	**Alert generated if…**
Daytime shortness of breath	More than usual

Night time shortness of breath	More than usual
Need for extra pillows	Yes
Swollen ankles	More than usual
Bloated stomach	More than usual
Dizziness	More than usual
Urine excretion	Less than usual + increased dizziness + increased bloating
Cough	New or worse than usual
Weight gain	If weight increases 2lbs since previous day or if weight increases 3lbs in a rolling 7 days
Systolic blood pressure	If below 100mmHg or drops by 20mmHg from previous reading
Pulse rate	If below 55 or above 120
Diet (i.e. eating well)	No or “less than usual” for a consecutive 7 days
Medication taken	No
Angina	More than usual

Table 2: Summary of patient characteristics

Population Characteristic	Participants (N=45)
Gender	
Male	83%
Female	17%
NYHA classification	
2	43%
2-3	17%
3	40%
Age (years)	
60-64	9.5%
65-69	26%
70-74	26%
75-79	24%
80-84	9.5%
85-89	5%
Living Arrangements	
Alone	26%
With partner/spouse	62%
More than 2 people in household	12%
Smoking	
Non-smoker	82%
Smoker	18%
Exercise	
None	67%
Light	28%
Moderate	5%
Number of prescribed medications	
<5	13%
5-10	59%
11-15	23%
>15	5%

Table 3. Variables considered as predictors of key medical events/interventions based on univariate statistical significance (Chi-squared test for binary predictors; independent t-test for continuous)

Variable	**Scale**	**p**
Number of system alerts	Continuous	<0.001
Quality of sleep	Binary (0-as usual or more; 1-less than usual or none)	<0.001
Extra pillows needed	Binary (0-no; 1-yes)	<0.001
Short of breath (day)	Binary (0-as usual or less; 1-more than usual)	<0.001
Diet	Binary (0-as usual or more; 1-less than usual)	0.001
Cough	Binary (0-as usual or none; 1-new or worse cough)	<0.001
Weight	Binary (0-within parameters; 1-outside of parameters)	<0.001
Fatigue	Binary (0-as usual or less; 1-more than usual)	<0.001
Self-rated health	Continuous	<0.001
Self-rated mobility	Binary (0-no problems; 1-some problems or unable)	0.001
Self-rated anxiety	Binary (0-none; 1-moderate or extreme)	0.002
Exercise	Binary (0-some exercise; 1-No exercise)	0.002
Self-rated pain	Binary (0-none; 1-moderate or extreme)	0.012

Table 4. Logistic regression model for prediction of key medical interventions/events.

Variables	Coefficient (β)	Standard Error	Wald	p	Odds Ratio (e^{β}) [95% CI]
System Alerts	0.179	0.008	545.1	<0.001	1.196 [1.178-1.214]
Self-rated Mobility	0.441	0.093	23.3	<0.001	1.559 [1.299-1.871]
Self-rated Health	-0.009	0.002	16.8	0.012	0.991 [0.986-0.995]
Self-rated Anxiety	0.144	0.084	3.6	0.192	1.157 [0.982-1.364]
Constant	-0.863	0.216	17.1		0.436

Figure 1. Performance of the logistic regression model predicting key medical interventions/events. (a) Receiver-operator curve (ROC) (b) Model sensitivity and specificity for a range of classification cut-off values.

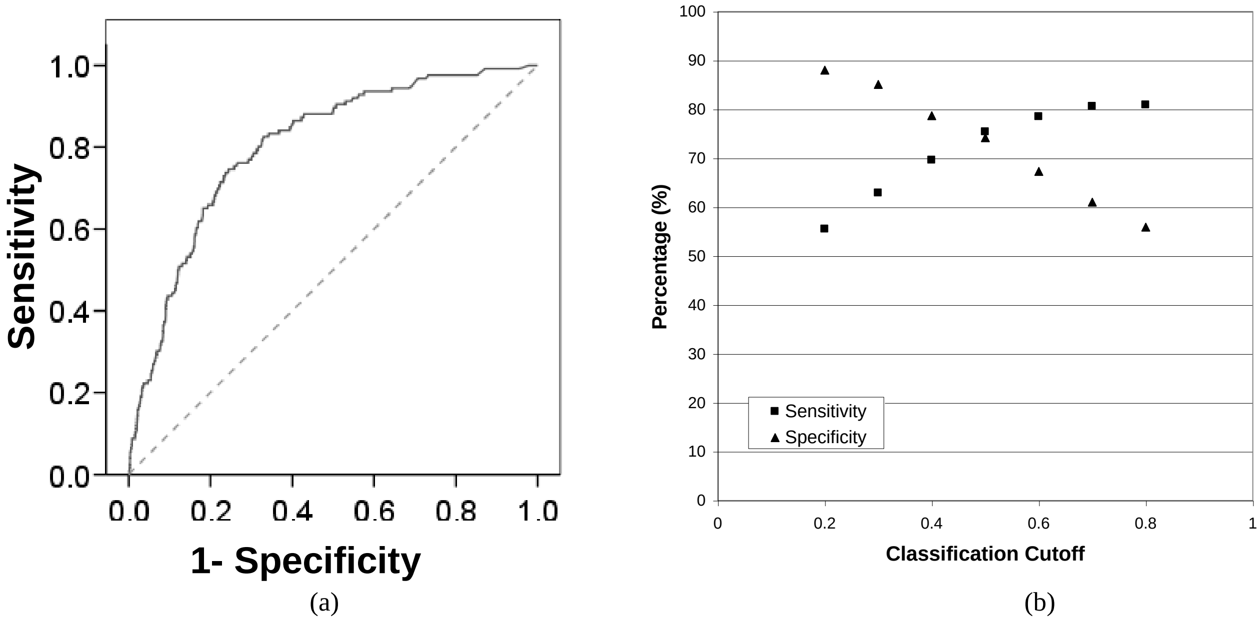

Figure 2. Breakdown of the total number of system alerts generated by class (e.g. weight gain, increased anxiety, worsened cough etc.). For each class of alert, the proportion with which it was associated with a key medical intervention/event is indicated (in black) as compared to a non-key alert (in grey). For example, reports of increased dizziness generated 339 alerts during the course of this study. In approximately 15% of cases, this symptom was associated with a key medical intervention/event, whereas in 85% of cases, it was not associated with a key medical intervention/event.