Application of Receiver Operating Characteristic Analysis to a Remote Monitoring Model for Chronic Obstructive Pulmonary Disease to Determine Utility and Predictive Value

Brunel University

Application of Receiver Operating Characteristic Analysis to a Remote Monitoring Model for Chronic Obstructive Pulmonary Disease to Determine Utility and Predictive Value

Nancy Elizabeth Brown Connolly R.N., M.S.N.

Department of Information Systems, Computing and Mathematics

A thesis submitted for the degree of Doctor of Philosophy

Awarded by Brunel University, Uxbridge, UK

December 12, 2013

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. I authorize Brunel University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Signature: _____ Date: December 12, 2013

Nancy Elizabeth Brown Connolly

ABSTRACT

This is a foundational study that applies Receiver Operating Characteristic (ROC) analysis to the evaluation of a chronic disease model that utilizes Remote Monitoring (RM) devices to identify clinical deterioration in a Chronic Obstructive Pulmonary Disease (COPD) population.

Background: RM programmes in Disease Management (DM) are proliferating as one strategy to address management of chronic disease. The need to validate and quantify evidence-based value is acute. There is a need to apply new methods to better evaluate automated RM systems. ROC analysis is an engineering approach that has been widely applied to medical programmes but has not been applied to RM systems. Evaluation of classifiers, determination of thresholds and predictive accuracy for RM systems have not been evaluated using ROC analysis.

Objectives: (1) apply ROC analysis to evaluation of a RM system; (2) analyse the performance of the model when applied to patient outcomes for a COPD population; (3) identify predictive classifier(s); (4) identify optimal threshold(s) and the predictive capacity of the classifiers.

Methods: Parametric and non-parametric methods are utilized to determine accuracy, sensitivity, specificity and predictive capacity of classifiers Saturated Peripheral Oxygen (SpO_2), Blood Pressure (BP), Pulse Rate (PR) based on event-based patient outcomes that include hospitalisation (IP), accident & emergency (A&E) and home visits (HH).

Population: Patients identified with a primary diagnosis of COPD, monitored for a minimum of 183 days with at least one episode of in-patient (IP) hospitalisation for COPD in the 12 months preceding the monitoring period.

Data Source: A subset of retrospective de-identified patient data from an NHS Direct evaluation of a COPD RM programme. Subsets utilized include classifiers, biometric readings, alerts generated by the system and resource utilisation.

Contribution: Validates ROC methodology, identifies classifier performance and optimal threshold settings for the classifier, while making design recommendations and putting forth the next steps for research. The question answered by this research is that ROC analysis can provide additional information on the predictive capacity of RM systems.

Justification of benefit: The results can be applied when evaluating health services and planning decisions on the costs and benefits. Methods can be applied to system design, protocol development, work flows and commissioning decisions based on value and benefit.

Conclusion: Results validate the use of ROC analysis as a robust methodology for DM programmes that use RM devices to evaluate classifiers, thresholds and identification of the predictive capacity as well as identify areas where additional design may improve the predictive capacity of the model.

DEDICATION

To my parents, Frank Joseph Brown and Elizabeth Jean Mc Sorley-Collins Brown, for creating an environment that supported curiosity and a sense of wonder in the world. To my late husband Michael P. Connolly for telling me on many occasions that I should do this, for always supporting my endeavours and for his love of life. To my children and grandchildren, Christopher Monaghan Connolly, Eric Monaghan Connolly, Tara Kelly Connolly-Carfrae and Alex Carfrae, Michael Seamus and Jacolyn Anne Carfrae, who bring me purpose and great joy with occasional exhaustion. To my husband, Steven R. Shannon for his astute observations and for his thoughtful comments, support and patience! You are all blessings in my life!

Thank you for your support and encouragement and for tolerating my absence and preoccupation with the process.

ACKNOWLEDGEMENTS

I wish to acknowledge my gratitude to the following people who have assisted me immeasurably along this journey:

Dr. Vincent Gallicchio, for his support and expertise in initiating this process.

Dr. Malcolm Clarke, who presented me with an opportunity and whose friendship and guidance have assisted me in completing this research.

Ms. Jacolyn Anne Connolly, for her friendship, support, advice and editorial expertise.

Monsignor Joseph Hernandez for his unwavering support, guidance and common sense.

Dr. Ellen-Hope Kearns, my friend and role model. Whose tireless belief and support are greatly appreciated.

Dr. Joseph P. Connolly, Patricia Connolly R.N., Dr. John Y. Connolly and Marybeth Connolly R.N., P.N.P., for their encouragement, unwavering support and love of their brother.

William K. Appelgate Ph.D., friend and colleague, who challenged my thought processes.

Joanna Fursse, Ph.Dc. and Russell Jones M.D., Chorleywood Health Centre, for their thoughtful discussions during this research process.

To the Department of Engineering staff, with special note to Ela Haney, for assisting me in staying on track with administrative tasks.

To all the many friends, colleagues and family who have both challenged, nudged and supported me through this process: colleagues Dr. Janet Grossman York and Janet Leifur R.N., M.S.N.; Eileen Brown Kohlhepp and Rosemary Galbally Endres, cousins who are true sisters of the heart; colleagues at the Telemedicine and Advanced Technology Research Center, Dr. Ron Gimbel, COL (ret.) Hon S. Pak, Daniel Blum, Dr. Jon Luis Belard, and to the coffee club on Daniel Island.

Thank you one and all.

GLOSSARY OF ABBREVIATIONS

A&E	Accident and Emergency
AHRQ	Agency for Healthcare Research and Quality
AUC	Area Under the Curve
BP	Blood Pressure
CCM	Chronic Care Model
CDC	Center for Disease Control and Prevention
CDS	Clinical Decision Support
CI	Confidence Interval
COPD	Chronic Obstructive Pulmonary Disease
CHF	Congestive Heart Failure
DBP	Diastolic Blood Pressure
DM	Disease Management
DSS	Decision Support System
EU	European Union
FN	False Negative
FPR	False Positive Ratio
FP	False Positive
GDP	Gross Domestic Product
HH	Home Health visit
HIPAA	Health Insurance Portability and Accountability Act

HR	Heart Rate
IP	In-Patient
ICT	Information Communications Technology
IOM	Institute Of Medicine
ISO	International Standardization Organization
IT	Information Technology
MS	Microsoft®
NHS	National Health Service
NHSD	National Health Service Direct
NIH	National Institutes of Health
PR	Pulse Rate
PCT	Primary Care Trust
PPV	Positive Predictive Value
TN	True Negative
TP	True Positive
QOL	Quality of Life
RM	Remote Monitor
ROC	Receiver Operating Characteristic
ROI	Return on Investment
SBP	Systolic Blood Pressure
SOB	Shortness of Breadth
SpO_2	Saturated Peripheral Oxygen

TN	True Negative
TP	True Positive
TPR	True Positive Ratio
UK	United Kingdom
US	United States of America
USD	United States Dollar
VHA	Veteran's Health Administration
WHO	World Health Organization
WSD	Whole System Demonstrator projects

DEFINITIONS	
Classifier	In this research, SpO_2, blood pressure and pulse rate
Area under the curve	Plot of the cumulative distribution function (area under the curve (*AUC*) of the detection probability in the y-axis versus the cumulative distribution function of the false alarm probability in x-axis (Swets 1996).
Clinical decision support	The use of a computer to bring relevant knowledge to bear on the health care and well-being of a patient (Greenes 2007).
Chronic disease	A disease that persists for a long time. A chronic disease is one lasting 3 months or more according to the CDC U.S. National Center for Health Statistics (CDC 2013).
Chronic Obstructive Pulmonary Disease	Chronic Obstructive Pulmonary Disease (COPD) is not one single disease but an umbrella term used to describe chronic lung diseases that cause limitations in lung airflow. The more familiar terms 'chronic bronchitis' and 'emphysema' are no longer used, but are now included within the COPD diagnosis. The most common symptoms of COPD are breathlessness, or a 'need for air', excessive sputum production, and a chronic cough. However, COPD is not just simply a "smoker's cough", but a under-diagnosed, life threatening lung disease that may progressively lead to death (WHO 2013).
Class skew	Statistics (of a statistical distribution) not symmetrical. (Oxford dictionary)
Confusion matrix	A confusion matrix summarizes the classification performance of a classifier with respect to some test data. It is a two-dimensional matrix, indexed in one dimension by the true class of an object and in the other by the class that the classifier assigns (Springer reference 2013).
Disease management programme	Disease management consists of a group of coherent interventions designed to prevent or manage one or more chronic conditions using a systematic, multidisciplinary approach and potentially employing multiple treatment modalities. The goal of disease management is to identify persons at risk for one or more chronic conditions, to promote self management by patients and to address the illnesses or conditions with maximum clinical outcome, effectiveness and efficiency regardless of treatment setting(s) or typical reimbursement patterns (Schrijvers 2009).
Decision model	Describe the relationship between all the elements of a decision - the known data (including results of predictive models), the decision and the forecast results of the decision - in order to predict the results of decisions involving many variables (Wikipedia, Predictive analytics 2010).
Diagnostic accuracy	Diagnostic accuracy is correctly classifying subjects into clinically relevant subgroups. Diagnostic accuracy refers to the quality of the information provided by the classification device (Pintea & Moldovan 2009).
Exacerbation	For this research exacerbation is defined as worsening respiratory symptoms requiring treatment and management as evidenced by a home visit, accident and emergency visit or in-patient hospitalisation.

DEFINITIONS	
False positive rate	Created by plotting the fraction of false positives out of the negatives (FPR = false positive rate), at various threshold settings. FPR is one minus the specificity or true negative rate. (Swets 1996)
Long-term Condition	Any health condition that cannot at present be cured, but can be managed with medicines and/or therapy. This includes conditions such as diabetes, heart failure, COPD, arthritis, depression (Department of Health UK 2012)
Negative likelihood ratio	The negative likelihood ratio is the ratio between the probability of a negative test result given the presence of the disorder and the probability of a negative test result given the absence of the disorder (Pintea & Moldovan 2009).
Non-parametric statistics	A statistical method wherein the data is not required to fit a normal distribution. Nonparametric statistics uses data that is often ordinal, meaning it does not rely on numbers, but rather a ranking or order of sorts (Investopedia 2013).
Negative predictive value	The negative predictive value is defined as the probability that the disorder is not present when the result of the test is negative (Pintea & Moldovan 2009).
Parametric statistics	Parametric statistics is a branch of statistics that assumes that the data has come from a type of probability distribution and makes inferences about the parameters of the distribution (Geisser 2006).
Positive likelihood ratio	The positive likelihood ratio is the ratio between the probability of a positive test result given the presence of the disorder and the probability of a positive test result given the absence of the disorder (Pintea & Moldovan 2009).
Positive predictive value	Positive predictive value, also called precision, is defined as the probability that the disorder is present when the result of the test is positive (Pintea & Moldovan 2009).
Remote monitoring	Electronic sensors or equipment that monitors vital health signs remotely, e.g. in your own home or while on the move (Department of Health UK 2012).
Sensitivity	Sensitivity, also called the true positive rate (when expressed as a percentage) is defined as the probability that a test result will be positive when the disorder is present (Pintea & Moldovan 2009).
Specificity	Specificity, also called the true negative rate (when expressed as a percentage), represents the probability that a test result will be negative when the disorder is not present (Pintea & Moldovan 2009).
SpO_2	Saturation of peripheral oxygen, Oxygen Saturation, Saturation of Hemoglobin with Oxygen as measured by Pulse Oximetry (Free dictionary 2013).
True positive rate	Created by plotting the fraction of true positives out of the positives (TPR = true positive rate) vs. the fraction of false positives out of the negatives (FPR = false positive rate), at various threshold settings. TPR is also known as sensitivity. (Swets 1996).
Verification bias	A type of measurement bias in which the results of a diagnostic test affect whether the gold standard procedure is used to verify the test result (Begg 1983).

LIST OF TABLES AND FIGURES

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

"... knowledge must continually be renewed by ceaseless effort, if it is not to be lost. It resembles a statue of marble which stands in the desert and is continually threatened with burial by the shifting sand. The hands of service must ever be at work, in order that the marble continue to lastingly shine in the sun. To these serving hands mine shall also belong."

Albert Einstein, On Education, 1950

1.0 INTRODUCTION AND OVERVIEW

The use of remote monitoring (RM) devices as a part of disease management (DM) programmes is proliferating. These devices are being implemented as part of a broader chronic DM strategy in healthcare. The goal of these DM models is to identify clinical deterioration early enough to allow clinicians to intervene and avoid episodes of clinical decompensation that impact negatively on the patient and attribute high costs to the healthcare system. This research will apply a rigorous methodology to validate the use of the receiver operating characteristic (ROC) in a chronic obstructive pulmonary disease (COPD) population, based on patient outcomes, to evaluate the performance of the DM system that uses RM devices.

1.1 Research purpose

The purpose of this research is to evaluate the utility, predictive capacity and overall performance of the DM system using ROC analysis, as a tool, to evaluate the DM model that uses RM devices for a COPD population, based on patient outcomes.

1.2 Research question

The research question is: does ROC analysis provide greater utility in the evaluation of model performance of DM programmes that use RM devices?

1.3 Problem addressed

The problem addressed in this research is that current evaluation methods for technology-assisted DM programmes that use RM devices lack rigorous methodology and processes to evaluate the efficacy of their performance and justify the long-term investment. In addition, the DM models monitor a number of biometric parameters or classifiers, but it is not known, based on specific diseases and patient outcomes, which of these are the most useful or predictive for the disease in question.

Assessment of the DM models that use RM devices in the COPD population remains inconclusive (Buntin 2006; Busse, Blumel, Scheller-Kreinsen & Zentner 2010; Mattke, Seid & Ma 2007; Webb & Howson 2006) and there are no studies that apply ROC analysis to RM systems.

ROC methodology is investigated to determine its potential to provide more robust evaluative data and analysis of the best classifiers and appropriate thresholds for the classifiers, as these are critical indices to evaluate the performance of RM systems to identify clinical deterioration. The utility of ROC analysis in other clinical areas that rely on the interpretation of signals lends this method to RM devices, which are essentially signalling devices that generate alerts based on a threshold. There is an absence of research in its use in DM systems, although ROC has the potential to provide salient knowledge to inform the design and operational features of these systems and to improve their efficacy and predictive capacity (Linden 2006). The results of this research support the use of ROC analysis, and identify areas where re-design would provide better value to both the patient and the healthcare system by adjustments to the classifiers for specific disease states and to the optimum thresholds.

The current lack of predictive value in the methods used to evaluate the performance of the new DM programmes that use RM devices leaves both the providers and the purchasers not knowing when systems and configurations will provide the best predictive value of a deteriorating clinical event for their population, what devices and classifiers work best and unable to evaluate the cost of investing in these systems relative to their expected value.

1.4 Objectives

The overarching goal of this research is to determine if ROC analysis can be applied to DM systems that use RM devices, and provide additional utility and more accurate assessment in evaluating these new models of care.

Objective 1: Apply ROC analysis to evaluation of a RM system

Objective 2: Analyse the performance of a disease management model for prediction of patient outcomes in a COPD population

Objective 3: Identify predictive classifier(s)

Objective 4: Identify the optimum threshold for classifier(s) of interest and the predictive capacity of the classifiers.

1.5 Hypotheses

This work will investigate the accuracy of the alert generated in a DM system using ROC analysis. Results are based on patient outcomes, and will indicate the degree of predictability and will determine the performance of the classifiers.

H_1: ROC analysis is applicable to the problem of evaluating a technology-assisted DM model that utilizes RM devices and will provide additional utility to evaluate the performance and predictive capacity for the specific classifiers being used to monitor patients with COPD.

H_0: The null hypothesis is that the ROC analysis will not provide any additional information in which to evaluate the DM models that utilize RM devices for COPD.

1.6 Research stages

This research is multi-faceted and was conducted in three stages. The first stage included the development of a proposal for a COPD programme evaluation for the NHSD. The proposal was completed in June 2010. The NHSD operationalised a RM programme in November 2010 for COPD in two primary care trusts (PCTs).

Stage two included an evaluation report prepared for the NHSD and the PCTs involved in the project. The evaluation was conducted by Brunel University and Chorleywood Health Centre for the NHSD and was completed in August 2012. The evaluation team members included Ms. Joanna Fursse and Russell Jones, M.D. from the Chorleywood Health Centre, Malcolm Clarke, Ph.D., and Nancy E. Brown Connolly, R.N., M.S., from Brunel University. Stage three utilized a

subset of data from the NHSD COPD programme evaluation to explore the use of ROC analysis in a RM system.

1.7 Structure of the thesis

This thesis describes the problems healthcare systems are facing due to our aging population, the changes that will be required in healthcare systems and how programmes may utilize RM technology in DM programmes. This research contributes to knowledge by providing a new tool to be used in the evaluation of the performance of DM programmes that utilize RM devices. Chapter 2 describes the background of the study and discusses details of the DM programme that are the basis for this study.

Chapter 3 presents a search of the current literature, specifically for RM technology used in DM programmes and the use of ROC analysis in evaluation. An in-depth description of ROC, its utility, a description of ROC space and interpretation of ROC space and measures are provided.

Chapter 4 outlines the methods, assumptions and processes used to create the data sets for ROC analysis. The methodology for ROC analysis is well defined mathematically and is also reviewed.

Chapter 5 presents an analysis of the data and results. Chapter 6 includes a discussion, Chapter 7 presents conclusions, contribution to knowledge and recommendations for the further use of ROC in the design and evaluation of DM programmes.

1.8 Potential contribution

This research will contribute to a more accurate evaluation process that can be applied to RM systems used in DM programmes. ROC analysis, as a methodological, validated tool has the potential to empirically illustrate the performance of the RM system and allows for comparisons to be made between two or more classifiers. The use of ROC curves enables comparison of the

performance of the classifiers at varying thresholds and the optimum operating point to be determined. This type of analysis will help identify design and operational process issues which may need further development or adjustment within the overall system.

Trade-offs in performance would be measurable and vendor claims would be verifiable based on an evaluation of system performance that is based on patient outcomes. Methods can be applied to programmes utilising RM systems for other chronic diseases to identify the performance of classifiers and the value that is added to the healthcare system. The generalizability of results to other health systems and countries is equally significant. It is expected that the results will be valuable in enabling informed decisions by purchasers, insurers, health systems and policy makers with regard to the effectiveness of these systems. For those organizations that purchase DM services, this research will provide a substantive background with which to discuss the inclusion of ROC as an integral component of the programme evaluation with their contracted vendors.

Health care planners and commissioners would be better able to make informed decisions based on outcomes and value, including predictive cost analysis when purchasing and planning disease management programmes. Medical resources are hard pressed to manage our current populations with chronic disease and as the aging population increases, they will need to integrate validated models for chronic conditions in order to be able to increase efficiencies without sacrificing quality and personal care for patients. Other variables such as the scalability of the programme and its long-term sustainability, local readiness factors and infrastructure will need to be considered in the cost valuation going forward, but first there is a need to find a better way to analyse and evaluate the performance of the RM systems.

The programmes that are implemented must exhibit value and benefit to patients and the healthcare system alike in order to responsibly manage healthcare resources. In addition, systems

must be flexible enough to support organizations ranging from solo practitioner offices to national integrated delivery networks, improve workflow, reduce cost, and improve the quality of care, while maintaining long-standing beneficial patterns of communication, collaboration, and care (Avison & Young 2007).

CHAPTER 2

BACKGROUND

"The task is not so much to see what no one yet has seen, but to think what nobody yet has thought about that which everybody sees."

Arthur Schopenhauer, 1788-1860

2.0 BACKGROUND

The use of technology to support the provision of medical treatment and management for persons with chronic disease conditions has been gaining acceptance and is viewed as a necessity if we as a society are to meet the needs of our aging population (Kalorama Information 2013).

Manuel Castells, the sociologist, advances the idea that we are in the midst of a third industrial revolution that has at its roots citizens, knowledge and information technology (Castells 1997). This is evident in the uptake of technology in all sectors of our social institutions as well as among individuals in society. Imagine having no cell phone today! No e-mail or computer at work! No cash machine during a bank holiday or while on vacation in another country. Information communications technology (ICT) and the Internet are changing the way we live and work. You have only to look at the generation below 18 years of age and their ease in using technology along with the new ideas that are being generated to realize that we are indeed in the midst of a revolution.

The provision of healthcare is a part of this revolution and one that has historically been slow to change, but those changes are in process, and the pace will be accelerated as we learn to absorb new applications into our healthcare systems and use them for the patients' benefit. We, our families and our neighbours are the patients. We will be using these technologies and it behoves us to ensure that they are the appropriate applications, developed, designed and used in ways that enhance our quality of life and well-being, since we will be the end-users.

As we age, we develop, either through lifestyle, environment or our genetic inheritance, medical conditions that we must learn to live with and manage. These chronic conditions, such as hypertension, heart disease, diabetes and respiratory diseases, necessitate medical advice and services. The management of chronic conditions can be time-consuming, inconvenient, costly and

difficult for people, especially for frail individuals and their families. Timely access to services is necessary to prevent acute episodes and the prompt identification and management of deteriorating health status. Delays in treatment can also lead to clinical deterioration and episodes of hospitalisation.

Most people would prefer to be treated at home, or at least enable easier management of their condition, without having to inconvenience family and friends; in other words, people would like to be able to maintain their independence. While working on a technology-assisted DM project in Iowa, in the US. I was asked by an older woman to explain what I was doing; when I finished a lengthy description of my work, she asked, quite simply, "Will this help me stay in my home?" (Personal communication, 2004) I answered, "Yes".

2.1 Scope of the problem

Healthcare systems are facing the onslaught of an aging population with increases in chronic disease rates that they are ill-prepared to manage. A study that included health trends for 50 countries in the next 25 years determined that trends would mainly be influenced by the aging of the world's population (Murray & Lopez 1997). COPD is projected to become the fourth leading cause of death globally by 2030 (Mathers & Loncar 2006).

The European Observatory on Health Systems and Policy, in a comprehensive review of multiple European countries and the United States notes,

> "Chronic diseases are the leading cause of mortality and morbidity in Europe, and research suggests that complex conditions such as diabetes and depression will impose an even larger burden in the future. Some years ago chronic diseases were considered to be a problem of the rich and elderly population. Today we know that within high-income countries, poor as well as young and middle-aged people are affected by chronic conditions. The economic implications of such diseases are also serious. Chronic diseases depress wages, earnings, workforce participation and labour productivity, as well as increasing early retirement, high job turnover and disability. Disease-related impairment of household consumption and educational performance has a negative effect on gross domestic product (GDP). As expenditure on chronic care rises across Europe, it takes up

increasingly greater proportions of public and private budgets." (Buss, Blümel, Scheller-Kreinsen & Zentner 2010, p.16).

A snapshot of the chronic disease burden from around the globe includes; Australia where 12 chronic diseases accounted for US $11.0 G, or 22.4% of the total allocated healthcare expenditure. Expenditure on heart disease was US $1.5 G alone. (Australian Institute of Health & Welfare, accessed 22 November 2010). In the UK COPD direct costs are estimated to be £492M in annual costs (Alder, Mayhew, Moody & Morris 2005); the United States spends 75% of its' healthcare expenditures in chronic disease care (CDC 2009). This is an increase of 5% from the CDC 2004 dataset (CDC 2004). The World Health Organization (WHO) estimated accumulated losses from 2005 to 2015 in USD, for China of $558 G, for India $236 G, and for the Russian Federation $303 G (WHO 2010).

With the current state of healthcare systems in some disarray, due in part to the complexity of our healthcare systems and the need to redesign these systems for our populations, what strategies are being tested to address the needs of the chronically ill and aging population?

2.2 Strategy to address chronic disease

Structured DM programmes are one strategy being used to manage chronic diseases. Developing and deploying DM programmes designed to monitor patients in their homes is being used to address inadequate access to care and escalating cost. Many countries with an aging population are struggling to provide the needed services. The complexity of the healthcare system, transportation and other logistics become problematic for the person with a chronic condition. Many healthcare systems have a demand versus capacity issue that is creating a barrier to access to medical care. The European Observatory on Health Systems and Policy presented an overview of the crisis faced by countries that are struggling to provide services, and outlined the responses

from budget cuts to increasing efficiencies, access and wait time increases, reform and private pay options (Figueras 2012).

RM has been introduced as an enabling technology to facilitate data gathering and provide clinical decision support as a part of a DM programme. DM programmes were originally developed as part of a chronic care model (CCM), Figure 2.1. The model has proliferated in European countries. Of note is the progression of the six areas of the model within the healthcare system. Self-management support and delivery system design precede the application of a decision support system (DSS) (Wagner 1998). An issue with RM systems has been the industry push for adoption without a rigorous design methodology or evaluation of the design for specific chronic diseases.

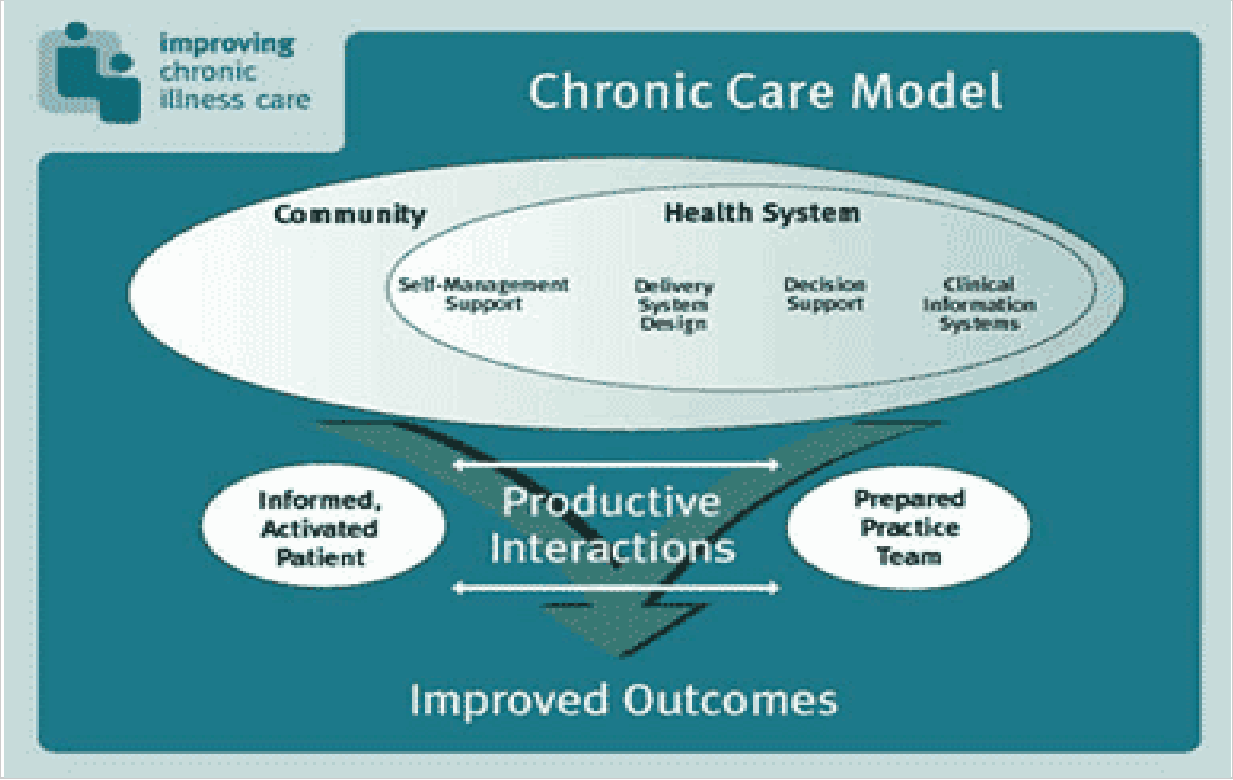


Figure 2.1. Chronic care model (Wagner 1998)

In 2008, the National Health Service (NHS) in the UK rolled out the Whole System Demonstrator (WSD) project using a randomized controlled trial in three PCTs to address and evaluate chronic disease care using remote monitoring (Department of Health Gateway 2011).

> "The early results indicate that telehealth can deliver a 15% reduction in A&E visits, 20% reduction in emergency admissions, 14% reduction in elective admissions, 14% reduction in bed days and an 8% reduction in tariff costs. More strikingly, results demonstrated a 45% reduction in mortality rates." (Department of Health 2011, p. 3)

The WSD project in the UK evidenced significant impact on the use of services in chronic disease care using RM systems (Department of Health 2013). In the US, the Veteran's Health Administration (VHA) deployed a range of RM programmes including COPD and estimated that 50 % of the patient population could be cared for with RM technologies by 2011 (Darkins, Ryan, Kobb, Foster, Wakefield & Lancaster 2008; New England Healthcare Institute 2013).

However, there remains an issue with the accuracy of the alerts generated by the DSS. It is not known how changes to the DSS, biometric measures (classifiers), and thresholds (biometric parameters) would impact on the accuracy of the alert and the effect on patient outcomes. Improving accuracy would include consideration of how design enhancements might impact on the overall performance of the RM system and its effect on the healthcare system.

2.3 System design

Developers of RM systems have advocated the use of their respective systems with little research on system design and performance, and instead design the programme from the perspective of the technology.

However, the physiology of each disease is consistent and a RM system designed to provide identifying data for specific chronic diseases that is based on medical indicators, should reflect those clinical indicators shown to be predictive for the disease.

It is essential that the technology being used is supported by good clinical care and it is only within the context of its application to a physiologic disease process, and capability to support treatment guidelines, that the technology can be of assistance. Without a framework for a well-designed programme, the programme and the technology will both fail to support the self-management by a patient while providing little value to the end-user or to the healthcare system. Although medical research has identified clinical indicators for COPD exacerbation, research to identify the predictive capacity of common classifiers and evaluate the design of the measures used in DM programmes that use RM has not been done. System design lags secondary to gaps in the research that has been conducted to identify the factors that may improve the DM system performance for patient outcomes, appropriate classifiers and measures in specific populations and acceptability by the health system and end user.

As an example, this research was unable to evaluate the self-reported respiratory measure for shortness of breath (SOB), cough or sputum production using ROC analysis due to the design. The design of the respiratory questions did not use a scale that could be translated within the DSS to quantify a change in respiratory effort or sputum production that could be issued as an alert for the healthcare professional nor be linked to patient outcomes. A scaled measure and routine timeframe for questions would be needed and could be included within the DSS as part of an algorithm to identify the risk for COPD exacerbation.

The two classifiers, of cough (Foreman, DeMeo, Hersh, Reilly & Silverman 2007) and sputum production, (Burgel, Nesme-Meyer, Chanez, Caillaud, Carré, Perez & Roche 2009) are known to be predictive of COPD exacerbation and indicate clinical deterioration. Inclusion of these measures into the design of the DSS has the potential to improve prediction of COPD exacerbation. Improved prediction would have further impact on patient outcomes and this may

decrease the costs associated with false alerts as well as free-up additional time in the Care Matron's schedule to focus on patients with a greater risk of clinical deterioration.

The health system must consider the cost and predictive capacity in order to allocate resources. Clinicians need to know that the system provides reliable information that is designed for the targeted population, and that can improve health outcomes through timely alerts. It must be easy to use for the patient and acceptable as a tool to improve the patient's access to healthcare services when needed.

2.4 Healthcare system change

Professional organizations are seeking change and identifying the need to adopt new strategies and methods that use engineering techniques and evaluative processes to become more efficient. Abbasi and colleagues note that professional organizations such as the Royal Society of Medicine in the UK, and the Institute of Medicine (IOM) in the US are calling for a systems approach to healthcare and leadership to manage the change (Abbasi 2010; Reid, Compton, Grossman & Fanjiang 2005).

The National Academy of Engineering in the US and the IOM have also called for the integration of engineering applications in healthcare, and to identify areas of research that could contribute to rapid improvements in healthcare. Both organizations recommend the use of currently available systems engineering tools, as well as the development of new tools through research (Reid 2005). The IOM in the United States noted that,

> "Healthcare is substantially underperforming on most dimensions: effectiveness, appropriateness, safety, cost, efficiency, and value. Increasing complexity in health care is likely to accentuate current problems unless reform efforts go far beyond financing, to foster significant changes in the culture, practice, and delivery of health care. If the effectiveness of health care is to keep pace with the opportunity of diagnostic and treatment innovation, system design and information technology must be structured to assure application of the best evidence, continuous learning, and research insights as a natural by-product of the care process. In effect, the nation needs to engineer the development of a

> learning healthcare system—one structured to keep the patient constantly in focus, while continuously improving quality, safety, knowledge, and value in health care. Striking transformations have occurred through systems and process engineering in service and manufacturing sectors—e.g. banking, airline safety, and automobile manufacturing. Despite the obvious differences that exist in the dynamics of mechanical versus biological and social systems, the current challenges in health care compel an entirely fresh view of the organization, structure, and function of the delivery and monitoring processes in health care." (IOM, Engineering a Learning Healthcare System 2005, p.1).

ROC is one engineering tool that has shown broad application in healthcare measurement and its validation as a tool to be used in RM systems, will be useful in developing designs that improve clinical decisions, resource allocation and ultimately improve patient outcomes.

2.4.1 Underlying Issues

"Why have healthcare systems been slow to change"? There are several underlying issues. These include: (1) a lack of empirical evidence, i.e. measurement of the effectiveness of the RM systems' overall performance in specific chronic diseases; (2) challenges to existing assumptions; (3) system design and; (4) cost and value to patients and the healthcare system (Buntin 2006; Duncan, Owen & Dove 2006; Mattke, Seid & Ma 2007; Webb & Howson 2006). This research addresses several issues in an effort to add some clarity and metrics to the discussion. These are effectiveness of the performance of the RM system and of the classifiers in relation to patient outcomes, and threshold settings for the classifier.

The first issue is empirical evidence. Claims by industry vendors and programmatic evaluations evidence shortcomings in the methodology. RM programmes are relatively new, and often leave gaps in the evaluation process (Buntin 2006; Duncan, Owen & Dove 2006; Mattke, Seid & Ma 2007; Webb & Howson 2006). Additionally, programmes are managed in a variety of healthcare settings and systems, such as doctor's offices, larger multi-specialty clinics, hospitals, as sub-contracted or commissioned services, in nationalized health services, and private health

insurance companies. This variety of settings does not lend itself to applying similar metrics and processes to an evaluation, since designs are reflective of the multiple environments.

Furthermore, the expertise to undertake the necessary evaluation in all settings is markedly absent. This lack of standard metrics for quality of service assessment impedes evaluation and obscures the progress of technology adoption and utility (Ackerman, Filart, Burgess & Poropatich 2010). A major difficulty is the inability to predict the potential impact that can be expected as well as the limitations of current methodology (Buntin 2006).

In a 2005 issue brief, the American Academy of Actuarials noted that,

> "Escalating health care costs and an increasing public focus on health care quality are causing employers and insurers to reassess the value and effectiveness of their medical management procedures. Many are looking at DM programs as a means for improving the treatment of major chronic diseases as well as the quality of life, while reducing the need for and the costs of medical care. … there is often a gap between the favourable clinical results and a clearly identifiable financial impact" (Duncan 2005, p. 1).

The Academy has not assessed or explored the new technology-assisted programmes that use RM, nor evaluated ROC analysis in its assessment procedures.

Another limitation is that not all technology-assisted programmes use the same technologies. The outcome however, in terms of the impact on a patient's well-being should be comparable for all programmes. Any programme that elicits a signal or "alert" should be amenable to ROC analysis. Analysis using ROC techniques allows the DSS to be evaluated to identify criteria changes such as the classifiers that have predictive capacity for the chronic disease that they are used to monitor. This might improve the outcome for a specific population and save health resources in terms of the personnel needed. Differences in other parameters such as scalability and long-term cost would be more easily identified and enable better decision-making.

This lack of good evaluation and research methodologies leaves the results of DM programmes that utilize RM systems open to question in terms of their efficacy. Healthcare

systems are slow to change and require validation in approaches and the use of these programmes that are methodologically strong and evidence positive effects on patient outcomes in order to change. Research to evaluate these models for DM is lacking both in specific diseases and in the methods employed to evaluate results that are based on patient outcomes and between different technologies. In addition, specific programme design elements such as, identifying the optimum threshold settings for the alerts in specific diseases, as well as the identification of the predictive capacity of the classifier has not been done. This decreases the effectiveness as well as the predictive capacity of the system.

There is a need to study the automated DSS in a way that can view several real-time or near real-time datasets simultaneously for chronic disease states using RM systems, that can assist providers with a timely detection of outliers (Ackerman, Filart, Burgess & Poropatich 2010). An outlier in the context of this research refers to a patient at-risk for clinical deterioration. This research will use patient outcomes to evaluate the effectiveness of the RM system. Outcomes are defined as home health visits (HH), accident & emergency (A&E) visits and in-patient (IP) episodes.

The second issue is that the development of new delivery models that use RM devices challenges the existing assumptions in the medical system. These include assumptions about the location where care can best be provided, as well as the episodic nature of that care. It will necessitate making further changes to the work processes (Speedie, Ferguson, Sanders & Doarn 2008).

The third issue is system design. Analysis based on patient outcomes can help to identify additional measures for specific diseases which if applied through re-design, would improve the

performance of the system, including improved prediction of clinical deterioration. First, we must evaluate current design more rigorously to inform future development of RM systems.

Lastly, health systems are spending billions of dollars implementing DM programmes (Alder 2005; CDC 2009). The UK invested £31 M in the Whole System Demonstrator (WSD) project that utilises RM. (Whole System Demonstrator Program, accessed December 1, 2010). The US market for remote patient monitoring is forecasted to reach US $296.5 M by 2019 (Monegain 2013). As the number of elderly individuals in the world continues to rise, chronic diseases, including COPD, are increasingly over-represented in hospital populations. In fact, according to a recent study, 63% of the patients with a chronic condition will need some form of home care. (Kane, Chen, Finch, Blewett, Burns & Moskowitz 2000).

However, technology and its evaluation is not the only barrier. The business model in a nationalized health system addresses one barrier to adoption which is evident in the US; this is the issue of reimbursement (New England Healthcare Institute 2009). National health systems can show immediate benefit with fewer IP hospitalisations and A&E visits. The ensuing cost reductions accrue directly to the NHS and PCTs. In the US, multiple agencies are involved in patient care and the costs and the benefits do not necessarily accrue to the agencies that are the most frequent service providers for in-home DM services and RM, with the exception of the VHA. Cost is a barrier to RM uptake in healthcare systems. "The VHA estimates the cost per patient to be $1,600 USD per year." (Darkins, Ryan, Kobb, Foster, Wakefield & Lancaster 2008).

This research uses cost as an optimization criterion in creating the ROC curve. The return-on-investment (ROI) analysis is based on the patient outcomes and is expected to provide another tool to evaluate the impact of the cost and value of RM systems as well as contribute to DM

system design, irrespective of the vendor system or configuration of the healthcare system, whether national or commercial.

2.5 Remote monitoring system and processes

DM models that utilize RM, most frequently use the telephone in the patient's home to transmit data from the RM devices to a central server. The devices connect to an interface that transmits the data via telephone to the DSS. Clinicians access data on an Internet website and respond to alerts.

DM models utilize a DSS consisting of an enterprise application-server framework that is combined with a rules engine and statistical analysis tools (Berner 2009). This clinical decision system analyses the data base using predetermined parameters for the patient. The DSS identifies variance in parametric values. If the data received from the RM devices are outside the set parameters or thresholds, the DSS elicits an "alert". The alert will be evaluated by a healthcare professional in the context of the patient's history and presentation. Based on the level of variance the alert is identified as 1 - no risk, 2 - moderate risk, or 3 - high risk. The alert identifies the probability of clinical decompensation based on the thresholds that are programmed into the system. Communications to the patient are triaged by the alert system with high-risk alerts having priority.

Access to daily information is controlled by user identification and authority to view functionality. The DM systems are secure and compliant with privacy standards and have redundancy to protect from data loss. The follow-up of alerts and non-responses are completed by clinical personnel who contact the patient and evaluate whether a change in treatment or management is needed. This entails a second level of decision-making. The information flow is illustrated in Figure 2.2.

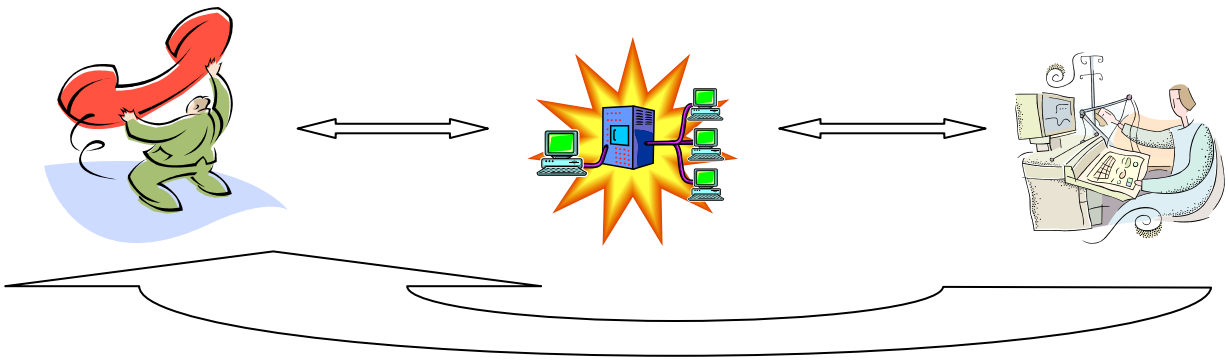


Figure 2.2. Information flow

The second level of decision-making occurs when the clinician assesses the patient by telephone. This second level of decision-making differs in that it includes judgmental decisions by the clinician based on their experience, training and valuation, so an element of uncertainty is introduced. It is determined that the patient may be deteriorating based on the classifier and clinician's expertise. This step in decision-making highlights the need for the right information based on a patient's clinical diagnosis and is critical in designing a patient-centred system. A decision is arrived at by clinical staff in favour of an HH visit, dispatching the patient to the A&E or IP hospitalisation. Performance of best classifiers set at the best threshold, based on the patient's clinical condition, is needed to provide clinical decision support (CDS) that assists clinicians in making treatment decisions.

2.6 Decision process

The decision begins with the signal of the patient's biometric data being transmitted to the DSS. Based on the value attributed to the signal, and the algorithm applied within the DSS, a risk level is established. If the risk level is outside the established parameters or thresholds, the DSS issues an alert that represents the probability of a clinical deterioration. This is followed by a response from the healthcare professional. This research evaluates the accuracy of the signal based on the patient outcome.

There are several decision points to consider. The first occurs through an algorithm or machine logic system within the DSS or "black box". The performance of the accuracy of this process will be visualized and evaluated but it is outside the scope of this research to evaluate the parameters or the architecture of the DSS. The next decision occurs when the patient's risk level is reviewed and a decision as to its accuracy is made. The Care Matron is alerted to the need for a clinical decision to either change treatment or management, or that no change is necessary. If the decision is that the patient is at risk for clinical deterioration, the Care Matron will schedule a home visit or refer the patient to A&E or IP services.

The next decision is a human decision process, which is influenced by the structure and site of the programme and the personnel administering the programme. For example, if operating under pre-determined protocols that use evidence-based practice guidelines with specialty nurses and prior knowledge of the patient's clinical history, the decision -- to treat or not to treat -- may differ from that taken in a call centre where the personnel managing the alerts may be non-licensed healthcare practitioners. In a call centre setting, non-professional personnel pass on information to another level of decision-makers (nurse or physician).

In the example of a call centre, there may be uncertainties in the human decision process that are introduced by a lack of prior knowledge of the patient's response to treatment or their level of knowledge and skills specific to the chronic disease or even interviewing skills. This is the situation for this research. The NHSD Telehealth Agent was the first line of decision-making that occurred in relation to the patient alarm. The final decision to visit or refer the patient was made by the clinical staff. All of these factors exert an influence on the decision to make an HH visit or referral to A&E or hospital. Evaluation and research for the secondary decision processes and influences are outside the parameters of this research but would be an area for further study.

2.7 Why ROC analysis?

The quote from Arthur Schopenhauer (1788-1860) that precedes this chapter is especially appropriate because ROC analysis is a well-known engineering application and a widely used method to evaluate the performance of binary classification models in engineering that has been applied extensively in clinical areas to visualize and analyse the behaviour of diagnostic systems, laboratory testing, and in the area of psychology (Swets 1988; Zou, O'Malley & Mauri 2007).

There are various methods to evaluate the performance of a system. Many are based on a 2x2 classification table known as a confusion matrix or decision matrix (the term confusion matrix will be used throughout this research). ROC analysis is one such method in this class and one that is widely used in healthcare. The use of ROC analysis in RM systems represents a maturity of thought in applying this method to a newly emerging area of healthcare. A detailed description of the confusion matrix and classification framework is given in Chapter 4, section 4.6. ROC analysis is described in detail in Chapter 3, section 3.4.

The attribute that makes ROC analysis especially appealing is that it is a method that is independent of the distribution of the data and therefore skew-agnostic. Skewing in the data for chronic disease is to be expected. There will be more true negative (TN) readings - this is due to the fact that people at home are chronically ill but stable, so more frequently will be within the biometric parameters programmed into the system. This skew in the data will increase the sensitivity and decreases the specificity, leading to results using a 2 X 2 classification table, that do not clearly demonstrate the performance of the classifiers (Krzanowski & Hand 2009).

ROC is ideally suited as a classification model for the types of data being assessed as COPD exacerbation presents as a binary classification problem. The goal is to identify the patient who will have a clinical exacerbation or not, and evaluation is seeking an unbiased estimate of how

well the RM model performs using a rigorous methodology. The aim of ROC analysis is to detect the presence of a particular signal, missing as few positive occurrences as possible while also identifying as few false alarms as possible (Krzanowski & Hand 2009). “Characteristic” refers to the characteristics of behaviour over the potential range of its operation (Krzanowski & Hand 2009, p.2). ROC analysis will graphically illustrate the performance of the classifiers at all thresholds allowing evaluation of how well the classifier predicts clinical deterioration.

CHAPTER 3

LITERATURE REVIEW

"If you would not be forgotten as soon as you are dead and rotten, either write something worth reading or do things worth the writing."

Benjamin Franklin, 1706-1790

3.0 LITERATURE REVIEW

This chapter reviews relevant studies and papers on ROC analysis in DM programmes that utilize RM. The topical areas include: ROC analysis applied to the assessment of RM programmes; ROC analysis processes; current evaluation methods for DM programmes; and clinical factors for COPD that need to be considered in the use of technology-assisted programmes in order to identify the potential areas for improvement in design.

3.1 Literature review process

A comprehensive search was conducted through electronic search facilitated by the Brunel University Library. The literature review included books and studies that appeared in peer-reviewed journals in the English language together with relevant reports and policy papers from the government, universities, and private organizations, Cochrane reviews in technology and healthcare, abstracts and meta-analysis. Additional searches included Google, Google Scholar and the Internet using keywords. The search covered the past ten years as the technology-enabled DM model has only emerged during that time frame. The background research on ROC applications and analysis was not time limited.

Identification of articles in the literature was approached in three stages: Stage one included a broad search of the Brunel library databases for journals and books on DM systems, ROC analysis, ROC analysis and medical programmes, DM evaluation and engineering applications in healthcare. In stage two, the search criteria were limited to MeSH terms, sub-headings or descriptors. Titles and abstracts were scanned for applicability to the research topic. Stage three included a review of full articles, reports and books. References in articles were used to generate additional sources of information.

The spelling of keywords was modified to include both American and British spelling, (e.g. program and programme). Keywords used included: RM, RM and DM programme, COPD RM programme, chronic disease and RM, receiver operating characteristic (ROC), ROC analysis and DM, evaluation and ROC, ROC analysis in healthcare, ROC and diagnostic performance, accuracy, sensitivity, specificity, predictive accuracy, ROC analysis and healthcare programme evaluation, ROC and model validation, ROC analysis and DM, DM evaluation, DM programme evaluation, DM and technology programme, healthcare technology programme, information communication technology (ICT) and healthcare programme, chronic disease programme evaluation, DM models, sensitivity and specificity, predictive modelling, DM outcomes and return-on-investment (ROI).

3.2 ROC analysis in disease management and remote monitoring

Despite the wide application of ROC analysis in other areas of healthcare, only one article was identified that used ROC analysis techniques in a RM system. Jensen and colleagues used ROC to analyse the predictive capacity of RM classifiers (SpO_2, BP, weight, and PR), to predict what the authors call, moving prediction of COPD exacerbation during a rehabilitation programme (Jensen, Cichosz, Dinesen & Hejlesen 2012). Moving prediction was defined as, "…prediction on a day-to-day basis." (Jensen, Cichosz, Dinesen & Hejlesen 2012, p.99). A major limitation of the study was that data were used from a study with another purpose and there was no consistent sampling protocol. Performance of the individual classifiers was not compared and the data provided in the article require further definition and information regarding methods for better evaluation. However, the conclusion was that the overall system had the capacity to discriminate. "The 70% sensitivity (SpO_2) is assessed as acceptable, since the alternative is no prediction at all." (Jensen, Cichosz, Dinesen & Hejlesen 2012, p.101). This statement highlights the need to further

refine RM systems to increase the predictive capacity. A sensitivity of 70% is a low level of performance and would be unacceptable in other areas of medical testing.

Research into the use of ROC did not identify any studies regarding the performance of the classifiers, overall assessment of RM systems using ROC to improve the predictive capacity or enhance design. This, despite the fact that there is a call in the literature to do so (Linden 2006). As Linden notes,

> "Receiver operator characteristic (ROC) analysis is a more appropriate and useful technique for assessing diagnostic and predictive accuracy in DM. Its advantages include; testing accuracy across the entire range of scores and thereby not requiring a predetermined cut-off point, easily examined visual and statistical comparisons across tests or scores, and independence from outcome prevalence. Therefore the implementation of ROC as an evaluation tool should be strongly considered in the various phases of a DM program." (Linden 2006, p. 132).

References and citations from the Linden and Jensen articles did not uncover any further studies that have applied ROC analysis to DM or RM programmes. Although Linden calls for ROC analysis for the programme elements, he does not address the use of ROC in DM programmes that use RM, nor does he identify the foundational research necessary to answer the question of whether ROC offers any advantage in predictive capacity for patient deterioration. The article does not address the question of identification of an optimum threshold nor identification of classifiers for any specific chronic disease.

One other pertinent article addressed the design of decision support architecture for patient management using RM and although there is discussion of signal quality, ROC analysis was not applied to evaluating the performance of the model. The article focused on development of the DSS with respect to patient outcomes (Basilakis, Lovell, Redmond & Celler 2010).

Basilakis et al. (2010) used a single case study extracted from an initial pilot trial of a DSS, in patients with COPD and CHF, to illustrate the potential benefit of integrating telehealth and

decision-support within a chronic DM technology-assisted programme. A description of the DSS, application server framework combined with a rules engine and statistical analysis tools, illustrates the value provided by this type of system in identifying the trends and shifts in parametric values for patients. The DSS provides a means to stratify health risk and target patients at risk for clinical deterioration and it is noted that this process influences changes in workflow by targeting scarce clinical resources (Basilakis, Lovell, Redmond & Celler 2010).

The article does not link the DSS performance to the outcomes for patients in the systems reviewed. This creates a gap in analysis since the end-point for all of these DM models needs to be based on the value provided to the patient. The accurate performance of the machine logic system within the DSS or "black box," is an important element to be visualized and evaluated, just as the article notes, but it is outside the scope of this research to evaluate the parameters and elements of the DSS. This would compromise the intellectual property rights of DM providers and the information is not freely available. However, DM providers would be able to follow-up with internal evaluations of their respective DSS based on patient outcomes following the methods employed in this research. Basilakis does identify the salient areas where additional research would be beneficial and notes,

> "Future DSS research work will focus on applying additional domain knowledge and improving the capability of the input and output modules, as well as improving and evaluating the robustness of the DSS analysis and risk stratification strategy. Research in this area is critical as telehealth systems become more widely adopted, and there is a need to screen large volumes of electronically monitored patient data effectively and efficiently. '

> 'At the signal level, refinement of the analysis techniques will identify specific pathological markers that would give further warning of a patient's health deterioration, such as a transition from a regular to an irregular heart-beat. Ongoing analysis of signal quality will further improve the robustness of the final DSS analysis. The use of machine-learning techniques will be investigated as an additional method for capturing domain knowledge, for instance, training sets of monitoring results against recorded clinical outcomes for a specific patient or clinical domain.'

> ‘Developmental work is currently aimed at expanding and refining the process flows for the workflow management system. As described earlier, the electronic clinical guideline or care pathway, will be critical in coordinating evidence based best practice telehealth management, in a way that can manage the resource needs of the health carers while at the same time being sensitive to the health requirements of the patient.’” (Basilakis, Lovell, Redmond & Celler 2010, p. 1224-1225).

There were no methods addressed that applied ROC analysis to evaluate the system performance nor was there mention of a process to apply ROC analysis to identify the optimum threshold values for a specific population.

3.3 Current evaluation methods in disease management programmes

Mattke et al. (2007) and colleagues at RAND in the US completed a meta-analysis of existing disease management programmes (Mattke, Seid & Ma 2007). They noted that the literature contained very little information about large, population-based disease management programmes that target the entire diseased populations and that use mass communication and information technology. There was no note of RM monitoring specifically or of ROC analysis in the evaluation. They were able to identify two meta-analyses and three reviews of 25 studies in COPD and noted that none of the five analyses addressed the effects on cost. There was also insufficient evidence to draw conclusions about the impact of DM programmes for COPD patients.

De San Miguel et al. (2010) completed a study in Australia to evaluate the impact and value of a COPD RM programme. The goal was to better evaluate cost and benefit. Results did not include an evaluation of the classifiers but did include the cost of the various monitoring devices (De San Miguel, Smith & Lewin 2010). This leaves a gap in the research. It is unclear if the right indicators for the disease are being measured. Cost analysis included potentially unnecessary devices that will have no contributory effect on the patient’s outcome.

A systemic analysis of respiratory conditions concluded that “evidence on the magnitude of clinical and structural effects remains preliminary, with variations in study approaches and an

absence of robust study designs and formal evaluations." (Jaana, Pare & Sicotte 2009). A subsequent article concluded that the knowledge of how to conduct a systematic review and meta-analyses in the area of home telemonitoring is not apparent (Spyros, Pare & Jaana 2013).

The National Institutes of Health (NIH) in the US funded a randomized control trial for a population-based technology-assisted disease management model that uses a DSS and a low-cost telephone voice over technology system (iVo). The system queries patients, using disease specific questions and patient specific parameters. The results indicated no significant difference in hospitalisation for CHF between the groups with and without telemonitoring. The study concluded that telemonitoring did not improve outcomes for heart failure patients (Chaudhry, Mattera, Curtis, Spertus, Herrin, Lin, Phillips, Hodshon, Cooper & Krumholz 2010). However, the study did not apply a consistent protocol for patient management and had a short six-month time frame for the data collection period.

The same iVo technology was used in a DM programme in Iowa, US (Hickman, Brown-Connolly, Garloff, Kunath & Appelgate 2004) and had consistent decreases in CHF specific IP hospitalisations of 86% over a three-year period, with significant decreases in all-cause hospitalisation as well. The difference in outcomes in the Iowa study may be explained by the inclusion of care management by nursing as a critical component of the system. This approach integrates the clinical decision process into the workflow using technology as clinical decision support, rather than the sole mechanism in identifying clinical deterioration. The logic systems in the DSS are not yet mature enough to perform as well as or better than experienced clinicians.

The difference in outcome from these studies identifies issues that cannot be answered by considering only the technology aspect, rather all factors need to be identified and considered.

However, systemic analysis, such as ROC, was not applied to evaluate the performance of the RM system nor was ROC utilized to evaluate the predictive capacity for clinical deterioration.

Current methods to determine the predictive value of DM programmes have clear limitations in the factors considered. The National Public Health Service for Wales in an international overview of effective service models in DM noted that although decision support and clinical information systems are an essential component in DM programmes, their efficacy as part of DM is not considered in peer reviewed journals (Webb & Howson 2006). Overall there is a lack of strong systematic evaluations for population-based chronic DM programmes (Buntin 2006; Busse, Blümel, Scheller-Kreinsen & Zentner 2010; Duncan 2005). As noted by Buntin, "Since the early 1990s, disease management has been one of the most heralded – and least rigorously evaluated developments in health service delivery." (Buntin 2006, p. 121)

Mattke et al. (2007) draw similar conclusions. Their comprehensive review found that many studies had methodological flaws, such as incomplete accounting for costs or a lack of a suitable control group. Even looking at the reported costs and the savings generated rarely identified any conclusive evidence that DM brought about net savings on direct medical costs (Busse, Blümel, Scheller-Kreinsen & Zentner 2010).

In review of the literature, it has been cited by multiple organizations and government level reviews that the ROI and efficacy of DM programme and ICT is unproven (Congressional Budget Office (CBO) 2004; Dove & Duncan 2005; OECD Health Policy Studies 2010).

This research considered the well-documented gaps in evaluation of DM models that use RM, and applied rigorous engineering evaluation method to technology-assisted DM model that use a DSS. This is highlighted by there being no further articles, papers or presentations that apply

the rigorous technique of ROC analysis to evaluate a RM programme linking results to the patient outcomes.

The gap in knowledge of programmes that utilize both RM technology and a DSS, and the lack of the use of engineering assessment methods, such as ROC analysis, is not surprising. DM models that use RM devices have been deployed over the past 12 years and indeed are still evolving. These technologies and new approaches require scrutiny and evaluation, like many technology applications that apply new paradigms to patient care; it will take time to fully realize their strengths and limitations.

3.4 Review of the receiver operating characteristic

ROC analysis is an engineering application and is a well-known method widely used to evaluate the performance of binary classification models. It has found wide use in healthcare areas with extensive use in psychology and in evaluation of diagnostic systems and laboratory tests (Lasko, Baghwat, Zou & Ohno-Machado 2005; Obuchowski 2003; Swets 1988). ROC analysis has also been applied to radiology, epidemiology and bioinformatics (Lasko, Baghwat, Zou & Ohno-Machado 2005; Lusted 1978; Obuchowski 2003).

ROC analysis is based on statistical decision theory and signal detection theory (Green & Swets 1966; Swets 1964; Wald 1950). Signal detection theory arose from the application of statistical decision theory to engineering problems, in particular, detecting a signal embedded in noise. (De Carlo 1998). Although initially developed in the field of statistical decision theory, ROC became well known for its use in signal detection theory during WWII where it was applied to distinguish radar signals from noise to identify enemy targets (Egan 1975; Peterson, Birdsall & Fox 1954; Swets 1973). There was a very real need to determine the probability that the signal was an airplane approaching London (a true positive), versus radar noise (a false positive) and the

ability to adjust settings to optimize performance toward desired outcomes. The number of correct and incorrect determinations is calculated as the true positive rate (TPR) versus the false positive rate (FPR).

Signal detectability specifies a mathematical ideal or optimal detection process, and provides a means of analysing the structure of the decision process while applying a quantitative method for the study of performance (Swets 1964). The assumption for signal detection is that all reasoning and decisions take place in the face of some uncertainty. The ROC curve is the graphic representation that provides a way of analysing the decision process (Heeger 2003). The mathematical calculations and theory behind ROC analysis are well documented and have been used since the 1950s. (Lasko, Bhagwat, Zou & Ohno-Marcado 2005; Swets 1986; Swets 1992; Zou, O'Malley & Mauri 2007). The calculations will be presented in Chapter 4.

An historical overview of the receiver operating characteristic, its refinement in signal detection and its acceptance in quantitative psychology is presented by Swets (1973). Green and Swets (1966), present an in-depth summary of the concepts and processes involved in signal detection theory.

> "The ROC graph was designed in the context of the theory of signal detectability by Peterson, Birdsall, and Fox (1954) to provide an index of accuracy consistent with their basic model of the detection process. They saw the detection task as one of discriminating occurrences of "signal plus noise" *(sn)* from occurrences of "noise alone" *(ri).* Given that noise is a random variable, the two alternatives can be considered as statistical hypotheses. The theory of statistical decision, or of testing statistical hypotheses (e.g., Wald, 1950), is the basis for a model that provides an accuracy index that is independent both of the probability of occurrence of the two alternatives *(s* and 1 — *s)* and of the discriminator's tendency to favor the choice of one or the other alternative. Neither variable, the detection theorists suggested, is usefully or properly regarded as part of the process of discrimination per se and neither should therefore influence an index of discrimination capacity or accuracy" (Swets 1986, p.104).

The overriding concept in applying ROC analysis to a RM system is that the RM elicits a signal and presents a binary decision problem. Elements of the decision problem in DM systems that use RM include: (1) identification of two possible states for the patient, i.e. stable or at-risk for decompensation; (2) information with which to make a decision (reading value and risk level based on a set threshold) and; (3) a decision.

Identification of the two states begins with an instance (I). The 'I' is the patient input to the decision support system i.e. BP, SpO_2, P. An -- alert or no-alert-- is generated within the DSS based on the 'I', the clinical parameters and the thresholds, as set by the patient's physician or nurse. A classification model maps an 'I' to a predicted class (Fawcett 2004, p.2). This has four possible outcomes (TP, TN, FP, FN) and Fawcett notes that, "Given a classifier and a set of instances a 2x2 confusion matrix can be constructed..." (Fawcett 2003, p.3). As noted in Chapter 1, ROC graphs visually represent a 2 X 2 classification model i.e. a confusion matrix to examine the performance of classifiers (Swets 1988).

ROC is a technique to visually represent, organize and select classifiers based on their performance (Fawcett 2004; Swets 1973; Swets 1986). ROC graphs provide another method in addition to confusion matrices to examine the performance of classifiers (Swets 1988).

Swets presents a clear, easily understood description of ROC,

> "'In signal detection theory, a receiver operating characteristic (ROC), or simply ROC curve, is a graphical plot which illustrates the performance of a binary classifier system as its discrimination threshold is varied. It is created by plotting the fraction of true positives out of the positives (TPR = true positive rate) vs. the fraction of false positives out of the negatives (FPR = false positive rate), at various threshold settings. TPR is also known as sensitivity ..., and FPR is one minus the specificity or true negative rate. In general, if both of the probability distributions for detection and false alarm are known, the ROC curve can be generated by plotting the cumulative distribution function (area under the probability distribution... *AUC*) of the detection probability in the y-axis versus the cumulative distribution function of the false alarm probability in x-axis.'

> ‘ROC analysis provides tools to select possibly optimal models and to discard suboptimal ones independently from (and prior to specifying) the cost context or the class distribution. ROC analysis is related in a direct and natural way to cost/benefit analysis of diagnostic decision making. ROC is also known as a relative operating characteristic curve, because it is a comparison of two operating characteristics (TPR and FPR) as the criterion changes.’” (Swets cited in Wikipedia 2013).

As noted by Fawcett,

> “The ROC graphs are two-dimensional graphs in which the TP ratio is plotted on the Y-axis and the FP ratio is plotted on the x-axis. The ROC graph depicts relative trade-offs between the benefits (TPs) and costs (FPs).” (Fawcett 2003, p.4)

The rules that determine the decision are systematic in the DM model and the relative merit of the decision procedure is evaluated based on the patient outcomes in this research (HH, A&E, IP). The objective is to make as many correct decisions as possible, alerting the clinician to the need to intervene. So the probability of the decision needs be determined.

> “In signal detection theory, the first quantity is termed the *hit rate* and the second, *the false-alarm rate.* (also indicating the frequent asymmetry between alternatives A and B are the corresponding terms "true-positive ratio" and "false-positive ratio". The two quantities in question vary together from low to high as the criterion for choosing alternative A is made more lenient (or the bias toward the choice of A becomes stronger) and, thus, for any particular degree of accuracy, an ROC curve is traced from left to right and low to high...and reflects all possible decision criteria or response biases, and hence is independent of any one.” (Swets 1973, p.100-101).

What is important is that there exist probability densities for each possible ‘I’ given an alert, or a no alert. Therefore ‘I’ is an element in a set, and a probability can be defined for each event. The number of correct and incorrect alerts is the dependent variable in the analysis of the model.

ROC curves are frequently used in clinical informatics to evaluate classification and predictive models that use a DSS and elicit a signal (Lasko 2005). Positive cases are identified as true positive (TP), and false negative (FN). Negative cases are identified as true negative (TN) and false positive (FP).

Lasko et al. (2005) notes,

> "ROC analysis investigates the accuracy of a model's ability to separate positive from negative cases (such as predicting the presence or absence of disease), the results are independent of the prevalence of positive cases in the study population. It is especially useful in evaluating predictive models or other tests that produce output values over a continuing range, since it captures the trade-off between sensitivity and specificity over the range." (Lakso, Bhagwat, Zou & Ohno-Machado 2005, p. 2).

In predicting the clinical decompensation leading to COPD exacerbation, the RM system documents a biometric measure that classifies the biometric parameters as falling above or below a certain threshold. This is interpreted as a probability of a patient having an exacerbation of their COPD. The clinician can adjust the threshold, which will in turn change the FP rate. Increasing the threshold would result in fewer FPs and more FNs. The actual shape of the curve is determined by the overlap of the two distributions. The relationship of the confusion matrix (TP, FN, FP, TN) with the bimodal curve and ROC curve is illustrated in Figure 3, p.40.

In order to describe how ROC analysis may be used to evaluate a RM system, it is first necessary to define the terms, structure and processes used in ROC analysis.

3.4.1 ROC space

Hamilton (2012) explains the ROC space in his course notes and provides a clear explanation of the graphic representation of the classifiers in ROC space,

> "The point (0, 1) is the perfect classifier: it classifies all positive cases and negative cases correctly. It is (0, 1) because the false positive rate is 0 (none), and the true positive rate is 1 (all). Point (1, 0) is the classifier that is incorrect for all classifications. In many cases, a classifier has a parameter that can be adjusted to increase *TP* at the cost of an increased *FP* or decrease *FP* at the cost of a decrease in *TP*. Each parameter setting provides a (*FP*, *TP*) pair and a series of such pairs can be used to plot an ROC curve. A non-parametric classifier is represented by a single ROC point, corresponding to its (*FP*, *TP*) pair." (Swets cited in Hamilton course notes 2012, np).

The classifier is the measurement parameter that when outside a set threshold generates the alert identifying the probability of a COPD exacerbation. However, in order to accomplish this

task, the appropriate classifier at an optimum threshold needs to be identified for the target population, which in this research, consists of persons afflicted with COPD. Research to identify the optimum threshold for RM systems has not been undertaken. This is important, as accuracy in identifying the probability of clinical deterioration is a matter of degree and lies along a continuum. Swets notes,

> "Diagnosis is probabilistic and diagnostic decisions are made with more or less confidence. Hence, making a positive or negative decision in a systematic way requires selecting a threshold along the scale of evidence, such that values above the threshold uniformly lead to a positive decision and values below it lead to a negative decision. In principle, one can set such a positivity criterion anywhere along the scale and so can aspire to choose the particular criterion that is best for a given purpose and best for the specific situation at hand." (Swets 1992, p.522).

ROC graphs are another way besides a 2x2 classification table or confusion matrix to examine the performance of classifiers (Swets 1988).

> "ROC graphs are bi-dimensional representations of the sensitivity [called the true positive rate (TPR) on the Y axis] and 1-specificity [called the false positive rate (FPR) on the X axis], corresponding to each possible cut-off point (classifying value). In other words, they represent the tradeoffs between benefits (TPs) and costs (FPs)" (Pintea & Moldovan 2009, p.53).

ROC analysis involves first determining the sensitivity and specificity of every individual in the sample group (i.e. both subjects with and without alerts), and then plotting sensitivity vs. 1-specificity across the full range of threshold values. The ROC curve or point is independent of class distribution or error costs (Kohavi & Provost 1998). A unique feature of ROC curves is that the results are independent of assumptions regarding distribution, so are "distribution-free. The ROC curve includes all information contained in the confusion matrix, since a false negative (FN) is the complement of a true positive (TP) and true negative (TN) is the complement of the false positive (FP) (Swets 1988).

Green and Swets (1966) note that an important aspect of the application of decision theory experiments and relationships with properties of the ROC curve and structure is that it represents the underlying decision process. "The slope of the ROC curve at any point is equal to the likelihood ratio criterion that generates the point." (Green & Swets 1966, p.36).

It is further noted that, "A ROC curve based on likelihood ratio criteria must have a hit probability (TP) that is a monotonically increasing function of the false alarm probability and a slope that is monotonically decreasing" (Green & Swets 1966, p.38).

Evaluation includes determining an alert level at which the DM model will best detect the potential for COPD exacerbation, and the corresponding threshold setting from the ROC curve that will provide a suitable number for detecting the 'I' correctly. The area under the curve (AUC) provides the actual measure of performance. The AUC varies from .50, representing a range where the hit probability equals the false alarm probability, to 1.0 where no errors occur.

Figure 3.1 visually represents the relationship between the bimodal curve (3.1a), the 2 X 2 classification model (3.1b), and the ROC curve (3.1c). Figure 3.1a depicts the classification of a two-valued variable having a normal distribution about each value with the shaded areas identifying the classification categories (TP, TN, FP, FN). Overlap of the curves results in FP and FN results. Figure 3.1b presents the areas of the bi-modal curve as a 2 X 2 confusion matrix. Note that the sum of the TP and FN, and of the FP and TN will equal 1 in each case representing the areas under their respective curves. The ROC curve in Figure 3.1c, represents the performance of the threshold as it is varied over the range of thresholds. The ROC is derived by calculating the confusion matrix at a succession of threshold values and plotting the outcomes with respect to FP against TP. The dot on the curve represents the optimum threshold value.

Call to mind that the aim of ROC analysis is to detect the presence of a particular signal, missing as few positive occurrences as possible while also raising as few false alarms as possible (Krzanowski & Hand 2009).

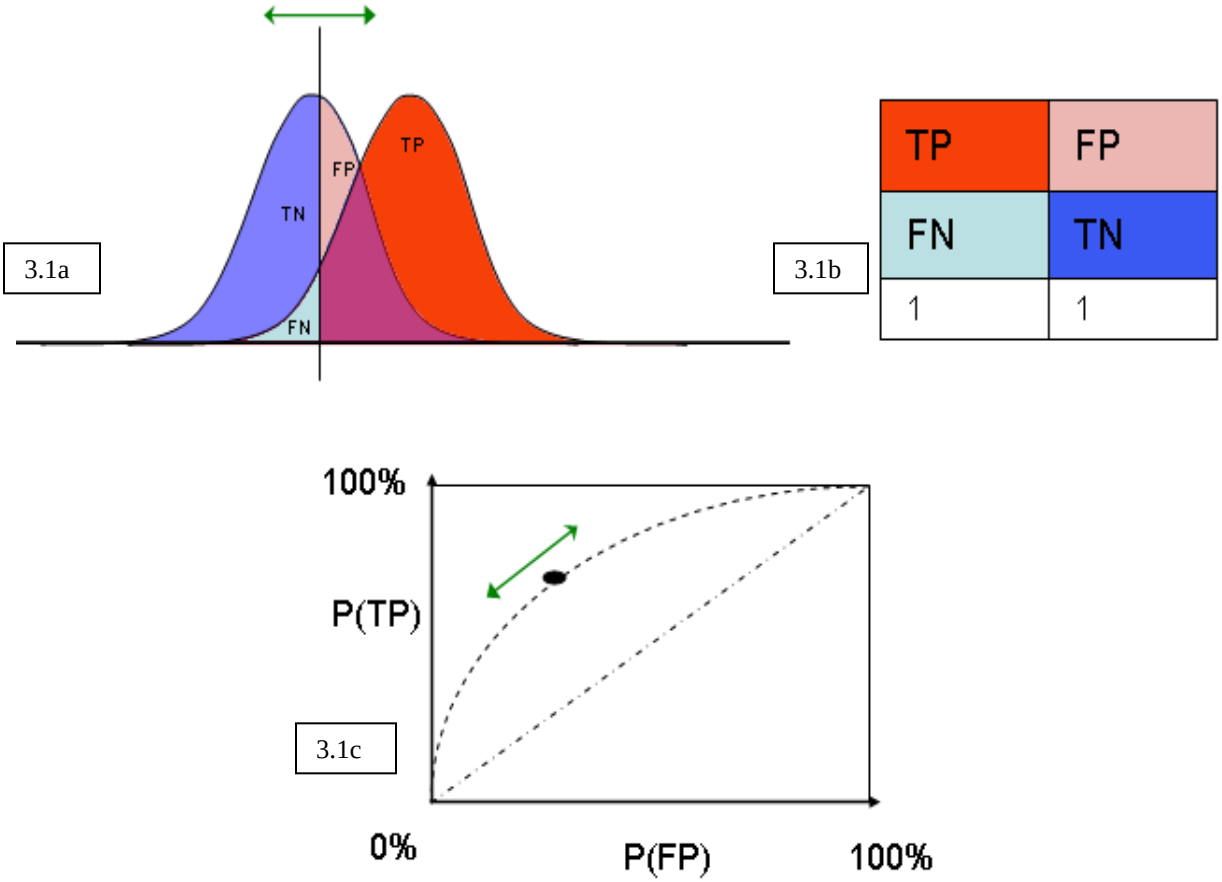


Figure 3. Bimodal curve (3.1a), confusion matrix (3.1b), ROC space (3.1c)
(Wikipedia, Receiver Operating Characteristic, accessed June 7, 2013)

Figure 3.2 illustrates ROC space and depicts three curves with varying predictive capacity; the higher the curve the better the classifier performance in predicting whether a patient has a disease or not. In the case of this research, it would identify the potential for COPD exacerbation.

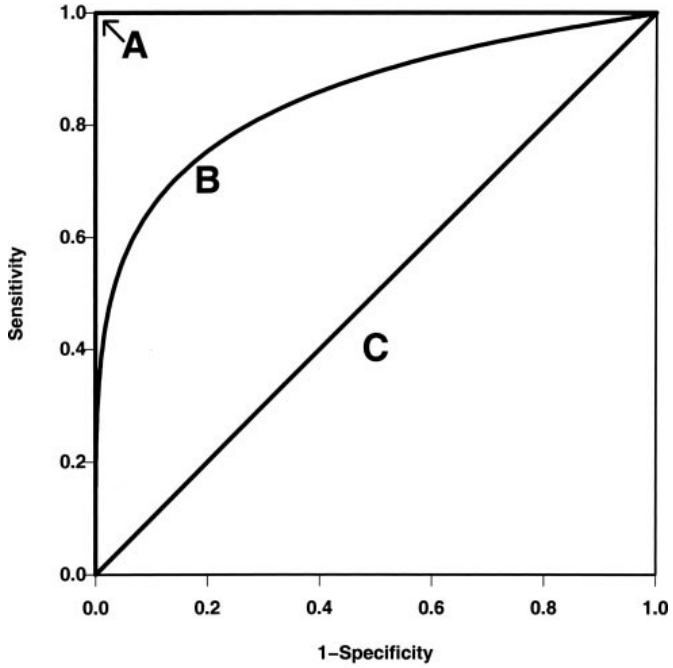


Figure3.2. Hypothetical ROC curves

The three hypothetical ROC curves represent the diagnostic accuracy of the gold standard (lines A; AUC = 1) on the upper and left axes in the unit square, a typical ROC curve (curve B; AUC = 0.85), and a diagonal line corresponding to random chance (line C; AUC = 0.5). As diagnostic test accuracy improves, the ROC curve moves toward A, and the AUC approaches 1 (Zou 2009, p. 656).

> "The ROC analysis can be used in determining the optimal cut-off point of a test. As mentioned above, the optimal cut-off point is the most north-western point in the ROC space. It is the cut-off point where the proportion of subjects that were accurately classified is maximal (cut-off point which has a high sensitivity and also a high specificity). In other words, as a rule, the optimal cut-off point is the one which maximizes TP+TN (or minimizes the FP+FN). However, this principle is based on the assumption that the cost of making a false positive mistake is equal to the cost of making a false negative mistake. In real life, these costs are rarely equivalent." (Pintea & Moldovan 2009, p. 54).

This research evaluates the performance of the classifiers commonly utilized in RM systems for a COPD population by using patient outcomes to determine performance. Patient outcomes are defined as HH, A&E or IP hospitalisation. The classifiers for this research are the biometric readings for BP, SpO_2 and P. As noted above, the ROC curve provides a graphic

illustration in which to examine the trade-off between the ability of a classifier to correctly identify positive cases and the number of negative cases incorrectly classified (Lakso, Bhagwat, Zou & Ohno-Machado 2005). Current thresholds for classifiers in RM systems are set at varying thresholds without an evaluation of the optimum setting.

As Basilakis et al. (2010) note, this type of analysis will enable an informed development of the DM system and research is critical as telehealth systems become more widely adopted and have a need to screen large volumes of electronically monitored patient data effectively and efficiently (Basilakis, Lovell, Redmond & Celler 2010).

3.4.2 ROC attributes

ROC analysis is a statistical technique that as noted, is widespread in medical testing and many areas of science. ROC analysis continues to evolve and be applied in new and different areas. Applying ROC to analyse DM models that use RM devices is such an extension. It is particularly helpful when evaluating a binary decision problem in a probability model. This research seeks to determine if the patient will have a COPD exacerbation and its probability.

The RM device elicits a signal and a threshold is applied within the DSS and a risk level for the classifier is determined. The clinician will make a secondary decision based on their experience and knowledge. The question is whether the signal is predictive for COPD exacerbation and to what degree. The predictive value of the classifier, signal and threshold is needed. ROC analysis is being investigated as a method to evaluate the performance of the model.

Basing the performance on patient outcomes specific to the disease is viewed as a necessary next step widespread use of these systems is to be encouraged. (Ackerman, Filart, Burgess, Lee & Poropatich 2010)

However, it must be determined if the right classifiers are being monitored in the system. ROC is especially useful in this context as it is distribution-free and skew-agnostic, and bearing in mind, that RM is evaluating populations where skew in the data does exist. It is inevitable that there will be more TNs in a stable but chronically ill population. This is the case for all chronic illnesses and makes ROC a valuable tool to use to evaluate the in-home RM system.

In summary, the following characteristics of ROC curves provide capability to assess RM systems.

- The ROC curve graphically captures the performance for the full range of threshold values in an easily viewed format
- The ROC graph visually represents the performance of classifiers
- The ROC graph includes and graphically represents all information contained in the confusion matrix
- Classifiers can be linked to the patient outcome to identify the TPR and FPR
- ROC provides a method that can be used to compare performance between classifiers
- The ROC curve or point is independent of class distribution or error costs (Provost et al., 1998). This is important since the distribution of outcomes for the COPD RM programme evidences "skew" with greater numbers of TNs
- Prevalence of the outcome is not a limiting factor

Chapter 4 will describe the methods used in this research that include the population criteria, data sources, processes and logic used to create the classification categories and the ROC curves.

CHAPTER 4

METHODS

"Concern for man and his fate

must always form the chief interest of all technical endeavours.

Never forget this in the midst of your diagrams and equations."

Albert Einstein, 1879-1955

4.0 METHODS

This research utilizes both parametric and non-parametric methods. All measures are identified and formulae provided in relevant sections. ROC analysis is applied to the performance of the DM model. Performance is measured for accuracy, sensitivity, specificity and probability of predicting clinical deterioration in a COPD population based on patient outcomes.

4.1 Research process

The framework for the application of ROC analysis follows the methods described in "Signal Detection Theory" (Green & Swets 1966), as noted in Chapter 3, and the model evaluation processes outlined by Zou and colleagues (Zou, O'Malley, & Mauri 2007). The approach and assumptions applied in this research were developed by the researcher. Databases were created in Microsoft Excel® in order to analyse the results of the signals elicited by the biometric data for the classifiers (SpO_2, BP, PR) and to link these to the resultant patient outcomes. Outcomes are identified as HH, A&E visits and IP hospitalisations.

4.1.1 Steps in the research process

The research process included:

1. Identification of a cohort of interest.
    a. Development of population inclusion criteria
    b. Development of population exclusion criteria
2. Development of a stratification process based on the documented resource usage.
3. Creation of a confusion matrix and measurement process based on patient outcome events.
4. Creation of ROC curves for each classifier.
5. Identification of the most effective classifier(s).
6. Defining the assumptions for cost/valuation based on the resource usage.

7. Developing a cost avoidance credit to be applied in the cost/valuation for the RM system.
8. Application of optimization criteria to the final ROC curve for the classifier(s) of interest.
9. Establishing an optimum threshold for each classifier shown to have predictive value.
10. Calculating the predictive capacity of the classifiers using ROC analysis.

4.1.2 Challenges

Evaluation was based on available data from the RM system. In order to minimize the intrusion in work flow for the PCT staff and to maintain the confidentiality of the patient's information, access to the patient specific data was unavailable. Due to this data limitation, the patient event-based outcomes of HH, A&E and IP have been used to evaluate the predictive capacity of the classifier. HH visits are documented in the biometric file, and the A&E and the IP data were provided by the PCT.

The limited access to patient data also impacted upon the stratification of the population. Without access to the medical records and pulmonary function information, it became necessary to seek an alternative method to stratify the population. Hospitalisation was used as a metric and the cohort included patients with at least one IP hospitalisation for COPD within a one year timeframe prior to the monitoring period.

4.2 Ethics Approval

This research was approved by the Brunel University Ethics Committee on December 8, 2010. The NHS Health Research Authority approval was received on April 15, 2013 and the permission to use programme data was received from the NHS Direct on April 16, 2013 (appendices A, B, C).

4.3 Statistical analysis tools

The tools for statistical analysis included: Microsoft Excel® spreadsheets for data input and

tabulation of the confusion matrices and MedCalc® version 12© 1993-2013. The MedCalc® software is used to generate the ROC curves and statistical measures. Statistics include accuracy, sensitivity measured by the area under the curve (AUC), specificity, confidence intervals, standard error and p value. The ROC curves were generated using the data input process stipulated in MedCalc®.

4.4 Data source and process

Data were acquired from the NHS Direct COPD monitoring programme evaluation conducted by Brunel University and the Chorleywood Health Centre. The NHSD report was finalized in August 2012. The de-identified secondary data were abstracted and analysed for the time period November 2010 through February 2, 2012. Access to all data was via a secure Brunel ftp site in the Department of Information Systems and Computing. A data use agreement (DUA) is in place between the researcher, Brunel University and the NHSD.

The MedCalc® statistical process assumes a negative event at a higher threshold. Values have been subtracted from 100 to account for this since a negative event occurs for a lower threshold for SpO_2 in the COPD population.

4.4.1 Data sets and description

The patient population included in this study consisted of patients with a confirmed diagnosis of COPD verified by the PCT and diagnostic codes. Patients were sorted by the number of days monitored and the number of hospitalisations occurring within a one year timeframe prior to the start date for telemonitoring. The data were limited to one PCT site.

The number of days monitored was determined by the start and end date in the readings file. Results are presented in Chapter 5. The start date is the first day that the RM device reported biometric data and the final date is the last reported monitoring date or February 2, 2011. Visits to

the home by Care Matrons were recorded in the readings data set and documents the date and time when home visit occurred. A master file was created in MS Excel® that contains all patients monitored for at least 183 days, exclusions, the population included in the research, resource usage, and readings for all biometric data, as well as the normalised resource usage data and process.

All of the participants were identified by an ID number that is created in the DM system. A simple numeric was substituted in the tables for each patient included in the research cohort in order to protect patient privacy.

Biometric readings for each patient for the classifiers (BP, SpO_2, PR) have been extracted from the DM system summary data sets and all readings contain a date and time stamp.

Identification of the use of healthcare services was extracted from the PCT data. Resource usage codes were identified using published code sets from the NSRC01 2011-12 (Department of Health 2011).

Data used to calculate the confusion matrices and formulae for event-based outcomes (TP, TN, FP, FN), are included in the MS Excel® spreadsheets.

ROC graphs and statistics were completed in MedCalc® for all classifiers and followed the processes and statistical methodology in the application.

4.4.2 Data inclusion criteria

1. All data from the PCT that meet the inclusion criteria.
2. All data in the readings file with a value and risk designation.
3. Resource usage for all A&E and IP events.
4. Resource usage for COPD respiratory codes.
5. All classifiers with documented measurement values i.e. SpO_2, BP, PR.

6. Care Matron visit.
7. First patient reading of the day, on days having a Care Matron visit.

4.4.3 Data exclusion criteria

1. Resource usage for non-COPD respiratory services is excluded. For example, accidents not related to COPD, cardiac services, scheduled services and surgeries, elective admissions, ophthalmology, psychiatry, urology and other gerontology services.
2. Data in the readings file by day/line that fail to provide an alarm rating, i.e. "0".
3. Multiple readings by the Care Manager during a HH visit.
4. IP data that occur outside of monitoring dates or more than five days from the last monitored date.

4.5 Population

The total population includes 34 patients. Patients were identified by the system ID, the start and end dates and the total number of days monitored. Each classifier includes more than 6,000 lines of data in the statistical tabulation. One patient with only SpO_2 data was included in the SpO_2 analysis. All other data sets for classifiers contain 33 patients.

4.5.1 Population inclusion criteria

1. Primary diagnosis of COPD was confirmed by the PCT and evidence of thoracic and respiratory codes are documented in the resource usage file.
2. Individuals who were monitored for a minimum of six (6) months (183 days) or more.
3. Patients with at least one IP stay for COPD within one year prior to the monitoring period.
4. Full data sets for the classifier.

4.5.2 Population exclusion criteria

1. The primary diagnosis is identified as other than COPD. Patients who have no documented COPD services within the monitoring timeframe and who are identified as having a primary medical issue that is not respiratory. As an example, cardiac or cancer services.
2. Individuals who are monitored for less than six months (183 days).
3. There are no IP COPD admissions or services identifiable as COPD within a one year timeframe prior to monitoring.
4. Missing data sets for classifiers.

4.6 Framework and processes

The underlying approach applied in this research follows the model evaluation methods for ROC analysis as described earlier. ROC analysis is applied to DM models that use RM in a similar manner and represents an extension of the use of ROC into disease management, another field of healthcare. The following sections provide definitions and descriptions of the processes that have been adapted for this research.

4.6.1 Confusion matrix

By definition, a classification model is the mapping of an alert to a predicted class or outcome (Fawcett 2005). This provides an estimate of the probability of the alerts predicting clinical deterioration (Fawcett 2005). There are four possible outcomes, TP, TN, FP and FN.

> **"**The confusion matrix ...is a table layout that allows visualization of the performance of an algorithm. Each column of the matrix represents the instances "I" in a predicted class, while each row represents the instances in an actual class. The name stems from the fact that it makes it easy to see if the system is confusing two classes." (Wikipedia, Confusion matrix 2013).

A simple statistical method for assessing diagnostic accuracy is through the use of the confusion matrix. The table has two rows and two columns identifying the number of TP, TN, FP

and FN instances. The confusion matrix (Table 4.1) contains information about the actual and the predicted classifications identified by a classification system (Kohavi & Provost 1998).

		actual value		
		p	***n***	**total**
predicted outcome	***p'***	TP=50	FP=100	P'
	n'	FN=50	TN=800	N'
total		P	N	

Table 4.1**.** Confusion matrix

The layout of the confusion matrix is depicted in Table 4.1 and an example is given that illustrates each section of the confusion matrix. For example, consider the DM model that predicts 1,000 patient alerts. Each alert requires a decision based on whether the alert requires an action or not. The model correctly predicts 50 true actionable alerts (TP) but incorrectly predicts 100 alerts that are non-actionable (FP). It incorrectly predicts 50 cases that are not alerts but are actionable (FN) and 800 cases that are not alerts and are not actionable (TN). The identification of positives and negatives is attributed based on the patient outcomes in this research. ROC space represents the trade-offs between the benefits (TPs) and the costs (FPs) (Pintea & Moldovon 2009, p.53), i.e. between the sensitivity and the specificity. Cost methodology applied in this research is described in section 4.9. The ROC graph will visually illustrate the performance of the classifiers in the DM model using the classification categories from the confusion matrix.

An explanation of the relationship of the sensitivity and specificity of the four classifications to threshold and accuracy is presented below and illustrated by the bi-modal curve in Figure 4.1.

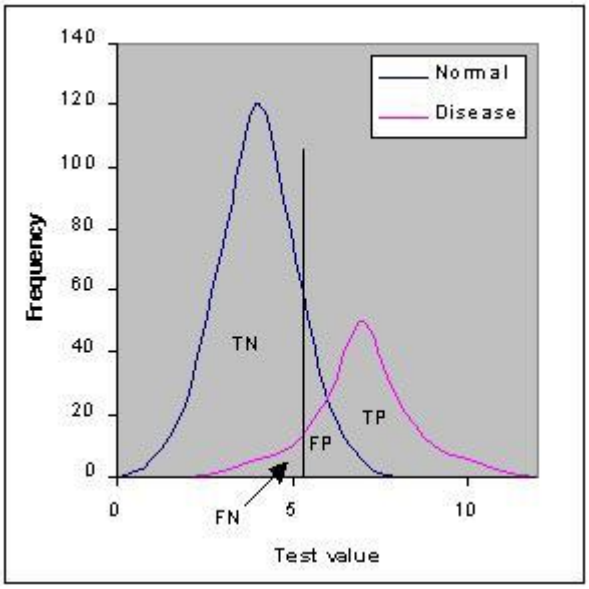


Figure 4.1. Bi-modal curve

> "The sensitivity and specificity of a diagnostic test depends on more than just the "quality" of the test--they also depend on the definition of what constitutes an abnormal test. Look at the idealized graph at right showing the number of patients with and without a disease arranged according to the value of a diagnostic test. This distributions overlap--the test (like most) does not distinguish normal from disease with 100% accuracy. The area of overlap indicates where the test cannot distinguish normal from disease. In practice, we choose a cutpoint (indicated by the vertical black line) above which we consider the test to be abnormal and below which we consider the test to be normal. The position of the cutpoint will determine the number of true positive, true negatives, false positives and false negatives. We may wish to use different cutpoints for different clinical situations if we wish to minimize one of the erroneous types of test results." (Tape 2013, n.p.)

4.6.2 Measures and formulae

Measures and formulae used to tabulate the confusion matrix and ROC curves are listed below with definitions. The ROC curves are created for each of the classifiers based on the patient outcomes, resource usage and cost optimisation methodology as defined in section 4.9. The

classifications for the category designation of TP, TN, FP, and FN that are used for the confusion matrix are the same as those used for the ROC curve.

1. The accuracy (AC) is the proportion of the total number of predictions that were correct. It is determined using the equation: AC = (TP + TN) / (TN + FN + FP + TP) (Fawcett 2006)
2. Sensitivity is the probability that a test result will be positive when the disease is present is defined as: Sensitivity = (TP) / (TP+FN). It is same as true positive rate which is expressed as a percentage. (MedCalc® Version 12.7.0 1993-2013).
3. Specificity, is the probability that a test result will be negative when the disease is not present i.e. the proportion of negatives cases that were classified correctly is defined as Specificity = (TN) / (TN + FP) = 1 - FPR. (It is same as true negative rate, which is expressed as a percentage) (Fawcett 2006; MedCalc® Version 12.7.0 1993-2013).
4. Precision (P) or true positive rate (TPR) is the proportion of positive cases that were correctly identified, as calculated using the equation: TPR = (TP) / (FN + TP) (Fawcett 2006).
5. The false positive rate (FPR) is the proportion of negatives cases that were incorrectly classified as positive, as calculated using the equation: FPR = (FP) / (FP + TN). This is a Type 1 error equivalent to a false alarm (Fawcett 2006).
6. The false negative rate (FNR) is the proportion of positives cases incorrectly classified as negative, and as calculated using the equation: FNR = (FN) / (FP + FN). This is a Type 2 error equivalent to a miss (Fawcett 2006).
7. Negative predictive value (NPV) is the probability that the disease is not present when the test is negative. NPV = (TN) / (FN+TN) (MedCalc® Version 12.7.0 1993-2013).

8. Positive predictive value (PPV) is equivalent with precision and is the probability that the disease is present when the test is positive. PPV = (TP) / (TP+FP)
   (Fawcett 2006; MedCalc®Version 12.7.0 1993-2013).
9. Positive likelihood ratio (PLR) is the ratio between the probability of a positive test result given the presence of the disease and the probability of a positive test result given the absence of the disease. PLR = (TPR) / (FPR) = Sensitivity / (1-Specificity)
   (MedCalc®Version 12.7.0 1993-2013).
10. Negative likelihood ratio (NLR) is the ratio between the probability of a negative test result given the presence of the disease and the probability of a negative test result given the absence of the disease, NLR = (FNR) / (TNR) = (1-Sensitivity) / Specificity (MedCalc® Version 12.7.0 1993-2013).
11. P value measures the strength of evidence in support of the null hypothesis.
12. Optimum threshold for an empirical ROC is the point at which a line with the above slope first intersects the ROC curve (Zweig & Campbell 1993). It is tangent to the ROC curve. Criterion values for the parameters of PLR, NLR and P value were calculated in MedCalc® and are provided in Appendices E, F, G, and H for all classifiers.

4.6.3 Accuracy

The DM model issues alerts that are based on a set threshold applied to a signal intended to predict a physiological event. The variance between the set threshold and the value indicates the potential for decompensation. The system is identifying the potential or probability for a negative clinical event and the alert may be thought of as predictive for COPD exacerbation. Although the patient's COPD diagnosis is known, the likelihood for clinical deterioration at the point in time that the signal is created, is not known.

> "Diagnostic accuracy studies are used to obtain how well a test, or a series of tests, is able to correctly identify diseased patients or, more generally, patients with the target condition, the condition of interest." (Bossuyt 2008, p. 2)

Generally, diagnostic test accuracy is expressed as sensitivity and specificity. However, accuracy statistics describe the performance under specific conditions, the sensitivity and specificity may change with the population being measured, the setting, and other extraneous or environmental variables. In predictive accuracy studies, the test is used to identify patients that benefit from treatment, and those that do not (Bossuyt & Leeflang 2008).

The accuracy of the model, indicated by the sensitivity and specificity is evaluated using established and validated mathematical formulae as noted in the previous section.

4.6.4 Classifiers

There are four (4) classifiers in the readings data set: (1) systolic blood pressure (SBP), (2) diastolic blood pressure (DBP), (3) pulse rate (PR) and, (4) saturated peripheral oxygen (SpO_2). Three of these, SBP, DBP and PR, have both high and low threshold settings in the DSS and the level of risk is determined using the deviation from these threshold settings.

4.6.5 Risk designation

The DSS does not present a simple binary decision, rather it assigns a number denoting risk levels based on the deviation from the threshold setting. Typically three risk levels are defined in RM systems and are used in this study. Risk level 1 is categorized as a non-alarm. There is no indication of COPD exacerbation based on the thresholds set in the DSS. A SpO_2 risk level of 2 or 3 is an alarm resulting from a reading that exceeds the threshold settings for each risk level. This indicates that some intervention may be appropriate to correct a deteriorating health condition.

The occurrence of an alarm triggers a decision process to provide a HH visit, send the patient to the A&E or for IP hospitalisation. These same actions may be expected for any threshold

that generates an alarm. The higher risk level may indicate prioritisation for intervention, but are treated equally in this research.

4.6.6 Threshold

The threshold value used in the RM system is an important factor to consider as it will influence the decision to provide healthcare resources and there is an associated cost to mobilizing limited healthcare resources based on alarms that have no relationship to the patient outcome. The accuracy of the threshold can also impact upon the healthcare system and attribute a high cost if the DM system fails to generate an alarm for a patient who has a deteriorating health status and is clinically decompensating; in this situation, a high level of impact for the patient and a high cost to the healthcare system for IP hospitalisation.

The ideal threshold for any individual patient may vary by disease progression and will ultimately be a clinical decision. However, this should be based on outcomes to focus resources where best needed.

In addition, patients may become anxious or experience “alarm fatigue” if the system is generating alarms that are set too high or too low for non-performing classifiers, i.e. measures that do not indicate a relationship to a deteriorating health status. In the case of non-performing classifiers, the patient, nurses and other healthcare workers may not pay attention to a classifier that can indicate deterioration if alarms are not predictive. The optimal threshold settings for classifiers can be determined in ROC space.

4.6.7 Outcomes

Patient outcomes, as defined for this research, are categorized as event-based clinical outcomes (HH, A&E, IP). The decision process applied depends on the Care Matron’s clinical judgment that the patient’s health status was deteriorating and whether a home visit is necessary or

that a visit to the A&E and/or an IP hospitalisation would be initiated for a person who was experiencing more severe respiratory difficulties.

4.6.7.1 Home health designation

The HH visit was attributed based on the documentation from the readings in the Master file. Tables 4.2 and 4.3, illustrate the layout of the data fields. The column labelled "Collection Type" when designated as "Care Manager Reported" indicates a HH visit. The protocol for HH visits is to document the biometric parameter while visiting the home so that the reading is entered into the system. One limitation was that there is potentially some data loss if the Care Matron did not follow the protocol. The assumption that is applied to the HH visit is that the secondary decision to visit the home is in response to the clinical decision being made by the Care Matron that the patient is deteriorating.

4.6.7.2 Resource usage identification process

The A&E visit and the IP admissions were confirmed by identifying the COPD code from the cost data sets, i.e. reason for admission. The cost data sets were provided by the PCT.

4.6.7.3 Code identification process

The service code identification process used the PCT resource utilisation data and paired the column labelled "Specialty_ M", (designates services by specialty area, i.e. thoracic, cardiology, endocrine etc.), with the column labelled "Agreement_Line_Number", (contains COPD DZ series codes). This identified the resources used and the COPD services.

4.6.8 Category designation process

Category designation occurs following a series of events. First, the signal that is generated is based on a threshold set into the DSS for each biometric classifier. The value is interpreted within the DSS and identified as a risk level in the RM system. The risk levels are identified as a 1

(no risk), 2 (medium risk) or 3 (high risk). Once a risk level is established the biometric classifier is designated into one of four categories (TP, TN, FP, FN) based on event-based patient outcomes (HH, A&E, IP).

Categories are listed below. True positive category designation also takes into account a range of three days for IP hospitalisation and a range of one day for a HH visit. This is because respiratory exacerbations are an infective process that takes three to five days to develop and a positive signal followed by an IP hospitalisation would be indicative of that timeframe. Further research with access to the medical record would clarify the best timeframe to use, i.e. three or five days. Likewise, a HH visit within one calendar day of an alert is also identified as a true positive event. This takes into account the data submission process, as patients may send biometric readings at any time of the day and the HH visit may not be possible on the same day, therefore. one day is allowed for the nursing response time to account for the work flow process for the Care Matrons.

Categories in this research are defined as follows:

1. TP is an alarm with a HH, A&E or IP hospitalisation.
2. FN has no alarm with a HH, A&E visit, or IP hospitalisation.
3. TN has no alarm and no HH, A&E visit, or IP hospitalisation.
4. FP has an alarm with no HH, A&E visit, or IP hospitalisation.

4.6.9 Signalling data spreadsheets

Spreadsheets were created in MS Excel® linking the signal and risk level for all classifiers with the patient outcomes. A confusion matrix was created from these spreadsheets and data from the confusion matrix was entered into MedCalc® to create the ROC curves for the classifiers. Formulae were applied to the MS Excel® spreadsheet to identify IP hospitalisations within three

days and HH visits within one day of a signal. All data were reviewed on a line by line basis to check formulae in the spreadsheet. The master file serves as the source document for all classifier readings and data. Table 4.2 and 4.3, identify the relevant patient data fields in the spreadsheet. This format is consistent throughout the data sets for all classifiers.

C	D	F	G	H	J	K	L	M	N
Type	**Value**	**Reading Date/ time**	**Collection Type**	**Risk Level**	**NO Signal**	**YES Signal**	**HHC**	**A&E**	**IP**
Diastolic BP	85	2/6/2012 12:33	Device-Reported	**3**		1			
Pulse Rate	75	2/6/2012 12:33	Device-Reported	1	1				
Systolic BP	166	2/6/2012 12:33	Device-Reported	**3**		1			
Diastolic BP	80	2/6/2012 11:30	Care Manager Reported	**3**		1	1		
Systolic BP	175	2/6/2012 11:30	Care Manager Reported	**3**		1			
Oxygen Sat.	94	2/6/2012 11:21	Self-Reported	1	1				

*HHC - home health visit
*A&E - accident and emergency
*IP - in-patient hospitalization

Table 4.2. Sample spreadsheet with data field labels

H	J	K	L	M	N	T	U	V	W
Risk Level	**NO Signal**	**YES Signal**	**HHC**	**A&E**	**IP**	**NS_NA =TN**	**NS_A = FN**	**S_NA = FP**	**S_A = TP**
2		1						1	
1	1					1			
1					1		1		
2		1						1	
2		1						1	
3		1	1						1
3		1			1				1

*NS – no signal
*NA – no action
*A – action
*S – signal
Table 4.3. Sample spreadsheet with data field labels-continued

4.7 Gold standard

In estimating the accuracy of the classification, the disease status of each patient is measured. This is called the gold standard and when evaluating a test result, it would be verified for the sake of accuracy (Zou, O'Malley & Mauri 2007). In this research, all patients have a confirmed diagnosis of COPD verified by the PCT. Confirmation of the COPD exacerbation is confirmed by patient outcomes for resource usage data. The codes are used to confirm the A&E and IP hospitalisations for respiratory care episodes. The HH visits are documented in the readings files and are shown in Table 4.3, column L.

4.8 Cost optimization

ROC analysis requires a cost optimization method to identify the value or estimated costs of a TP, TN, FP, and FN. There are different optimization criteria that could be used.

The resource usage in this study is quantified by cost. There are different resource utilisation costs applicable to each decision and costs are not equal. In this analysis a high cost is applied to a FN because missing the COPD exacerbation will eventually result in a negative event

for the patient and healthcare system (A&E or IP). The costs associated with a FP and TN are moderate to low respectively, as these may or may not result in a HH visit. In the case of the TN, no alert is generated and the cost is low as no untoward event occurs and no services are utilized. A FP has a moderate cost as health resources are utilized with a HH visit but the cost to the patient is low. A TP has a moderate to high cost as health resources are utilized, with the most frequent being a HH visit. However, the avoidance of the use of higher-cost resources for an A&E visit and/or an IP hospitalisation is a frequent outcome. Therefore there is a cost avoidance associated with the TP as cost could have been quite high if resulting in A&E visits and/or IP hospitalisation.

The method used for cost valuation was calculated using direct resource usage cost data incurred during the monitoring period. The average cost of services for a HH visit and an A&E visit uses the UK Reference Costs 2011-2012. The IP cost used in this study reflects the average cost calculated in the NHSD program evaluation (UK reference costs 2011-2012; NHSD Evaluation Report 2012).

A cost avoidance value was developed and applied as the cost valuation method. Cost avoidance was chosen because the use of RM is predicted to prevent the cost of avoidable healthcare resources and result in a savings to the healthcare system reflecting the avoidance of HH, IP and A&E events. Commissioners and vendors project that the savings will offset the cost to support the investment in RM systems over time. Using the cost avoidance method more clearly evaluates the reality of this assumption. A weighted value and the method used to calculate the cost is presented in the next section.

The start-up and on-going technical support costs are not included in calculations as they are not available. However these should be considered for inclusion in future cost models.

4.8.1 Cost valuation methodology

Comparing data from the pre-monitoring period with the data from the monitored period shows a decrease in the use of A&E and IP resources (this is covered in Chapter 5).

This decrease in resource usage has a cost avoidance value. The cost avoidance could be recognized as a cost reduction factor for A&E and IP hospitalisation or attributed to HH. For the purposes of this study, the avoidance value is assigned to HH because it is assumed that the early intervention by the Care Matrons during a HH visit is the reason for the reduction in A&E and IP hospitalisation during the monitoring period. This approach is used because it can be quantified and because the data show a recognizable pattern in the form of HH visits in response to alarms during the monitoring period, exclusive of other documented actions.

The calculation method is as follows: A&E visits decreased from 86 in the pre-monitored period of 12,045 patient-days to 42 during the monitored period of 9,593 patient days. The effective cost avoidance was calculated by weighting the reduction by the patient-days in the two periods. The same technique was used for the decrease in IP costs due to the reduction in bed days.

The reductions were then applied to the HH cost resulting in an effective savings to the overall system of £226 for every HH visit. The HH cost of a negative £-226 was applied, the A&E cost of £119 and the IP cost of £640 were also applied to the monitored data of the SpO_2 classifier in order to develop a confusion matrix for resource usage (as expressed by cost). Cost valuation is needed to identify the optimum threshold and was applied when ROC curves were created for the classifiers. Each cost was assigned to one of the confusion matrix quadrants (TP, FP, TN, FN) previously calculated for SpO_2 readings and associated alarms and was applied to the

cost/value calculations in the final ROC curve. Chapter 5, Table 5.7, presents the weighted values applied when generating the ROC curves.

This study assumes that any IP usage within three days of an SpO_2 reading applies to the confusion matrix. The cost of the IP hospitalisation was divided equally between all readings that occurred within the three days prior to the date for the IP hospitalisation. Therefore, the cost of a hospitalisation was, in some cases, divided between TP (for a SpO_2 reading below the alarm threshold) and FN (for readings above the alarm threshold).

The results of this research are presented in Chapter 5.

CHAPTER 5

RESULTS

"Measure what is measurable, and make measurable what is not so."

Galileo Galilei, 1564-1642

5.0 RESULTS

5.1 Population

The sample size includes (N_T=34) patients listed in Table5.1. The table contains patients identified as meeting all of the eligibility criteria, the start date and end date of monitoring, and the total number of days monitored. A numeric has been substituted for the ID in this study to protect patient confidentiality. All patient diagnoses are confirmed as COPD by the PCT and by the code sets included in the PCT cost data files. The cost data files document at least one (1) IP admission for respiratory illness within a 12 month period prior to the monitoring start date for all patients included in the cohort.

ID	Start Date	End Date	Monitor Days	ID	Start Date	End Date	Monitor Days
1.	10/11/2010	06/02/2012	454	18.	04/05/2011	06/02/2012	279
2.	24/11/2010	06/02/2012	440	19.	10/05/2011	06/02/2012	273
3.	25/11/2010	06/02/2012	439	20.	11/05/2011	06/02/2012	272
4.	25/11/2010	06/02/2012	439	21.	12/05/2011	06/02/2012	271
5.	25/11/2010	06/02/2012	439	22.	24/05/2011	06/02/2012	259
6.	21/02/2011	06/02/2012	351	23.	01/06/2011	06/02/2012	251
7.	21/02/2011	06/02/2012	351	24.	31/03/2011	30/11/2011	245
8. **	24/02/2011	06/02/2012	348	25.	08/06/2011	06/02/2012	244
9.	10/03/2011	02/02/2012	330	26.	10/03/2011	17/10/2011	222
10.	25/03/2011	06/02/2012	319	27.	20/04/2011	15/11/2011	210
11.	29/03/2011	06/02/2012	315	28.	26/07/2011	06/02/2012	196
12.	29/03/2011	06/02/2012	315	29.	27/07/2011	06/02/2012	195
13.	30/03/2011	06/02/2012	314	30.	27/07/2011	06/02/2012	195
14.	30/03/2011	03/02/2012	311	31. *	28/07/2010	06/02/2012	194
15.	19/04/2011	06/02/2012	294	32.	10/05/2011	15/11/2011	190
16.	20/04/2011	06/02/2012	293	33.	20/04/2011	25/10/2011	189
17.	21/04/2011	06/02/2012	292	34.	04/08/2011	02/02/2012	183

* SpO_2 data only
** No alarms for pulse during the monitoring period

Table 5.1. Patient population

5.2 Programme results

The overall impact of the COPD RM programme, for this cohort of patients, evidences a decrease of 50% fewer IP bed days during the monitoring period, representing 437 fewer days as compared to the 12 months prior to monitoring. The number of A&E visits decreased by 49%

representing 44 fewer A&E visits (Table 5.2). There was no baseline available for the number of pre-monitoring HH visits and consequently the change during the RM period could not be evaluated.

ID	Monitored Days	# Home Visits	Pre A&E	# A&E Monitored	Δ A&E	Pre IP Days	# IP Days Monitored	Δ IP
1.	454	88	1	1	0	10	6	-4
2.	194	11	5	3	-2	45	19	-26
3.	440	35	1	0	-1	13	0	-13
4.	439	102	2	1	-1	2	1	-1
5.	439	76	3	1	-2	23	3	-20
6.	439	41	4	1	-3	56	7	-49
7.	351	126	5	0	-5	27	0	-27
8.	351	86	1	1	0	10	3	-7
9.	348	56	1	0	-1	1	0	-1
10.	330	95	3	2	-1	24	14	-10
11.	319	62	0	0	0	12	1	-11
12.	315	81	2	3	+1	26	0	-26
13.	315	78	3	0	-3	19	15	-4
14.	314	50	1	0	-1	4	0	-4
15.	311	115	7	1	-6	61	7	-54
16.	294	47	1	0	-1	59	0	-59
17.	293	12	5	0	-5	11	0	-11
18.	292	58	1	1	0	4	3	-1
19.	279	20	1	1	0	4	6	+2
20.	273	21	3	0	-3	27	0	-27
21.	272	77	3	1	-2	12	4	-8
22.	271	14	1	0	-1	9	0	-9
23.	259	39	2	0	-2	8	0	-8
24.	251	6	3	0	-3	3	0	-3
25.	245	23	1	4	+3	16	80	+64
26.	244	15	1	0	-1	2	0	-2
27.	222	22	2	0	-2	8	0	-8
28.	210	53	11	8	-3	290	154	-136
29.	196	25	5	0	-5	29	0	-29
30.	195	69	1	6	+5	3	45	+42
31.	195	49	1	0	-1	13	2	-11
32.	190	34	2	2	0	16	32	+16
33.	189	23	2	4	+2	12	16	+4
34.	183	25	1	1	0	5	9	+4
TTotal		**1734**	**86**	**42**	**-44**	**864**	**427**	**-437**

Table 5.2 Resource utilisation

5.3 Summary data by classifier

The RM system considers a continuous output classification problem with two classes. The confusion matrices for the classifiers illustrates the relationship of the signal to the patient event-based outcome in order to identify the accuracy into the two classes of interest, i.e. COPD exacerbation or no COPD exacerbation. As discussed in Chapter 4, HH visits within one day of an

alert are designated as TPs and IP hospitalisation within 3 days of an alert are designated as TPs. Data are inclusive for all submitted data that meet the data inclusion criteria as noted in Chapter 4, section 4.4.2 and has been compiled following the processes outlined in Chapter 4, section 4.6. Table 5.3 presents the summary data for the classifiers as a first step in the research process prior to ROC analysis being applied.

It should be noted that data for all classifiers were not submitted by all patients. The summary numbers represent all submitted data that meet the criteria outlined in Chapter 4. For example, the discrepancy between the SBP and DBP numbers is a result of there being separate parameters. If the system does not register one of the parameters, it is not included in the summary data set. Tables 4.2 & 4.3 in Chapter 4 illustrate the process of assigning a category based on a signal and an event-based outcome. Table 5.3 below summarizes the results for each classifier following the processes and formulae presented in Chapter 4.

	NS_NA =TN	**NS_A= FN**	**S_NA= FP**	**S_A= TP**	**Accuracy**	**TPR Sensitivity Precision PPV**	**FPR 1- Specificity**	**TNR Specificity**	**FNR**
SpO_2	4233	763	783	705	**0.77**	0.48	0.16	0.84	**0.52**
SBP	4637	1547	1310	2014	0.66	0.57	0.30	0.70	0.43
DBP	4546	1547	1328	1799	0.69	0.54	0.23	0.77	0.46
PR	4737	721	148	192	0.51	0.21	0.03	0.97	0.79
All	18453	4578	3569	4710	0.72	0.51	0.19	0.81	0.49

NS – no signal, NA –no action, A –action, S – signal
Table 5.3 Classifier summary measures

The statistical results that relate to the performance of the DM system include accuracy, precision, the TPR, representing the sensitivity, the FPR, representing 1-specificity, TNR and FNR. Note that the results presented in Table 5.3 can present a misleading impression of the

relative performance of the tests due to the skew in the data. However, SpO_2 is identified as the most accurate classifier and this will also be illustrated with the ROC curve in the next section.

Accuracy is the proportion of the total number of correct predictions, and precision is the proportion of positive cases correctly predicted. The accuracy equals 0.72 for all of the classifiers combined and 0.77 for SpO_2 alone. The mean difference in accuracy between all classifiers combined and SpO_2 is 0.05. Inclusion of the additional biometric classifiers of BP and PR does not improve the overall accuracy of the model.

The precision of the model for all classifiers combined is 0.51 and for SpO_2 is 0.48. Although SBP is identified as the most precise classifier with a rate of 0.57, the ROC curve in the next section will show that SBP has little or no relationship to patient outcomes and fails as a classifier. Distribution skew in the data accounts for higher sensitivity and lower specificity as seen in Table 5.3 (Krzanowski & Hand 2009).

The diagnostic accuracy is used to identify how well a test, in this case a classifier, is able to identify the target condition (Bossuyt & Leeflang 2008) and is expressed in sensitivity and specificity. Statistical results using the confusion matrix identify the highest sensitivity as SBP at 0.57 and the lowest false positive rate as pulse rate at 0.03 (Table5.4). However, SpO_2 is identified as the most accurate classifier and this will also be illustrated with the ROC curve in the next section.

This difference in outcome for the two approaches, statistical and ROC curve, will illustrate the need to use the appropriate method for the system being evaluated. As the ROC curve is distribution-free and therefore skew-agnostic, the results in the next section will better represent the performance of the classifiers and their predictive capacity.

5.4 ROC analysis of classifier performance

The characteristics of ROC analysis summarised in Chapter 3 are depicted by the ROC curve and the quantitative results are provided. All ROC curves were produced in MedCalc® and the scales (normalised to 100) are those produced by the MedCalc® application. Table 5.4 summarises classifier performance. A summary of all criterion values as calculated in MedCalc® for sensitivity, specificity, likelihood ratio, P value, confidence interval and cost, for all thresholds, are provided in Appendices E, F, G, and H. Characteristics visually represented by the curve include:

1. The trade-off between sensitivity and specificity.
2. The AUC as a measure of the accuracy of the classifier.
3. The slope of the tangent line at a cut-off point is the likelihood ratio (LR) denoting the optimum threshold for the classifier being measured.

Performance of the classifiers is visually represented in Figures5.1-5.4. The ROC curves illustrate the relationship between the classifiers and the prediction of COPD exacerbation based on the patient outcomes. The only classifier that correlates with clinical deterioration, i.e. COPD exacerbation and is predictive is SpO_2. The classifiers for SBP, DBP and PR are only slightly above the diagonal indicating poor performance, i.e. they are only a little better than random chance. It should be noted that each point on the curves corresponds to the true frequency of clinical deterioration as measured by an event-based outcome (HH, A&E or IP). The ROC curve presents all thresholds for the classifiers.

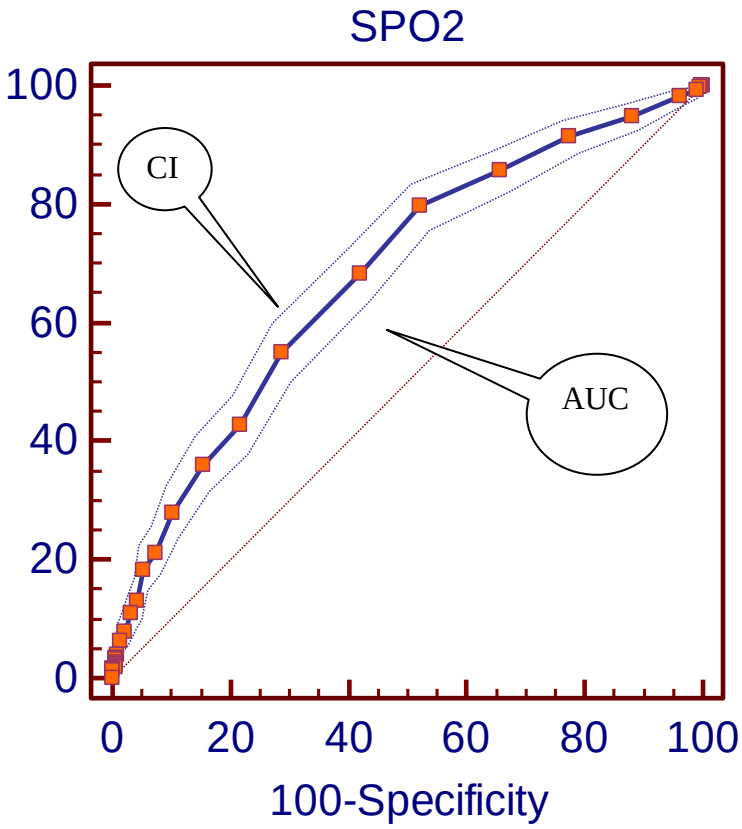


Figure 5.1. ROC curve for SpO_2
*Confidence interval (CI) and the diagonal are denoted by dotted lines and the AUC is labelled for clarity.

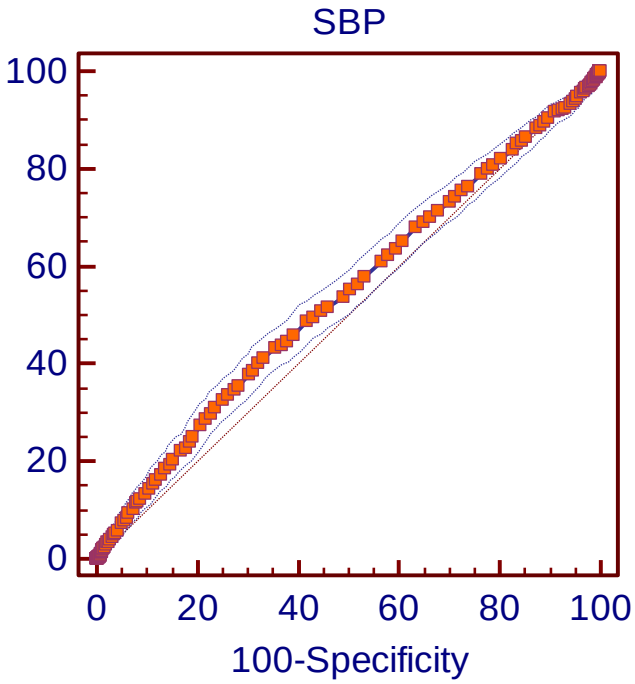


Figure 5.2. ROC curve for SBP

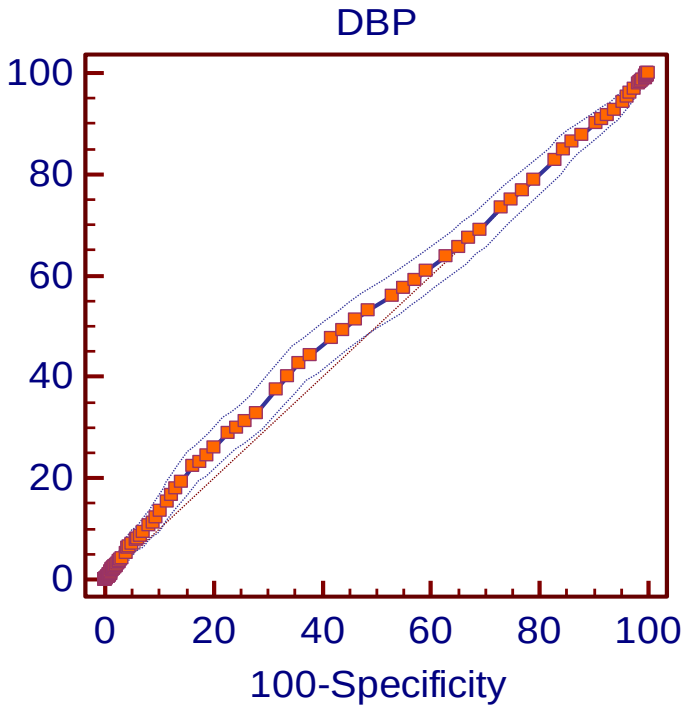


Figure 5.3. ROC curve for DBP

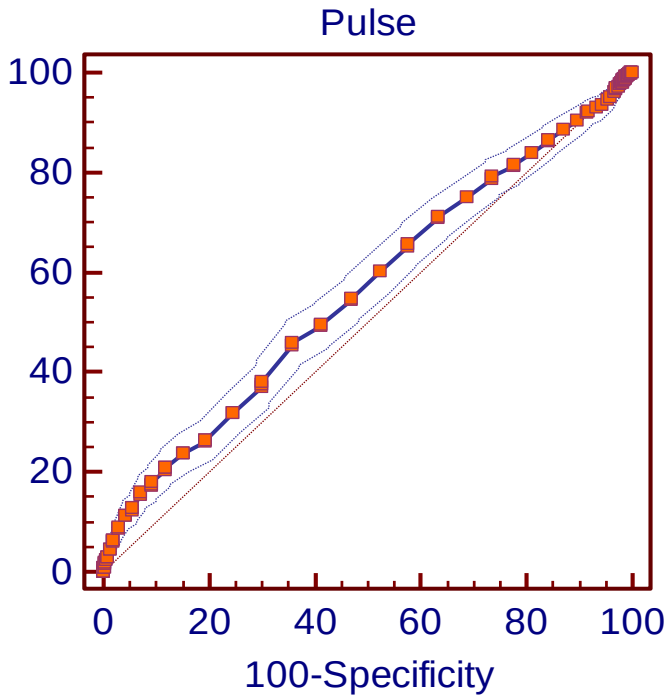


Figure 5.4. ROC curve for pulse rate (high and low)

Variable	SpO_2	SBP	DBP	Pulse Rate
Classification variable	Outcome	Outcome	Outcome	Outcome
Sample size	6533	10039	10034	6427
Positive group :	674	1422	1344	831
Negative group :	5859	8617	8690	5596
Disease prevalence (%)	10.3	14.2	13.4	12.9
Area under the curve (AUC)	0.693	0.540	0.527	0.553
Standard Error[a]	0.0107	0.00849	0.00881	0.0117
95% Confidence interval[b]	0.682 to 0.704	0.530 to 0.550	0.517 to 0.537	0.541 to 0.566
Significance level P (Area=0.5)	<0.0001	<0.0001	0.0020	<0.0001

a DeLong et al., 1988

b Binomial exact

*Taking into account disease prevalence and estimated costs:
cost False Positive: -0.003; cost False Negative: -0.255
cost True Positive: -0.742; cost True Negative: 0

Table 5.4. ROC measures of classifier performance

5.5 The Area under the curve

The accuracy of a test using the ROC curve is measured by the AUC. AUC for this research is the probability that the system correctly identifies the clinical deterioration of a COPD exacerbation. The AUC for each of the classifiers is presented in section 5.5 and quantifies their accuracy. To give some perspective the following analogy and scale is used.

> "A rough guide for classifying the accuracy of a diagnostic test is the traditional academic point system:
>
> - .90-1 = excellent (A)
> - .80-.90 = good (B)
> - .70-.80 = fair (C)
> - .60-.70 = poor (D)
> - .50-.60 = fail (F)" (Tape, accessed July 8, 2013).

The ROC curves illustrate that the classifiers for SBP, DBP and PR are only slightly above the diagonal. BP and PR fail the usefulness test as classifiers. Performance is quantified between .50 and. 60. Comparing the AUC for the classifiers indicates that SBP, DBP and PR with AUCs respectively of 0.540, 0.527 and 0.553 are deemed poor predictors for COPD exacerbation. The only classifier that demonstrates predictive capability is SpO_2. It is the best of the classifiers, however it is rated as a poor classifier, using the scale above, with an AUC of 0.693. As noted in Chapter 3, section 3.2, Jensen et al. (2012) identified the sensitivity for SpO_2 in their study as 0.70 providing concordance with the results of this work. And, as has been noted, this level of performance and predictive capability is only acceptable as the alternative is no predicition.

5.6 Optimum threshold

In order to determine the threshold at which the alarms were set, a manual review of the data sets was conducted based on the SpO_2 reading at which alarms occurred to establish the threshold for each patient. The optimum threshold was only identified for SpO_2 as it is the only classifier with predictive capability. The optimal threshold was calculated using the sensitivity

(TPR) and specificity (TNR). The threshold range for SpO_2 of 85-86 evidences the highest sensitivity and specificity as an average for the population. This is illustrated in Table 5.5 with sensitivity (TPR) of 0.62, and specificity of 0.91 (TNR).

TH VALUE SpO_2	**FPR (1-specificity)**	**TPR Sensitivity**	**TNR Specificity**	**POS (TP+FN)**	**TP**	**NEG (TN+FP)**	**TN**
80-82	0.18	0.61	0.82	77	47	360	297
85-86	0.09	*0.62	*0.91	109	68	1069	968
87-88	0.20	0.51	0.80	312	159	1038	834
89-90	0.19	0.39	0.81	698	275	1349	1086
91-92	0.23	0.07	0.77	270	156	1202	1050
Total				1466		5018	

Table 5.5. SpO_2 results by threshold

The graph of SpO_2 thresholds (Figure 5.5) indicates the level of performance for each threshold. The diagonal line (0, 0) to (1, 1) represents the performance of the classifier that is no better than random chance. A comparison of thresholds in Figure 5.4 indicates that a threshold set at 92, which is a common standard setting for respiratory illness, performs below the random chance diagonal when evaluated based on the patient outcomes- there is no relationship. Each point in the graph corresponds to a threshold for SpO_2 referenced in Figure 5.5 on the next page.

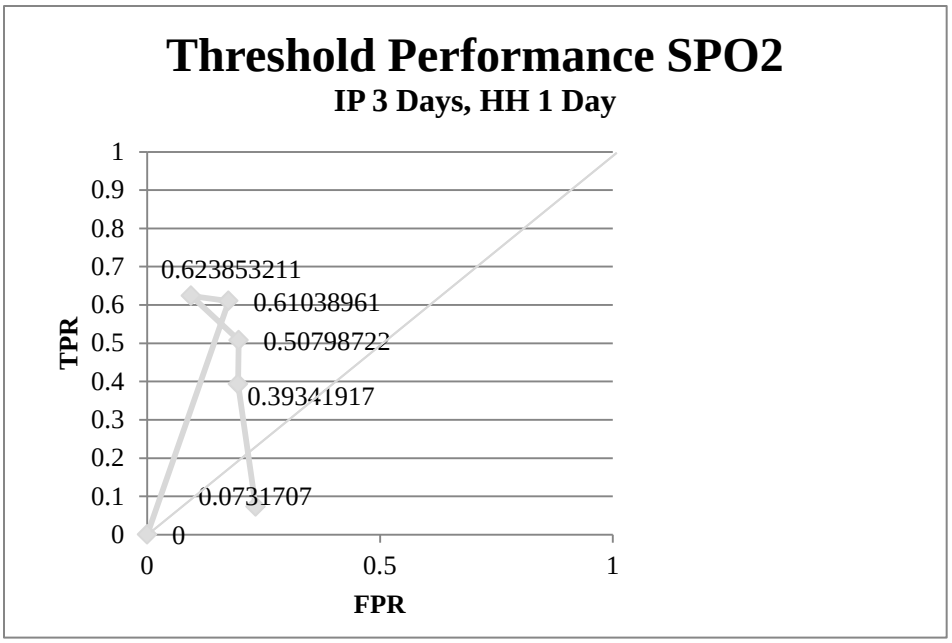


Figure 5.5. Threshold performance - SpO_2

5.7 Cost valuation

The process of assigning cost involves identification of the benefit/value of each outcome and assigning a weighted value to each category (TP, TN, FP, FN) in constructing the ROC curve. As previously noted in Chapter 4, clinical cost data is used to weight the values as they are available. The cost associated with the on-going management of the devices and contracts is outside the scope of this research but should be considered in future research. These start-up and on-going technical system costs provide additional indices to consider in evaluation of value for purchasing. Table 5.6 identifies the cost applied. The cost optimisation process is outlined in section 4.8. Apportioned values based on cost that were used to produce the ROC curves in MedCalc® are listed in Table 5.7.

COST (£)		
Specialist Nurse Home visit	**A&E**	**IP Bed day (Avg. cost)**
*£58 (CN217AF)	*£119	**£640

*UK Reference Costs 2011-2012
** NHSD Evaluation Report 2012
Table 5.6. Cost applied to ROC curves

Disease prevalence and estimated costs	
True Positive: -0.742	False Positive: -0.003
False Negative: -0.255	True Negative: 0

Table 5.7. ROC data input - cost by outcome

5.8 Limitations

This study has limitations that are identified below. Future research can be designed to modify these limitations and improve model evaluation and design.

The limitations include:

1. A lack of access to the patient record and notes limits a more detailed analysis of the degree of respiratory compromise for the individual patient. Areas where this would be helpful would be in stratification of the population, identification of the past history of exacerbation that was treated at the PCT, identifying the decision process and work processes applied in managing the RM of patient alerts. Access to the patient record would also provide a Forced Expiratory Volume (FEV1) value for lung function. The FEV1, as a measurement parameter would identify the true disease status of the patient. In addition, the lack of access to the medical record introduces the potential loss of data. Information contained in the record would quantify primary physician visits at the PCT and incorporate these costs into the cost valuation. The documentation of these visits was unavailable and was not identified in the data used for this study. In addition, specialty physician visits are

generally scheduled in advance but access to the medical record would also allow those specialty visits that occur as a result of an alarm to be calculated.

2. Classification processes have been applied consistently across all patients. However, COPD is a progressive disease and not all patients in the population measured represent the same level of disease. Stratification by level of disease with adjustment of the threshold may be needed to optimize the performance of the DM system and improve the predictive capability of the classifiers and the DSS.
3. Selection bias - DM programmes are dependent on the cooperation of the patient, and a willingness to participate. The patients in the RM programme all consented to RM indicating a willingness to participate. However, there was no matching cohort of COPD patients that had declined participation and that could be used for comparison. Outcomes may also be influenced by the choice of patients chosen for RM. The process of selection was not stipulated by the PCT and in future research a selection process should be designed at the beginning of the programme.
4. Reporting errors - all HH are entered into the system at the time of the visit and there may be some visits that were not recorded. Access to the patient record and notes would alleviate this potential data loss.
5. Verification bias - use of event based outcomes from cost data resource usage (HH, A&E, IP) did introduce the possibility of verification bias. This may influence results and introduce verification error.
6. Cost valuation is limited to direct cost with a cost-avoidance applied. However, this is subject to some data loss, including the cost of a visit to the primary care clinician at the

PCT for respiratory illness. The cost of prior HH services for each patient prior to the RM was unavailable and these costs should be included in future studies.

CHAPTER 6

DISCUSSION

"Good ideas are not adopted automatically. They must be driven into practice with courageous patience."

Hyman Rickover, 1900-1986

6.0 Discussion

The simplistic threshold model of single stratification and no selection of an optimum threshold is currently the state of the art for RM systems. However, this limits their evaluation as well as their predictive capability. The RM systems are being driven from a technology perspective and what is needed, is a reframing of the questions being asked in the context of the clinical decision support, i.e. information being provided to the clinician from which to make a decision to treat or not to treat in order to improve patient outcome. The RM system must provide correct and useful information to the clinician. This is CDS. Further research on the thresholds is needed to identify the optimum for each specific disease and subgroups of patients with co-morbid conditions. Machine learning should be linked within the DSS for clinical indices and the classifiers must be appropriate to the chronic disease as well as set at the correct threshold. These design developments can not only contribute to the benefit of the patient but also to the healthcare system.

The rigorous evaluation method of ROC analysis has been applied in this research to the current state of measures. Although the findings are specific to the COPD population measured, the methods can be applied to the broader chronic disease populations. Results indicate that current systems require design enhancements in order to improve performance and potentially, the predictive capability. Specific areas of the research are discussed in the next sections.

6.1 Classifier predictive capability

The results indicate that only SpO_2 has predictive value in this COPD population and that the relationships of the other classifiers, BP and PR, are not predictive. The results for BP and PR indicate that they perform only slightly better than random chance and are deemed to be failed classifiers. Remote monitoring can provide an avenue for communication between the patient and

the healthcare provider before deterioration makes hospitalisation inevitable. Clinicians, both doctors and nurses, must have a system that provides accurate CDS.

6. 2 Clinical decision support

CDS is provided as a result of the alerts that are generated in the DM system and which are based on the thresholds and classifiers. Non-predictive classifiers may set up a situation where home visits are made when unnecessary, as will threshold values that are either too high or too low. Alarms identifying a risk level of 2 or 3 for any of the classifiers trigger a second decision process by the Care Matron to either make or not to make a HH visit. Table 6.1 identifies the total number of HH visits made during the monitoring period and the number of visits made for each classifier.

What is important about potentially unnecessary visits is that a cost is incurred in the time and personnel providing home care services. There are visits made as a result of an FP signal for classifiers (BP and PR) that provide little value and are not predictive of COPD exacerbation. These alarms trigger potentially unnecessary visits that will impact upon the workflow process. Evaluation of workflow is outside the scope of this research, however, because this has a real impact on health services utilisation, it represents an opportunity for further research. Future research should therefore evaluate whether focusing on a specific classifier performs better and so improves outcomes and releases nursing time to concentrate on the sickest patients. The patterns of care embedded in the analysis can be used to assess the workflow to better target patients at risk for deterioration. The cost of services is impacted upon by the number of responses to alarms with no correlation to the outcome and this also needs further evaluation.

TOTAL	Total # Home visits	SpO2 # Home visits	Pulse # Home visits	SBP # Home visits	DBP # Home visits
	1734	1139	794	1354	1348

* There may be alarms for multiple classifiers that initiated the decision to visit the home and totals tabulated for all classifiers are therefore greater than the total number of visits.
Table 6.1. HH visits by classifier

Non-predictive classifiers can add cost to the healthcare system through the purchase of equipment as well as answering false alarms. This is the case for the classifiers of PR and BP for COPD. Rather design should include multiple clinically proven indicators for the specified disease based on medical research. The example of the respiratory questions from this research illustrates the need for a rigorous approach to incorporate additional measures that are targeted to specific disease states and designed in a manner that makes them usable.

6. 3 Design implications

RM can provide near real-time data to assist the clinician in making a decision that can potentially prevent a negative clinical event. However, the classifiers must be correct and accurate. In this research, SpO_2 is a marginal classifier and is rated as poor. Yet, this provides a level of clinical decision support that when linked to home health services decreases the IP bed days by 50%. This performance of the classifiers leaves room for improvement, but identifies the need to continue research into more clinically accurate classifiers by disease state using the more rigorous methodology of ROC analysis in order to improve the predictive capability.

Although disease specific measures can be designed into the DSS to allow automatic algorithms to more accurately identify clinical deterioration, additional identification of critical factors in the management of patients with COPD in the home environment using RM is needed. The iVo research by Chaudhry et al. (2010) indicates that simply applying technology to a problem evidences no value in patient outcomes. Patterns of care coordination with nursing management

coupled with the CDS supplied by RM appears to be a critical need when designing the DM programme and is supported by the results of this study and of the Iowa iVo study referenced in Chapter 3. Additional research is needed to identify the impact of the two combined services.

ROC analysis can identify features that would enhance the DM model. An interview with Dr. Jan Van Emelen, Association Internationalé de la Mutualité, illustrated the design issue and the impact on care when he stated,

> "We need to first concentrate on content. What does the programme look like which delivers the best achievable results for patients with COPD? Then we need to focus on processes and the role for the different providers. Only then should we be designing the IT systems. Otherwise, all we end up with is a series of IT projects which are incompatible and which cannot be compared to each other and where the content and processes have not been thought through properly." (Association Internationalé de la Mutualité 2010, np).

An example is the aforementioned respiratory questions in the RM system that were not amenable to ROC analysis as they were not designed with a measurement scale that could be interpreted with an alert threshold and were sporadically delivered. As noted earlier, respiratory effort, cough and sputum are important indicators of clinical deterioration and of COPD exacerbation and they should be included in the algorithm within the DSS (Burgel, Nesme-Meyer, Chanez, Caillaud, Carre, Perez & Roche 2009; Schlecht, Schwartzman & Bourbeau 2005). In order for system design to be more effective, it needs to reflect the medical knowledge for specific disease conditions.

RM can provide a communications avenue to the healthcare provider before deterioration reaches the point of needing hospitalisation. Clinicians, both doctors and nurses, must have a system that is reliably providing the needed information. So again, the appropriate classifiers for the disease process need to be built into the DSS as well as the optimal threshold for alerts generated by the system when readings exceed set thresholds.

The best approach for the design of classifiers is to integrate the clinical research on the best predictive measures for the targeted disease with design of the DSS to incorporate the most predictive clinical measures (Jurado, Feu, Jurado, Garcia, Munoz, Jimenez & Munoz 2013). This concept is starting to appear in the literature, McKinstry (2013) comes to the conclusion that,

> "Key to the success of future telemonitoring interventions will be establishment of the utility of different physiological measures and the construction of accurate predictive algorithms which can take into account individuals' risk factors, patterns of symptom and physiological parameters and recent therapy changes." (McKinstry 2013, p. 1)

This research points out the need to identify the correct parameters for measurement and link them to real world outcomes so that we do not waste time and resources on poorly designed systems.

Medical research should drive the design. The results of this research highlight the value of designing the DM system based on the disease profile and medical research, rather than only the technology. The technology should be a tool and therefore is dependent on the requirements, in this case the medical and clinical profile for COPD.

A review of the literature for COPD exacerbation indicates that there are several design elements that can be added to the system and the DSS. Examples of indicators investigated in other research include respiratory effort, dyspnea, exercise tolerance and temperature, pulmonary function, prior history of exacerbation (Burgle, Nesme-Meyer, Chanez, Caillaud, Carré, Parez & Roche 2009; Dijk, Bemt, Haak-Rongen, Bischoff, Weel, Veen & Schermer 2011; Sundh, Janson, Lisspers, Stallberg & Montgomery 2012). Designing these with a measurement scale that allows for a threshold and decision process to be applied would allow comparative data to be analysed in the DSS and with ROC analysis.

As diagnostic devices mature other classifiers can be added. However, current designs need further development to increase the predictive capability of the system and to support their use and

continual development. Linking additional measures with SpO_2 through DSS redesign has the potential to enhance accuracy and predictive capability.

6. 4 Health policy considerations

The patients in the measured cohort were monitored over an extended period of time based on clinical needs in a nationalized system that provides home health visits that are not time limited. In some countries, RM programmes include designs and payment structures that restrict the timeframe for use of the RM, as well as limiting the number of HH visits. In a seasonal and cyclic disease such as COPD, this approach limits the usefulness of these systems and may also decrease their value in cost and human benefit. Additional research is needed to inform governments and private payers of the value of these systems. The annual cost of providing RM has been measured by the VHA in the US as $1,600 per annum (Darkins, Ryan, Kobb, Foster, Edmonson, Wakefield & Lancaster 2008). One IP hospitalisation is on average, twice that amount. This is another area for further research so that public policy decisions are better able to develop service standards that include the long term cost of RM systems and value provided to patients.

Chapter 7 presents the conclusions in relation to the hypothesis, the contribution to knowledge and identifies future areas for research.

CHAPTER 7

CONCLUSIONS

"There is great satisfaction in discovery."

Nancy E. Brown Connolly, 2013

7.0 Conclusion

In conclusion, ROC analysis provides a robust and rigorous approach to measuring programme performance for technology-assisted DM models that use RM devices and as this research illustrates, can be successfully applied to this domain. Software applications to perform ROC analyses are readily available and easy to use. It is hoped that this research, by providing evidence of its value, will encourage additional effort to identify classifier performance, inform design for specific chronic diseases and provide commissioners and health systems with indicators to assist in choosing the most efficacious RM systems for their patients.

7.1 Results in relationship to the hypothesis and objectives

The hypothesis of this research has been confirmed. The process of ROC analysis is a valuable tool that can be applied to DM systems that use RM and does provide additional utility and a more accurate depiction of model performance. The results of this research meet the objectives outlined in Chapter 1 and are referenced below for clarity.

Objective 1: Apply ROC analysis to evaluation of a RM system

Application of ROC analysis methods and processes are applied successfully to a DM system that uses RM devices. Chapter 4 outlines the methods and processes applied in this research and further identified a framework that can be used for other chronic disease programs that use RM devices. The process scales to small and large volumes of data. The hypothesis under test was that the accuracy of the alert, generated in the RM system, can be measured through ROC analysis and that ROC would provide broader utility in the assessment of RM systems. The validation of the hypothesis is illustrated in the difference between the results for accuracy using a confusion matrix and the ROC curve which is agnostic to skew distribution. Measurement based solely on the confusion matrix indicates a higher sensitivity and lower specificity for the

classifiers, which arises from the skew-distribution (Krzanowski & Hand 2009). The ROC curve presents a more accurate measure of the specific performance of the system in relation to the classifiers for the COPD population in this study. Therefore, ROC analysis did provide additional utility to evaluate the RM model in a COPD population.

Objective 2: Analyse the DM model performance model based on patient outcomes

ROC analysis predicts more clearly the performance of the DM model and classifiers when based on the patient outcome. The relationship of the biometric classifiers to patient outcomes is clearly depicted by ROC and will allow health professionals and planners to better evaluate which devices are needed to improve outcomes and provide value to the patient and healthcare system. The processes outlined in the methods and results sections, illustrate how event-based outcomes are associated with the signal that is generated. This process provides a measure of the accuracy of the alert in relation to real event-based patient outcomes and health resources. The degree of predictability for the RM system establishes a benchmark for the performance of the classifier and further identifies the best classifier for prediction of COPD exacerbation. ROC analysis provides a better method to predict the performance of the classifiers than the simple statistics of a 2 X 2 classification table. One attribute of ROC that is very important to remote monitoring is its characteristic of being distribution free and "skew-agnostic". Skew in the data for chronic disease is a factor in all chronic diseases monitored in the home environment and so ROC analysis provides a more accurate depiction of model performance.

Objective 3: Identify the predictive classifiers for COPD

ROC analysis identifies SpO_2 as the most predictive classifier for COPD exacerbation based on patient outcomes, and identifies SBP, DBP and PR as failed classifiers for COPD exacerbation based on patient outcomes. When classifier performance is linked to the patient

outcome, as in this research, appropriate classifiers can be identified. This should inform the design of RM systems.

Objective 4: Identify the optimum threshold for predictive classifiers

ROC analysis identifies an optimum SpO_2 threshold of 85-86 as an average for all patients in this COPD population. Setting a threshold value is a clinical decision and as COPD is a progressive disease, lower thresholds may be necessary for patients with a more severe disease. The sensitivity and specificity of the threshold set to 87-88 is similar in value and Figure 5.5 and Table 5.5 indicate that the peak is fairly flat and a range of values may provide similar performance. Additional research is required to understand how patient specific clinical data, such as the FEV1, may be applied to provide more accurate prediction for patients as their disease progresses.

Additional findings and observations in cost valuation and design

In addition to meeting the research objectives outlined above, a cost valuation methodology was developed that includes cost avoidance. As previously noted, the savings are apportioned to HH. This research identifies additional aspects for inclusion in cost valuation that can potentially improve the assessment of cost in relationship to the value provided.

The results from this research impact on the design of the RM system and may improve predictive capability with design changes to the DSS. Design changes that incorporate disease specific measurement indices and appropriate classifiers may further enhance the capacity of the system to identify the likelihood of a COPD exacerbation. As previously noted, future design should incorporate disease specific indicators into the DM system while paying particular attention to the medical literature. This would ensure that critical measures for the specific disease are incorporated in a manner that aids clinicians in making treatment decisions and best utilises health

resources. The addition of specific classifiers for respiratory measurement, based on clinical factors such as cough, sputum, prior exacerbation and lung capacity, can also be evaluated using ROC analysis.

7.2 Research contribution

This research contributes to our knowledge by providing a new tool to be used in the evaluation of the performance of DM programmes that utilize RM devices. ROC analysis and processes have empirically illustrated the performance of the RM system, and the predictive capacity of the classifiers used to monitor patients with COPD. The use of ROC methods have enabled comparison of the performance of the classifiers at varying thresholds and the optimum operating point to be determined for SpO_2. This type of analysis can assist in the development of policies and procedures related to the workflow for patient support services. It has further identified issues, such as the design of the respiratory questions that need changes so that they can be integrated into the clinical decision process in order to improve patient outcomes.

Trade-offs in performance are more accurately and measurably predicted using ROC analysis. Vendor claims are verifiable based on an evaluation of system performance that are based on patient outcomes. Methods are applicable to programmes utilising RM systems for other chronic diseases and can be used to identify the performance of classifiers for specific medical conditions and the value that is added to the healthcare system can be quantified using cost evaluation processes. The generalizability of results to other health systems and countries is equally significant. Results enable purchasers, insurers, health systems and policy makers to evaluate the effectiveness of these systems during the planning process. This research further provides a substantive background with which to discuss the inclusion of ROC as an integral

component of the programme evaluation with their contracted vendors. Future research efforts that are needed to improve the predictive capacity of RM systems are presented in the next section.

7.3 Future research recommendations

Recommendations stemming from this research include both COPD specific recommendations and broader design recommendations for other areas of disease management.

1. Redesign respiratory measures for COPD and integrate into the DSS.
2. Develop and design classifiers based on medical research for known variables that lead to acute episodes by disease.
3. Identify classifiers based on predictive performance for specified chronic disease states.
4. Identify optimum thresholds by classifier and disease based on patient profiles and level of disease.
5. Analysis of data patterns should be included in the research to identify critical factors in patient management.
6. Focus research on specific chronic disease designs to improve predictive capacity and target resources to patients with the greatest need.
7. Develop workflow analysis of design changes.
8. Increase decision support research in the telehealth and remote monitoring areas to evaluate designs and systems for DM.
9. Calculate the cost for multiple scenarios i.e. implementation, start-up, and on-going administration.

REFERENCES AND BIBLIOGRAPHY

Abbasi, K (ed.) (2010), Doctors and managers: radical thinking required, *Journal of the Royal Society of Medicine,* vol. 103, no. 10, pp. 383-383.

Ackerman, MJ, Filart, R, Burgess, LP, Lee, I & Poropatich, RK (2010), Developing next-generation telehealth tools and technologies: Patients, systems, and data perspectives. *Telemedicine and e-Health,* vol. 16, no. 1, pp. 93-95.

Acronyms. Available from: <http://acronyms.thefreedictionary.com/Spo2>, [accessed May 22, 2012].

Alder J, Mayhew, L, Moody, S, & Morris, R (2005), *The Chronic Disease Burden – An Analysis of Health Risks and Health Care Usage,* Cass Business School, City University, London UK.

Association Internationale de la Mutualité, Interview with Dr Jan Van Emelen (n.d.), Available from: <http://www.healthcareeuropa.com/articles/janvanemelen>, [accessed December 1, 2010].

Australian Institute of Health and Welfare (2010), *Healthcare Costs*. Available from: <http://www.aihw.gov.au/cdarf/data_pages/health_care_costs/index.cfm>, p.24, [accessed November 22, 2010].

Avison, D & Young, T (2007), Time to rethink healthcare and ICT? *Communications of the ACM,* vol.50, no. 6, pp. 69-74.

Bartoli, L, Zanaboni, P, Masella, C & Ursini, N (2009), Systematic Review of Telemedicine Services for Patients Affected by Chronic Obstructive Pulmonary Disease (COPD), *Telemedicine and e-Health*, vol.15, no. 9, pp. 877-883.

Basilakis, J, Lovell, NH, Redmond, SJ & Celler, BG (2010), Design of a decision-support architecture for management of remotely monitored patients, *IEEE Transactions on Information Technology in Biomedicine*, vol.14, no. 5, pp. 1216-1226.

Begg, CB & Greenes, RA (1983), Assessment of diagnostic tests when disease verification is subject to selection bias, *Biometrics*, vol. 39, no. 1, pp. 207–215.

Berner, ES (2009), Clinical decision support systems: State of the Art, *AHRQ Publication No. 09-0069-EF*, Agency for Healthcare Research and Quality. Available from: <http://www.healthit.ahrq.gov/sites/default/files/docs/page/09-0069-EF_1.pdf>, [accessed February 5, 2010].

Bossuyt, P & Leeflang (2008), Chapter 6: Searching for Studies, Cochrane handbook for systematic reviews of diagnostic test accuracy, *Cochrane Collaboration*, version 0.4, pp. 1-7.

Brown-Connolly, NE, (2002), Patient satisfaction with telemedical access to specialty services in rural California, *Journal of Telemedicine and Telecare*, vol. 8 Suppl. 2, pp. 7-10.

Busse, R, Blümel, M, Scheller-Kreinsen, D & Zentner, A (2010), Tackling Chronic Diseases in Europe: Strategies, interventions and challenges, European Observatory on Health Systems and Policies Studies Series, *WHO Press*, no. 20. Available from: <http://apps.who.int/bookorders/anglais/detart1.jsp?codlan=1&codcol=34&codcch=97>, [accessed November 15, 2011].

Buntin, MB (2006), Rigorous disease management evaluation. *Journal of Evaluation in Clinical Practice*, vol.12, no. 2, pp. 121-123.

Burgel PR, Nesme-Meyer P, Chanez P, Caillaud D, Carré P, Perez T & Roche N (2009), Cough and sputum production are associated with frequent exacerbations and hospitalisations in COPD subjects. Initiatives Bronchopneumopathie Chronique Obstructive Scientific Committee, *Chest*, vol. 135, no. 4, pp.975-982.

Castells, Manuel (1997, second edition, 2004). The Power of Identity, The Information Age: Economy, Society and Culture Vol. II. Malden, MA; Oxford, UK: Blackwell. ISBN 978-1-4051-0713-6.

CDC (2004 & 2009), The burden of chronic diseases and their risk factors: national and state perspectives, *US Department of Health and Human Services.* Available from: <http://www.cdc.gov/chronicdisease/resources/publications/AAG/chronic.htm>, [accessed June 3, 2013].

Chaudhry, S, Barton, B, Mattera, J, Spertus, J & Krumholz, H (2007), Randomized Trial of Telemonitoring to Improve Heart Failure Outcomes (Tele-HF): Study Design, *Journal Cardiac Failure*, vol. 13, no. 9, pp. 709–714.

Chaudhry, S, Mattera, J, Spertus, J, Curtis, JP, Herrin, J, Zhenqiu L, Phillips, C, Hodshon, B, Cooper, L & Krumholz, H (2010), Telemonitoring in Patients with Heart Failure, New England Journal of Medicine, vol.363, pp.2301-2309.

Clarke, M, Brown-Connolly, N, Smith, R & Fursse, J (2012), NHS Direct Telehealth Evaluation Report. Available from NHS Direct, Uxbridge, UK.

Congressional Budget Office (CBO) (2004), An Analysis of the Literature on Disease Management Programs. Available from:

<http://www.cbo.gov/search/apachesolr_search?keys=disease+management&op=search, [accessed January 10, 2013].

Darkins A, Ryan, P, Kobb, R, Foster, L, Edmonson, E,Wakefield, & Lancaster, A (2008), Care Coordination/Home Telehealth: The Systematic Implementation of Health Informatics, Home Telehealth, and Disease Management to Support the Care of Veteran Patients with Chronic Conditions, *Telemedicine and e-Health*, vol. 14, no. 10, pp. 1118-1126.

De Carlo, LT (1998), Signal Detection Theory and Generalized Linear Models Psychological Methods, *American Psychological Association*, vol. 3, no. 2, pp. 186-205.

Department of Health UK (2011), *Reference Costs 2011-2012*, Available from: <https://www.gov.uk/.../213060/2011-12-reference-costs-publication.pdf>, [accessed March 13, 2013].

Department of Health Gateway (2011), *Whole System Demonstrator Programme, Headline Findings,* December 2011, Ref: 16972, p. 3. Available from: <https://www.gov.uk/government/publications/whole-system-demonstrator-programme-headline-findings-december-2011>, [accessed June 3, 2013].

Department of Health UK (2012), Whole System Demonstrator Programme Report. Available from: <http://www.dh.gov.uk/dr_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_131689.pdf>, [accessed January 7, 2012]

De San Miguel, K, Smith, J & Lewin, G (2010), Telehealth Research across the community – remote monitoring of chronic obstructive pulmonary disease, Telehealth Final Report, *Silver*

Chain. Available from: <http://www.eostech.com.au/documents/01_Final_report_COPD_DOHA.pdf>, [accessed August 30, 2013].

De Vet HCW, Eisinga, A, Riphagen, I, Aertgeerts, B & Pewsner, D (2008), Chapter 7: Searching for Studies. Cochrane handbook for systematic reviews of diagnostic test accuracy, Version 0.4 [updated September 2008]. *The Cochrane Collaboration*. Available from: <http://srdta.cochrane.org/handbook-dta-reviews>, [accessed September 30, 2010].

Dove, HG & Duncan, I (2005), Paper 3: Estimating Savings, Utilization Rate Changes, and Return on Investment from Care Management Interventions, Selective Literature Review of Care Management Interventions, *US Society of Actuaries.* Available from: <http://www.soa.org/search.aspx?go=True&q=&page=1&pagesize=10&or=True&taxid=4294967447>, [accessed October 27, 2010].

Dijk, WD, Bemt, L, Haak-Rongen, S, Bischoff, E, Weel, C, Veen, JC & Schermer, TR (2011), Multidimensional prognostic indices for use in COPD patient care. A systematic review, *Respiratory Research*, vol.12, pp. 151-9921-12-151.

Duncan, I (2004), Part2: Actuarial Issues in Care Management Interventions, Paper 4: Understanding the Economics of Disease Management Programs, *US Society of Actuaries*. Available from: <http://www.soa.org/search.aspx?go=True&q=&page=1&pagesize=10&or=True&taxid=4294967447>, [accessed October 27, 2010].

Duncan, I (2005), Paper 5:Evaluating Disease Management Savings Outcomes 5, *US Society of Actuaries.* Available from: <http://www.soa.org/search.aspx?go=True&q=&page=1&pagesize=10&or=True&taxid=4294967447>, [accessed October 27, 2010].

Duncan, I (2005), Part 2: Actuarial Issues in Care Management Interventions, Paper 6: An Actuarial Method for Evaluating disease Management Options, *US Society of Actuaries*. Available from: <http://www.soa.org/search.aspx?go=True&q=&page=1&pagesize=10&or=True&taxid=4294967447>, [accessed October 27, 2010].

Duncan, I, Owen, R & Dove, H (2006), Paper 8: Testing Actuarial Methods for Evaluating Disease Management Savings Outcomes, US Society of Actuaries' Health Section and the Committee on Knowledge Extension Research, *US Society of Actuaries.* Available from: <http://www.soa.org/search.aspx?go=True&q=&page=1&pagesize=10&or=True&taxid=4294967447>, [accessed October 28, 2010].

Duncan, I (2008), An Introduction to Care Management Interventions and Their Implications for Actuaries, *US Society of Actuaries*. Available from: <http://www.soa.org/search.aspx?go=True&q=&page=1&pagesize=10&or=True&taxid=4294967447>, [accessed October 27, 2010]. [accessed October 27, 2010].

Egan, JP (1975), *Signal Detection Theory and ROC Analysis, Series in Cognition and Perception*, Academic Press. New York, New York.

Fawcett, T (2004). *ROC Graphs: Notes and Practical Considerations for Researchers*. Kluwer Academic Publishers, Netherlands.

Fawcett, T (2006), An introduction to ROC analysis, Pattern Recognition Letters, vol. 27, no. 8, June 2006, Pages 861-874, ISSN 0167-8655. Available from: <http://dx.doi.org/10.1016/j.patrec.2005.10.010>, <http://www.sciencedirect.com/science/article/pii/S016786550500303X>, [accessed October 3, 2013).

Fawcett T & Mishra N (ed) (2003), Proceedings of the 20th International Conference on Machine Learning (ICML'03), *AAAI Press*, pp. 416–423. Available from: <http://www.aaai.org/Library/ICML/icml03contents.php>, [accessed May 8, 2013].

Field, MJ (ed.) (1996), Committee on Evaluating Clinical Applications of Telemedicine, Institute of Medicine, *Telemedicine: a guide to assessing telecommunications in health care*, National Academy Press. Available from: <http://www.nap.edu/openbook.php?isbn=0309055318>, [accessed September 27, 2010].

Figueras, J (2012), A response to the crisis in the EU? Elements for a decision, *Paper presentation Tel-Aviv*. European Observatory on Health Systems and Policies Series. Available from: <www.healthobservatory.eu>, [accessed May 6, 2013].

Foreman, MG, DeMeo, DL, Hersh, CP, Reilly, JJ & Silverman, EK (2007), Clinical determinants of exacerbations in severe early-onset COPD, *European Respiratory Journal*, vol.30, no.6, pp. 1124-30.

Geisser, S & Johnson, WM (2006), *Modes of Parametric Statistical Inference*, New York; London, John Wiley & Sons.

Gomaa, WH, Morrow, T & Muntendam, P (2001), Technology-Based Disease Management: A Low-Cost, High-Value Solution for the Management of Chronic Disease, *Disease Management & Health Outcomes,* vol. 9, no. 10, pp. 577-588.

Green, DM & Swets, JA (1966), *Signal Detection Theory and Psychophysics*, New York ; London, Wiley.

Greenes, RA (ed.) (2007), *Clinical Decision Support: The Road Ahead*, Boston, Elsevier Academic Press.

Hamilton, HJ (2000-2009), *Knowledge Discovery in Databases,* Course notes in Computer Science 831: University of Regina, Last updated June 8, 2012, section 4. Available from: <http://www2.cs.uregina.ca/~dbd/cs831/index.html>, [accessed June 6, 2013].

Interview with Dr. Jan Van Emelen (2010), *Healthcare Europa*, p.39. Available from: <http://www.healthcareeuropa.com/articles/janvanemelen>, [accessed December 1, 2010].

Heeger, D (2003), *Signal Detection Theory*, Lecture notes, Department of Psychology, New York University. Available from: <http://www.cns.nyu.edu/~david/handouts/sdt/sdt.html> [accessed September 9, 2012].

Hickman, D, Brown-Connolly, NE, Garloff, D, Kunath, K & Appelgate, WK (2004), Congestive Heart Failure Telemanagement Demonstrations, Congestive Heart Failure Case Management in

Iowa. *Report to the Office for the Advancement of TeleHealth*, HRSA Grant Number 1D1BTH01054-01-00.

Holland, A (2013), Telehealth reduces hospital admission rates in patients with COPD. *Journal of Physiotherapy*, vol. 59, no. 2, pp. 129-9553(13)70168-1.

Hughes, McRoberts, Burnett & Hughes, (1999), Decision-making and diagnosis in disease management, *Plant Pathology*, vol. 48, no. 2, pp. 147-153.

Humphreys, LG & Swet, JA (1991), Comparison of Predictive Validities Measured With Biserial Correlations and ROCs of Signal Detection Theory, *Journal of Applied Psychology*, vol. 76, no. 2, pp. 316-321.

Hwang, W, Weller, W, Ireys, H & Anderson, G (2001), Out of pocket medical spending for care of chronic conditions, *Health Affairs*, vol. 20, pp.268-269.

Institute of Medicine (2008), *IOM Executive Office Roundtable on Value & Science-Driven Health Care, Meeting 6: Engineering a Learning Healthcare System - A Look at the Future*. Available from: <http://iom.edu/Activities/Quality/VSRT/2008-APR-29.aspx> [accessed November 1, 2010].

Investopedia, <http://www.investopedia.com/terms/n/nonparametric-statistics.asp>, [accessed July 13, 2013].

Jensen, MH, Cichosz, SL, Dinesen, B & Hejlesen, OK (2012), Moving prediction of exacerbation in chronic obstructive pulmonary disease for patients in telecare. *Journal of Telemedicine and Telecare*, vol. 18, no. 2, pp. 99-103.

Jurado-Gamez, B, Feu-Collado, N, Jurado-Garcia, JC, Garcia-Gil, F, Munoz-Gomariz, E, Jimenez Murillo, L & Munoz-Caberra, L (2013), Home intervention and predictor variables for rehospitalisation in chronic obstructive pulmonary disease exacerbations, *Archivos de Bronconeumologia*, vol.49, no. 1, pp. 10-14.

Jaana, M, Pare, G & Sicotte, C (2009), Home Telemonitoring for Respiratory Conditions: A Systematic Review, *The American Journal of Managed Care*, vol. 15, no. 9, pp.313-320.

Kalorama Information, Advanced Remote Patient Monitoring Systems, *Report Linker*, March 2013. <http://www.reportlinker.com/p01172037/Advanced-Remote-Patient-Monitoring-Systems.html>, [accessed September 30, 2013].

Kane, RL, Chen, Q, Finch, M, Blewett, L, Burns, R & Moskowitz M (2000), The optimal outcomes of post-hospital care under Medicare, *Health Services Research*, vol. 35, pp.615–661.

Kocks, JW, Van Den Berg, JW, Kerstjens, HA, Uil, SM, Vonk, JM, De Jong, YP, Tsiligianni, Ig & Van Der Molen, T (2013), Day-to-day measurement of patient-reported outcomes in exacerbations of chronic obstructive pulmonary disease, *International journal of chronic obstructive pulmonary disease*, vol.8, pp. 273-286.

Kohavi, R &, Provost, F (eds.), (1998), Introduction: On Applied Research in Machine Learning, *Machine Learning*, vol. 30,no. 2, pp. 127-132. Available from: <http://citeseerx.ist.psu.edu/showciting?cid=753041>, [accessed November 10, 2010].

Kristensen, FB, Makel, M, Neikter, SA, et al. (2009), European network for health technology assessment, EUnetHTA: Planning, development, and implementation of a sustainable European network for health technology assessment, *International Journal Technology Assessment Health*

Care, vol. 25, Suppl. 2, pp.107-116. Available from: <http://www.eunethta.eu>, [accessed July 1, 2011].

Krzanowski, WJ & Hand, DJ. (2009), *ROC curves for continuous data: monographs on statistics and applied probability*, CRC Press, Taylor and Francis Group, London, U.K.

Lange, P, Marott, JL, Vestbo, J, Olsen, KR, Ingebrigtsen, TS, Dahl, M & Nordestgaard, BG (2012), Prediction of the clinical course of chronic obstructive pulmonary disease, using the new GOLD classification: a study of the general population, *American Journal of Respiratory and Critical Care Medicine*, vol.186, no. 10, pp. 975-981.

Lasko, TA, Bhagwat, JG, Zou, KH & Ohno-Machado, L (2005), The use of receiver operating characteristic curves in biomedical informatics, *Journal of Biomedical Informatics*, vol. 38**,** no. 5, pp. 404-415.

Lee, SJ, Lee, SH, Kim, YE, Cho, YJ, Jeong, YY, Kim, HC, Kim, JH, You, JJ, Yoon, CH, Lee, JD & Hwang, YS (2012), Clinical Features according to the Frequency of Acute Exacerbation in COPD, *Tuberculosis and Respiratory Diseases*, vol. 72, no. 4, pp. 367-373.

Lefebvre C, Manheimer E & Glanville J (2009), Chapter 6: Searching for studies in Higgins JPT & Green S, (eds.) *Cochrane Handbook for Systematic Reviews of Interventions, version 5.0.2, The Cochrane Collaboration*. Available from: <http://www.cochrane-handbook.org>, [accessed July 1, 2010].

Levine, S, Reyes, JY, Schwartz, R, Schmidt, D, Schwab, T & Leung, M (2006), Disease Management of the Frail Elderly Population, *Disease Management & Health Outcomes*, vol. 14, no. 4, pp. 235.

Linden, A (2006), Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis, *Journal of Evaluation in Clinical Practice,* vol. 12, no. 2, pp. 132-139.

Luck, J, Parkerton, P & Hagigi, F (2007), What is the business case for improving care for patients with complex conditions? *Journal of General Internal Medicine,* vol. 22, suppl. 3, pp. 396-402.

Lusted, LB (1971), Signal detectability and medical decision making, *Science,* vol. 171, pp. 1217-1219.

Lusted, L (1978), General problems in medical decision-making with comments on ROC analysis , *Seminars in Nuclear Medicine,* Elsevier Inc., vol. 8, no. 4, pp. 299-306. Available from: <http://dx.doi.org/10.1016/S0001-2998(78)80015-4>, [accessed June 20, 2013].

Lorenzi, NM (2004), Beyond the gadgets, *British Medical Journal,* vol. 328, no. 7449, pp. 1146-1147.

Macstravic, S (2007), Evaluating disease management results: individuals and cohorts vs populations, *Disease Management,* vol. 10, no. 4, pp. 185-188.

Mandic, S, Go, C, Aggarwal, I, Myers, J & Froelicher, V (2008), Relationship of predictive modeling to receiver operating characteristics, *Journal of Cardiopulmonary Rehabilitation Prevention,* vol. 28, no. 6, pp. 415-419.

Mathers, CD & Loncar, D (2006), Projections of global mortality and burden of disease from 2002 to 2030, *PLoS Medicine,* vol. 3, no. 11, e442.

Mattke, S, Seid, M & Ma, S (2007), Evidence for the Effect of Disease Management: Is $1 Bn a Year a Good Investment? *The American Journal of Managed Care,* vol.12, no. 12, pp. 670-676.

McFall, R & Treat, TA (1999), Quantifying the information value of clinical assessment with Signal Detection Theory, *Annual Review of Psychology*, vol. 50, pp. 215-241.

McKinstry, B (2013), The use of remote monitoring technologies in managing chronic obstructive pulmonary disease, *QJM: Journal of the Association of Physicians,* vol. 106, no. 10, pp. 883-885., doi:10.1093/qjmed/hct06. Available from: <http://qjmed.oxfordjournals.org/content/106/10/883.full.pdf+html?sid=b30a5f02-6bff-44b4-9ad8-51fa6bb4da49>, [accessed August 24, 2013].

MedCalc®Version 12.7.0 (2013), Reference manual, Last modified: June 16, 2013 , © 1993-2013 MedCalc Software bvba MedCalc Software, Acacialaan 22, B-8400 Ostend, Belgium, [accessed April 21, 2013].

Medterms. Available online: <http://www.medterms.com/script/main/art.asp?articlekey=33490>, [accessed June 27, 2013].

Metz, CE (1978), Basic principles of ROC analysis, *Seminars in Nuclear Medicine*, vol.8, pp. 283-298.

Monegain, B (ed.) (2013), Remote monitoring market growing fast, Healthcare IT News. Available from: < http://www.healthcareitnews.com/news/remote-monitoring-market-growing-fast>, [accessed May 15, 2013].

Motegi, T, Jones, RC, Ishii, T, Hattori, K, Kusunoki, Y, Furutate, R, Yamada, K, Gemma, A & Kida, K (2013), A comparison of three multidimensional indices of COPD severity as predictors of future exacerbations, *International journal of chronic obstructive pulmonary disease,* vol. 8, pp. 259-271.

Murray, CJL & Lopez, AD (1997), Alternative projections of mortality and disability by cause 1990–2020: Global Burden of Disease Study, *Lancet,* vol. 349, pp. 1498-1504.

New England Healthcare Institute (2009), Fast Adoption of Significant Technologies. Detailed Technology Analysis: Home Telehealth. Available online: <http://www.nehi.net/publications/42/detailed_technology_analysis_home_telehealth>, [accessed June 3, 2013].

NSRC2 Primary Care Trust reference cost schedules 2010-11 Available online: <http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_131146.xls>, [accessed February 15, 2013].

Obuchowski, NA (2003), Receiver Operating Characteristic Curves and Their Use in Radiology, *Radiology* , vol. 229, no. 1, pp. 3-8.

Organisation for Economic Co-operation and Development (OECD) (2010), *OECD Health Policy Studies Improving Health Sector Efficiency The Role of Information and Communication Technologies,* France: OECD Publishing.

Papi A, Bellettato CM, Braccioni F, RomagnoliM, Casolari P, Caramori G, Fabbri LM & Johnston SL (2006), Infections and airway inflammation in chronic obstructive pulmonary disease severe exacerbations, *American Journal Respiratory Critical Care Medicine,* vol. 173, pp. 1114–1121.

Pepe, M, (1998), Three approaches to regression analysis of receiver operating characteristic curves for continuous test results, *Biometrics,* vol. 54, no. 1, pp. 124-135.

Personal communication, Iowa, US, November 2004

Peterson, WW, Birdsall, T G & Fox, WC (1954), The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory,* vol.4, pp. 171-212. Reprinted Luce, RD, Bush RR & Galanter, E [eds.], (1963), *Readings in mathematical psychology,* pp. 167-211, New York, Wiley.

Pintea, S & Moldovan, R (2009), The Receiver-Operating Characteristic (Roc) Analysis: Fundamentals and Applications in Clinical Psychology, *Journal of Cognitive & Behavioral Psychotherapies,* vol. 9, no. 1, pp. 49-66.

Quotes. Available from: <http://www.brainyquote.com>, [accessed December 2, 2010].

Reid, PP, Compton, WD, Grossman, JH & Fanjiang, G (2005), Building a better delivery system: a new engineering health care partnership, *The National Academy of Engineering, Institute of Medicine and National Academies Press,* Washington, D.C., National Academies Press.

Roca, M, Verduri, A, Corbetta, L, Clini, E, Fabbri, LM & Beghe, B (2013), Mechanisms of acute exacerbation of respiratory symptoms in chronic obstructive pulmonary disease, *European Journal of Clinical Investigation,* vol. 43, no. 5, pp. 510-521.

Schlecht, NF, Schwartzman, K & Bourbeau, J (2005), Dyspnea as clinical indicator in patients with chronic obstructive pulmonary disease, *ChronicRrespiratory Disease,* vol. 2, no. 4, pp. 183-191.

Schrijvers, G (2009), Disease management: a proposal for a new definition, *International Journal of Integrated Care*. Published online: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2663707/>, [accessed May 20, 2013].

Seemungal T, Harper-Owen R, Bhowmik A, Moric I, Sanderson G, Message S, Macallum P, Meade TW, Jeffries DJ, Johnston SL & Wedzicha JA (2001), Respiratory viruses, symptoms, and inflammatory markers in acute exacerbations and stable chronic obstructive pulmonary disease, *American Journal of Respiratory Critical Care Medicine*, vol. 164, pp.1618–1623.

Selecky, CE (2008), Disease Management of Chronic Obstructive Pulmonary Disease from a Disease Management Organization Perspective, *Disease Management & Health Outcomes,* vol. 16, no. 5, pp. 319-325.

Shiu SY & Gatsonis, C (2008), The predictive receiver operating characteristic curve for the joint assessment of the positive and negative predictive values, *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences,* vol. 366, no. 1874, pp. 2313-2333.

Speedie, SM, Ferguson, AS, Sanders, J & Doarn, C (2008), Telehealth: The Promise of New Care Delivery Models, *Telemedicine and e-Health*, vol. 14, no. 9, pp. 964-967.

Springer reference (2013), *Confusion matrix.* Available online: <http://www.springerreference.com/docs/html/chapterdbid/178869.html>, [accessed July 8, 2013].

Spyros, K, Paré, G & Jaana, M (2013), Systematic Reviews and Meta-Analyses of Home Telemonitoring Interventions for Patients With Chronic Diseases: A Critical Assessment of Their Methodological Quality, *Journal Medical Internet Research,* vol. 15, no.7,e150. Available online: <http://www.jmir.org/2013/7/e150/>, [accessed August 30, 2013].

StatTrek (2009), Statistics and Probability Dictionary. Available online: <http://stattrek.com/statistics/dictionary.aspx?definition=z_score>, [accessed July 10, 2013].

Stehman, SV (1997), Selecting and interpreting measures of thematic classification accuracy, *Remote Sensing of Environment,* vol. 62, no. 1, pp. 77–89.

Streiner, DL & Cairney, J (2007), What's under the ROC? An introduction to receiver operating characteristics curves, *Canadian Journal of Psychiatry (Revue Canadienne de Psychiatrie),* vol. 52, no. 2, pp. 121-128.

Stuart, SM, Ferguson, AS, Sanders, J & Doarn, CR, (2008), Telehealth: The Promise of New care Delivery Models, *Telemedicine and e-Health,* vol. 14, no. 9, pp. 964-967.

Sundh, J, Janson, C, Lisspers, K, Stallberg, B & Montgomery, S (2012), The Dyspnoea, Obstruction, Smoking, Exacerbation (DOSE) index is predictive of mortality in COPD. *Primary care respiratory journal: journal of the General Practice Airways Group,* vol. 21, no. 3, pp. 295-301.

Swets, JA (1964), *Signal Detection and Recognition by Human Observers: Contemporary Readings.* New York; London: Wiley.

Swets, JA (1973), The Relative Operating Characteristic in Psychology: A technique for isolating effects of response bias finds wide use in the study of perception and cognition. *Science,* vol. 182, no. 4116, pp. 990-1000.

Swets, JA (1979), ROC analysis applied to the evaluation of medical imaging techniques, *Investigative Radiology,* vol. 14, no. 2, pp. 109-121.

Swets, JA (1986), Indices of Discrimination or Diagnostic Accuracy: Their ROCs and Implied Models, *Psychological Bulletin*, vol.99, no. 1, pp. 100-117.

Swets, JA (1986), Form of Empirical ROCs in Discrimination and Diagnostic Tasks: Implications for Theory and Measurement of Performance, *Psychological Bulletin*, vol. 99, no. 2, pp. 181-198.

Swets, JA (1988) cited in Hamilton course notes, 2012, sec. 4, Available online: <http://www2.cs.uregina.ca/~dbd/cs831/notes/ROC/ROC.html>, [accessed June 6, 2013].

Swets, JA, Getty, DJ, Pickett, M, D'Orsi, CJ, Seltzer, SE & McNeil, BJ (1991), Enhancing and evaluating diagnostic accuracy, *Medical Decision-Making: an International Journal of the Society for Medical Decision Making*, vol.11, no. 1, pp. 9-18.

Swets, JA (1992), The Science of Choosing the Right Decision Threshold in High-Stakes Diagnostics, *American Psychologist*, vol. 47, no. 4, pp. 522-532.

Swets, JA (1996), *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers.* Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey.

Tape, TG (n.d.), *Interpreting Diagnostic Tests,* University of Nebraska Medical Center. Available online: <http://gim.unmc.edu/dxtests/Default.htm>, [accessed July 8, 2013].

Tripepi, G, Jager, KJ, Dekker, FW & Zoccali, C (2009), Diagnostic methods 2: receiver operating characteristic (ROC) curves, *Kidney International*, vol.76, no. 3, pp. 252-256.

Wagner EH (1998), Chronic disease management: What will it take to improve care for chronic illness? *Effective Clinical Practice*, vol. 1, no. 1, pp. 2-4. Available from: <http://www.ihi.org/IHI/Topics/ChronicConditions/>, [accessed May 29, 2011].

Wald, A (1950), *Statistical Decision Functions*. New York, Wiley.

Wan, ES, DeMeo, DL, Hersh, CP, Shapiro, SD, Rosiello, RA, Sama, SR, Fuhlbrigge, AL, Foreman, MG & Silverman, EK (2010), Clinical Predictors of Frequent Exacerbations in Subjects with Severe Chronic Obstructive Pulmonary Disease (COPD), *Respiratory Medicine*, vol. 105, no. 4, pp. 588–594.

Webb, M & Howson, H (2006), International overview of the evidence on effective service models in chronic disease management, V09-05-06, *National Public Health Service for Wales*, Wales, UK.

Whole System Demonstrator Programme, Available from: <http://www.fastuk.org/research/projview.php?id=1436>, [accessed December 1, 2010].

Wikipedia (2013), *Receiver Operating Characteristic*. Available from: <http://en.wikipedia.org/wiki/Receiver_operating_characteristic>, [accessed June 7, 2013].

Wikipedia (2013), *Predictive Analytics*. Available from: <http://en.wikipedia.org/wiki/Predictive_analytics >, [accessed June 15, 2013].

World Health Organization, (2010), Noncommunicable Diseases: Preventing Chronic Disease: a vital investment, *WHO Press*. Available from: <http://www.who.int/chp/chronic_disease_report/en/>, [accessed November 3, 2010]

World Health Organization, *Definition COPD*. Available from: <http://www.who.int/respiratory/copd/en/>, [accessed August 19, 2013].

Zhou, XH (1998), Correcting for verification bias in studies of a diagnostic test's accuracy, *Statistical Methods in Medical Research*, vol. 7, no. 4, pp. 337-353.

Zou, KH & O'Malley, AJ (2005), A Bayesian Hierarchical Non-Linear Regression Model in Receiver Operating Characteristic Analysis of Clustered Continuous Diagnostic Data, *Biometrical Journal*, vol. 47, no. 4, pp. 417-427.

Zou, KH, O'Malley, AJ & Mauri, L (2007), Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models, *Circulation*, vol. 115, no. 5, pp. 654-657.

Zweig, MH & Campbell, G (1993), Receiver operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, *Clinical Chemistry*, vol. 39, pp.561-577.

APPENDIX A

School of Information Systems, Computing and Mathematics
David Gilbert, Head of School, Professor of Computing
Jasna Kuljis, Head of Information Systems and Computing, Professor of Computing
Tony Rawlins, Head of Mathematical Science, Professor of Mathematics

Brunel UNIVERSITY WEST LONDON

Brunel University, Uxbridge,
Middlesex UB8 3PH, UK
Telephone: +44(0) 1895 274000
Fax: +44(0) 1895 251686
Email:
Laurence.Brooks@brunel.ac.uk
Lampros.Stergioulas@brunel.ac.uk
Annette.Payne@brunel.ac.uk

Date: 5th September 2011

STATEMENT OF ETHICS APPROVAL

Proposer: Nancy Connolly

Title: Application of Receiver Operating Characteristic (ROC) Analysis to Evaluate Predictive Value of Technology Assisted Disease Management Model

The school's research ethics committee has considered the proposal recently submitted by you. Acting under delegated authority, the committee is satisfied that there is no objection on ethical grounds to the proposed study. Approval is given on the understanding that you will adhere to the terms agreed with participants and to inform the committee of any change of plans in relations to the information provided in the application form.

Yours sincerely,

Dr. Laurence Brooks, Chair of the Research Ethics Committee
SISCM

APPENDIX B

Health Research Authori

NRES Committee West Midlands - Solih
East Midlands REC Ce
The Old Ch
Royal Standard F
Notting
NG1

Telephone: 0115 883

15 April 2013

Mrs. Nancy E. Brown Connolly
5907 Irving Blvd. NW
Albuquerque
New Mexico, USA
87114

Dear Mrs. Brown Connolly

Study title:	**Receiver Operating Characteristic (ROC) analysis of a remote monitoring disease management system to determine optimum threshold and predictive value.**
REC reference:	**13/WM/0172**
Protocol number:	**N/A**
IRAS project ID:	**124782**

The Proportionate Review Sub-committee of the NRES Committee West Midlands - Solihull reviewed the above application on 10 April 2013.

We plan to publish your research summary wording for the above study on the NRES website, together with your contact details, unless you expressly withhold permission to do so. Publication will be no earlier than three months from the date of this favourable opinion letter. Should you wish to provide a substitute contact point, require further information, or wish to withhold permission to publish, please contact the Co-ordinator Maria Morledge, NRESCommitteeWesttMidlands.Solihull@nhs.net.

Ethical opinion

On behalf of the Committee, the sub-committee gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission ("R&D approval") should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements.

Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of approvals from host organisations.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

You should notify the REC in writing once all conditions have been met (except for site approvals from host organisations) and provide copies of any revised documentation with updated version numbers. The REC will acknowledge receipt and provide a final list of the approved documentation for the study, which can be made available to host organisations to facilitate their permission for the study. Failure to provide the final versions to the REC may cause delay in obtaining permissions.

Approved documents

The documents reviewed and approved were:

Document	*Version*	*Date*
Investigator CV	Nancy Brown-Connolly	
Investigator CV	Dr Malcolm Clarke	
Other: Statement of Ethics Approval - Brunel University		05 September 2011
Protocol	Designated version 1	12 January 2012
REC application	124782/433155/1/161	25 March 2013

Membership of the Proportionate Review Sub-Committee

The members of the Sub-Committee who took part in the review are listed on the attached sheet.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

Feedback

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.
information is available at National Research Ethics Service website > After Review

13/WM/0172	**Please quote this number on all correspondence**

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

With the Committee's best wishes for the success of this project.

Yours sincerely

pp

Dr Rex J Polson
Chair

Email: NRESCommittee.WestMidlands.Solihull@nhs.net

Enclosures: *List of names and professions of members who took part in the review*

"After ethical review – guidance for researchers" [SL-AR2]

Copy to: *Mrs. Nancy E. Brown Connolly*
N/A. R&D contact not specified in database.

NRES Committee West Midlands - Solihull

Attendance at PRS Sub-Committee of the REC meeting on 10 April 2013

Committee Members:

Name	*Profession*	*Present*	*Notes*
Mrs Irene Linder	Assistant Manager, Local Authority - Retired	Yes	
Dr Rex J Polson	Consultant Physician - Chair	Yes	
Dr Timothy Priest	Consultant in Anaesthesia & Pain Management - Vice Chair	Yes	

Also in attendance:

Name	*Position (or reason for attending)*
Ms Maria Morledge	
Ms Trish Wheat	Committee Coordinator

APPENDIX C

Ms N Brown Connolly
5907 Irving Boulevard NW
Albuquerque
NM 87114

Research, Service Evaluation and
Clinical Audit Office
NHS Direct
Berrywood Business Village
Tollbar Way
Hedge End
Southampton
SO30 2UN

Telephone: 0208 676 3211
Fax: 01489 773721
E-mail: shirley.large@nhsdirect.nhs.uk

16 April 2013

Dear Nancy

Re: Receiver Operating Characteristic (ROC) analysis of a remote monitoring disease management system to determine optimum threshold and predictive value
CI: Nancy Brown Connolly
Sponsor: Brunel University
Ref no: NHS Direct Ref 137

NHS permission for the above research has been granted on the basis described in the application form, protocol and supporting documentation. The documents reviewed were:

- NHS Direct Research Governance Approval Form
- Potential addendum
- NBC Resume
- Brunel University ethics approval letter
- REC Permission letter

Permission is granted on the understanding that the study is conducted in accordance with the Research Governance Framework and NHS Direct policies and procedures.

Permission is only granted for the activities for which a favourable opinion has been given by the REC and for those listed in the application.

Please note that it is a condition of this permission that you should, where applicable, provide the following notifications:

- Any deviation from the study protocol; this includes any delay or deviation from the proposed start and end dates of your study, which we confirm to be March 2013 and December 2013.
- Any research Adverse Incident.
- Six-monthly research progress reports.
- The final report.
- Any dissemination arising from the research in the form of conference presentations and academic papers prior to presentation or publication.

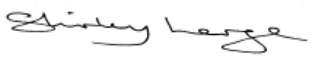
In addition, the research sponsor or the Chief Investigator may take appropriate urgent safety measures in order to protect research participants against any immediate hazard to their health or safety.
The R&D office should be notified that such measures have been taken. The notification should also include the reasons why the measures were taken and the plan for further action.
The R&D Office should be notified within the same time frame of notifying the REC and any other regulatory bodies.

All amendments need to be submitted in accordance with guidance in IRAS.

Please note that NHS Direct is required to monitor research to ensure compliance with the Research Governance Framework and other legal and regulatory requirements. This is achieved by random audit of research.

I wish you every success with your research.

Yours sincerely



Dr Shirley Large CSci
Head of Research and Clinical Audit

Cc Dr Malcolm Clarke

APPENDIX D

Population and resource usage

ID	Monitored Days	# Home Visits	Pre A&E	# A&E Monitored	Δ A&E	Pre IP Days	# IP Days Monitored	Δ IP
1. .	454	88	1	1	-0	10	6	-4
2.	194	11	5	3	-2	45	19	-26
3.	440	35	1	0	-1	13	0	-13
4.	439	102	2	1	-1	2	1	-1
5.	439	76	3	1	-2	23	3	-20
6.	439	41	4	1	-3	56	7	-49
7.	351	126	5	0	-5	27	0	-27
8.	351	86	1	1	-0	10	3	-7
9.	348	56	1	0	-1	1	0	-1
10.	330	95	3	2	-1	24	14	-10
11.	319	62	0	0	0	12	1	-11
12.	315	81	2	3	+1	26	0	-26
13.	315	78	3	0	-3	19	15	-4
14.	314	50	1	0	-1	4	0	-4
15.	311	115	7	1	-6	61	7	-54
16.	294	47	1	0	-1	59	0	-59
17.	293	12	5	0	-5	11	0	-11
18.	292	58	1	1	0	4	3	-1
19.	279	20	1	1	0	4	6	+2
20.	273	21	3	0	-3	27	0	-27
21.	272	77	3	1	-2	12	4	-8
22.	271	14	1	0	-1	9	0	-9
23.	259	39	2	0	-2	8	0	-8
24.	251	6	3	0	-3	3	0	-3
25.	245	23	1	4	+3	16	80	+64
26.	244	15	1	0	-1	2	0	-2
27.	222	22	2	0	-2	8	0	-8
28.	210	53	11	8	-3	290	154	-136
29.	196	25	5	0	-5	29	0	-29
30.	195	69	1	6	+5	3	45	+42
31.	195	49	1	0	-1	13	2	-11
32.	190	34	2	2	0	16	32	+16
33.	189	23	2	4	+2	12	16	+4
34.	183	25	1	1	0	5	9	+4
TOTAL		**1734**	**86**	**42**	**-44**	**864**	**427**	**-437**

APPENDIX E

Criterion Values - SpO_2

Criterion SpO_2	Sensitivity	Specificity	+LR	-LR	+PV	95% CI	-PV	95% CI	Cost
≥0	100.00	0.00	1.00		10.3	9.6 - 11.1			-0.0792
>0	100.00	0.017	1.00	0.00	10.3	9.6 - 11.1	100.0	2.5 - 100.0	-0.0792
>1	100.00	0.051	1.00	0.00	10.3	9.6 - 11.1	100.0	29.2 - 100.0	-0.0792
>2	99.85	0.31	1.00	0.48	10.3	9.6 - 11.1	94.7	73.1 - 99.9	-0.0792
>3	99.26	0.90	1.00	0.82	10.3	9.6 - 11.1	91.4	80.9 - 97.2	-0.0788
>4	98.37	3.77	1.02	0.43	10.5	9.8 - 11.3	95.3	91.7 - 97.6	-0.0783
>5	95.55	11.85	1.08	0.38	11.1	10.3 - 11.9	95.9	94.1 - 97.2	-0.0767
>6	92.58	22.46	1.19	0.33	12.1	11.2 - 13.0	96.3	95.2 - 97.3	-0.0749
>7	86.80	34.32	1.32	0.38	13.2	12.2 - 14.2	95.8	94.8 - 96.6	-0.0717
>8	81.01	47.76	1.55	0.40	15.1	14.0 - 16.3	95.6	94.8 - 96.3	-0.0684
>9	69.73	57.98	1.66	0.52	16.0	14.7 - 17.4	94.3	93.5 - 95.1	-0.0625
>10	56.82	71.16	1.97	0.61	18.5	16.8 - 20.2	93.5	92.7 - 94.2	-0.0556
>11	44.81	78.22	2.06	0.71	19.1	17.2 - 21.2	92.5	91.7 - 93.2	-0.0494
>12	38.13	84.45	2.45	0.73	22.0	19.7 - 24.5	92.2	91.5 - 92.9	-0.0459
>13	30.12	89.52	2.87	0.78	24.8	21.9 - 28.0	91.8	91.0 - 92.5	-0.0417
>14	23.89	92.49	3.18	0.82	26.8	23.3 - 30.5	91.4	90.6 - 92.1	-0.0385
>15	21.22	94.49	3.85	0.83	30.7	26.5 - 35.1	91.2	90.5 - 91.9	-0.0371
>16	16.02	95.56	3.61	0.88	29.3	24.7 - 34.3	90.8	90.1 - 91.5	-0.0345
>17	13.50	96.57	3.94	0.90	31.2	25.9 - 36.8	90.7	89.9 - 91.4	-0.0332
>18	10.39	97.71	4.54	0.92	34.3	27.8 - 41.3	90.5	89.7 - 91.2	-0.0316
>19	8.61	98.48	5.67	0.93	39.5	31.5 - 47.9	90.4	89.6 - 91.1	-0.0307
>20	6.23	98.99	6.19	0.95	41.6	31.9 - 51.8	90.2	89.4 - 90.9	-0.0295
>21	5.64	99.04	5.90	0.95	40.4	30.4 - 51.0	90.1	89.4 - 90.8	-0.0292
>22	5.34	99.10	5.90	0.96	40.4	30.1 - 51.4	90.1	89.3 - 90.8	-0.0290
>23	4.90	99.11	5.52	0.96	38.8	28.4 - 50.1	90.1	89.3 - 90.8	-0.0288
>24	4.60	99.13	5.28	0.96	37.8	27.3 - 49.3	90.0	89.3 - 90.8	-0.0286
>25	4.30	99.20	5.36	0.96	38.2	27.2 - 50.0	90.0	89.3 - 90.7	-0.0285
>28	4.15	99.20	5.18	0.97	37.3	26.4 - 49.3	90.0	89.2 - 90.7	-0.0284
>30	3.86	99.74	15.07	0.96	63.4	46.7 - 78.0	90.0	89.3 - 90.7	-0.0283
>32	3.86	99.76	16.14	0.96	65.0	48.1 - 79.5	90.0	89.3 - 90.7	-0.0283
>41	3.71	99.76	15.52	0.97	64.1	46.9 - 79.0	90.0	89.3 - 90.7	-0.0282
>100	0.00	100.00		1.00			89.7	88.9 - 90.4	-0.0263

APPENDIX F

Criterion values - SBP

Criterion SBP	Sensitivity	Specificity	+LR	-LR	+PV	95% CI	-PV	95% CI	Cost
<-33	0.00	100.00		1.00			85.8	85.1 - 86.5	-0.0361
≤-33	0.00	99.99	0.00	1.00	0.0	0.0 - 97.5	85.8	85.1 - 86.5	-0.0361
≤-19	0.00	99.98	0.00	1.00	0.0	0.0 - 97.5	85.8	85.1 - 86.5	-0.0361
≤-13	0.00	99.97	0.00	1.00	0.0	0.0 - 70.8	85.8	85.1 - 86.5	-0.0361
≤-12	0.00	99.94	0.00	1.00	0.0	0.0 - 52.2	85.8	85.1 - 86.5	-0.0361
≤-10	0.070	99.93	1.01	1.00	14.3	0.2 - 61.9	85.8	85.1 - 86.5	-0.0362
≤-8	0.14	99.91	1.51	1.00	20.0	2.5 - 55.6	85.8	85.1 - 86.5	-0.0362
≤-7	0.14	99.88	1.21	1.00	16.7	2.1 - 48.4	85.8	85.1 - 86.5	-0.0362
≤-5	0.14	99.87	1.10	1.00	15.4	1.7 - 47.0	85.8	85.1 - 86.5	-0.0362
≤-4	0.14	99.85	0.93	1.00	13.3	1.7 - 40.5	85.8	85.1 - 86.5	-0.0362
≤-3	0.14	99.84	0.87	1.00	12.5	1.4 - 39.5	85.8	85.1 - 86.5	-0.0362
≤-2	0.21	99.81	1.14	1.00	15.8	3.2 - 40.4	85.8	85.1 - 86.5	-0.0363
≤-1	0.21	99.78	0.96	1.00	13.6	2.8 - 35.6	85.8	85.1 - 86.5	-0.0363
≤0	0.21	99.72	0.76	1.00	11.1	2.4 - 29.2	85.8	85.1 - 86.5	-0.0363
≤1	0.21	99.70	0.70	1.00	10.3	2.1 - 27.8	85.8	85.1 - 86.5	-0.0363
≤2	0.28	99.63	0.76	1.00	11.1	3.1 - 26.1	85.8	85.1 - 86.5	-0.0363
≤3	0.28	99.61	0.71	1.00	10.5	2.9 - 24.8	85.8	85.1 - 86.5	-0.0363
≤4	0.28	99.56	0.64	1.00	9.5	2.6 - 22.8	85.8	85.1 - 86.5	-0.0363
≤5	0.28	99.55	0.62	1.00	9.3	2.6 - 22.1	85.8	85.1 - 86.5	-0.0363
≤6	0.35	99.49	0.69	1.00	10.2	3.4 - 22.2	85.8	85.1 - 86.5	-0.0364
≤7	0.35	99.45	0.64	1.00	9.6	3.2 - 21.2	85.8	85.1 - 86.5	-0.0364
≤8	0.49	99.36	0.77	1.00	11.3	4.6 - 22.0	85.8	85.1 - 86.5	-0.0365
≤9	0.49	99.34	0.74	1.00	10.9	4.5 - 21.3	85.8	85.1 - 86.5	-0.0365
≤10	0.63	99.30	0.91	1.00	13.0	6.1 - 23.3	85.8	85.1 - 86.5	-0.0366
≤11	0.63	99.27	0.87	1.00	12.5	5.9 - 22.4	85.8	85.1 - 86.5	-0.0366
≤12	0.77	99.18	0.94	1.00	13.4	6.9 - 22.8	85.8	85.1 - 86.5	-0.0367
≤13	0.91	99.12	1.04	1.00	14.6	8.0 - 23.7	85.8	85.1 - 86.5	-0.0368
≤14	0.98	99.08	1.07	1.00	15.1	8.5 - 24.0	85.8	85.1 - 86.5	-0.0368
≤15	1.05	99.04	1.10	1.00	15.3	8.8 - 24.0	85.8	85.1 - 86.5	-0.0369
≤16	1.55	98.91	1.42	1.00	19.0	12.3 - 27.3	85.9	85.2 - 86.6	-0.0372
≤17	1.69	98.87	1.50	0.99	19.8	13.1 - 28.1	85.9	85.2 - 86.6	-0.0373
≤18	1.76	98.82	1.49	0.99	19.7	13.1 - 27.7	85.9	85.2 - 86.6	-0.0374
≤19	1.90	98.76	1.53	0.99	20.1	13.7 - 28.0	85.9	85.2 - 86.6	-0.0375
≤20	2.11	98.61	1.51	0.99	20.0	13.9 - 27.3	85.9	85.2 - 86.6	-0.0376
≤21	2.18	98.54	1.49	0.99	19.7	13.8 - 26.9	85.9	85.2 - 86.6	-0.0377
≤22	2.32	98.44	1.49	0.99	19.8	14.0 - 26.6	85.9	85.2 - 86.6	-0.0378
≤23	2.53	98.32	1.50	0.99	19.9	14.3 - 26.5	85.9	85.2 - 86.6	-0.0379
≤24	3.02	97.96	1.48	0.99	19.6	14.6 - 25.5	86.0	85.3 - 86.6	-0.0383
≤25	3.45	97.78	1.55	0.99	20.4	15.5 - 26.1	86.0	85.3 - 86.7	-0.0386
≤26	3.59	97.62	1.51	0.99	19.9	15.2 - 25.4	86.0	85.3 - 86.7	-0.0387
≤27	4.01	97.32	1.50	0.99	19.8	15.3 - 24.9	86.0	85.3 - 86.7	-0.0390
≤28	4.64	96.81	1.45	0.99	19.4	15.3 - 24.0	86.0	85.3 - 86.7	-0.0394
≤29	4.99	96.46	1.41	0.98	18.9	15.1 - 23.2	86.0	85.3 - 86.7	-0.0397
≤30	5.34	96.24	1.42	0.98	19.0	15.3 - 23.2	86.0	85.3 - 86.7	-0.0399

Criterion SBP	Sensitivity	Specificity	+LR	-LR	+PV	95% CI	-PV	95% CI	Cost
≤31	5.84	95.74	1.37	0.98	18.4	15.0 - 22.3	86.0	85.3 - 86.7	-0.0403
≤32	7.52	94.85	1.46	0.97	19.4	16.2 - 23.0	86.1	85.4 - 86.8	-0.0414
≤33	7.95	94.43	1.43	0.97	19.1	16.0 - 22.5	86.1	85.4 - 86.8	-0.0417
≤34	8.51	93.99	1.42	0.97	18.9	16.0 - 22.2	86.2	85.4 - 86.9	-0.0421
≤35	9.42	93.49	1.45	0.97	19.3	16.4 - 22.4	86.2	85.5 - 86.9	-0.0428
≤36	10.34	92.64	1.41	0.97	18.8	16.1 - 21.7	86.2	85.5 - 86.9	-0.0434
≤37	11.46	92.13	1.46	0.96	19.4	16.8 - 22.2	86.3	85.6 - 87.0	-0.0442
≤38	11.81	91.71	1.43	0.96	19.0	16.5 - 21.8	86.3	85.6 - 87.0	-0.0445
≤39	12.31	91.24	1.40	0.96	18.8	16.4 - 21.5	86.3	85.6 - 87.0	-0.0448
≤40	13.43	90.12	1.36	0.96	18.3	16.0 - 20.8	86.3	85.6 - 87.0	-0.0456
≤41	14.35	89.49	1.36	0.96	18.4	16.1 - 20.8	86.4	85.6 - 87.1	-0.0463
≤42	15.40	88.79	1.37	0.95	18.5	16.3 - 20.8	86.4	85.7 - 87.1	-0.0470
≤43	16.24	88.24	1.38	0.95	18.6	16.4 - 20.8	86.5	85.7 - 87.2	-0.0476
≤44	17.44	87.06	1.35	0.95	18.2	16.2 - 20.3	86.5	85.7 - 87.2	-0.0485
≤45	18.71	86.31	1.37	0.94	18.4	16.4 - 20.5	86.5	85.8 - 87.3	-0.0494
≤46	19.34	85.40	1.32	0.94	17.9	16.0 - 20.0	86.5	85.8 - 87.2	-0.0498
≤47	20.46	84.74	1.34	0.94	18.1	16.3 - 20.1	86.6	85.8 - 87.3	-0.0506
≤48	22.22	83.13	1.32	0.94	17.9	16.1 - 19.7	86.6	85.9 - 87.4	-0.0519
≤49	22.86	82.29	1.29	0.94	17.6	15.9 - 19.4	86.6	85.8 - 87.3	-0.0523
≤50	24.12	81.47	1.30	0.93	17.7	16.0 - 19.5	86.7	85.9 - 87.4	-0.0532
≤51	25.18	80.76	1.31	0.93	17.8	16.1 - 19.5	86.7	86.0 - 87.5	-0.0540
≤52	27.50	79.22	1.32	0.92	17.9	16.3 - 19.6	86.9	86.1 - 87.6	-0.0556
≤53	28.90	78.38	1.34	0.91	18.1	16.5 - 19.7	87.0	86.2 - 87.7	-0.0566
≤54	29.89	77.24	1.31	0.91	17.8	16.3 - 19.4	87.0	86.2 - 87.7	-0.0573
≤55	31.08	76.52	1.32	0.90	17.9	16.4 - 19.5	87.1	86.3 - 87.8	-0.0582
≤56	32.56	74.76	1.29	0.90	17.6	16.1 - 19.1	87.0	86.3 - 87.8	-0.0592
≤57	33.83	73.73	1.29	0.90	17.5	16.1 - 19.0	87.1	86.3 - 87.9	-0.0601
≤58	34.74	72.59	1.27	0.90	17.3	15.9 - 18.7	87.1	86.3 - 87.8	-0.0608
≤59	35.44	71.63	1.25	0.90	17.1	15.7 - 18.5	87.1	86.2 - 87.8	-0.0613
≤60	37.90	69.65	1.25	0.89	17.1	15.8 - 18.4	87.2	86.4 - 88.0	-0.0630
≤61	38.75	68.76	1.24	0.89	17.0	15.7 - 18.3	87.2	86.4 - 88.0	-0.0637
≤62	40.23	67.80	1.25	0.88	17.1	15.8 - 18.4	87.3	86.5 - 88.1	-0.0647
≤63	41.28	66.69	1.24	0.88	17.0	15.7 - 18.3	87.3	86.5 - 88.1	-0.0655
≤64	43.25	64.47	1.22	0.88	16.7	15.5 - 18.0	87.3	86.5 - 88.1	-0.0669
≤65	43.88	63.19	1.19	0.89	16.4	15.3 - 17.7	87.2	86.4 - 88.0	-0.0673
≤66	44.73	61.99	1.18	0.89	16.3	15.1 - 17.5	87.2	86.3 - 88.0	-0.0680
≤67	45.85	60.71	1.17	0.89	16.1	15.0 - 17.3	87.2	86.3 - 88.0	-0.0688
≤68	48.80	58.31	1.17	0.88	16.2	15.1 - 17.3	87.3	86.5 - 88.2	-0.0709
≤69	49.58	56.77	1.15	0.89	15.9	14.8 - 17.0	87.2	86.3 - 88.1	-0.0714
≤70	50.91	55.30	1.14	0.89	15.8	14.8 - 16.9	87.2	86.3 - 88.1	-0.0724
≤71	51.76	54.07	1.13	0.89	15.7	14.7 - 16.8	87.2	86.2 - 88.1	-0.0730
≤72	53.73	51.00	1.10	0.91	15.3	14.3 - 16.4	87.0	86.0 - 87.9	-0.0744
≤73	55.27	49.63	1.10	0.90	15.3	14.4 - 16.3	87.1	86.1 - 88.0	-0.0755
≤74	56.26	48.15	1.09	0.91	15.2	14.2 - 16.2	87.0	86.0 - 87.9	-0.0763
≤75	58.02	46.64	1.09	0.90	15.2	14.3 - 16.2	87.1	86.1 - 88.0	-0.0775
≤76	61.18	43.40	1.08	0.89	15.1	14.2 - 16.1	87.1	86.1 - 88.1	-0.0798
≤77	62.45	41.94	1.08	0.90	15.1	14.2 - 16.0	87.1	86.1 - 88.1	-0.0807

Criterion SBP	Sensitivity	Specificity	+LR	-LR	+PV	95% CI	-PV	95% CI	Cost
≤78	63.71	40.48	1.07	0.90	15.0	14.1 - 15.9	87.1	86.0 - 88.1	-0.0816
≤79	65.12	39.17	1.07	0.89	15.0	14.1 - 15.9	87.2	86.1 - 88.2	-0.0826
≤80	68.00	36.47	1.07	0.88	15.0	14.1 - 15.9	87.4	86.2 - 88.4	-0.0847
≤81	69.20	35.08	1.07	0.88	15.0	14.1 - 15.8	87.3	86.2 - 88.4	-0.0855
≤82	70.25	33.63	1.06	0.88	14.9	14.0 - 15.7	87.3	86.1 - 88.4	-0.0863
≤83	71.59	32.18	1.06	0.88	14.8	14.0 - 15.7	87.3	86.1 - 88.4	-0.0873
≤84	73.28	29.85	1.04	0.90	14.7	13.9 - 15.5	87.1	85.9 - 88.3	-0.0885
≤85	74.33	28.66	1.04	0.90	14.7	13.9 - 15.5	87.1	85.8 - 88.3	-0.0892
≤86	75.74	27.41	1.04	0.89	14.7	13.9 - 15.5	87.3	85.9 - 88.5	-0.0902
≤87	76.51	26.11	1.04	0.90	14.6	13.8 - 15.4	87.1	85.7 - 88.3	-0.0908
≤88	78.97	23.59	1.03	0.89	14.6	13.8 - 15.4	87.2	85.8 - 88.5	-0.0926
≤89	79.96	22.30	1.03	0.90	14.5	13.7 - 15.3	87.1	85.6 - 88.5	-0.0933
≤90	80.73	21.14	1.02	0.91	14.5	13.7 - 15.2	86.9	85.4 - 88.3	-0.0938
≤91	82.14	19.65	1.02	0.91	14.4	13.7 - 15.2	87.0	85.4 - 88.4	-0.0948
≤92	84.04	17.42	1.02	0.92	14.4	13.6 - 15.2	86.9	85.2 - 88.4	-0.0962
≤93	85.16	16.41	1.02	0.90	14.4	13.6 - 15.2	87.0	85.3 - 88.6	-0.0970
≤94	85.72	15.57	1.02	0.92	14.4	13.6 - 15.1	86.9	85.1 - 88.5	-0.0974
≤95	86.50	14.63	1.01	0.92	14.3	13.6 - 15.1	86.8	84.9 - 88.5	-0.0980
≤96	88.33	12.64	1.01	0.92	14.3	13.6 - 15.0	86.8	84.8 - 88.6	-0.0993
≤97	88.89	11.83	1.01	0.94	14.3	13.5 - 15.0	86.6	84.5 - 88.5	-0.0997
≤98	89.73	11.04	1.01	0.93	14.3	13.6 - 15.0	86.7	84.5 - 88.6	-0.100
≤99	90.58	10.29	1.01	0.92	14.3	13.6 - 15.0	86.9	84.6 - 88.9	-0.101
≤100	91.77	8.89	1.01	0.93	14.3	13.5 - 15.0	86.7	84.3 - 88.9	-0.102
≤101	91.98	8.10	1.00	0.99	14.2	13.5 - 14.9	86.0	83.4 - 88.3	-0.102
≤102	92.41	7.42	1.00	1.02	14.1	13.4 - 14.9	85.5	82.8 - 88.0	-0.102
≤103	92.69	6.87	1.00	1.06	14.1	13.4 - 14.8	85.1	82.2 - 87.6	-0.102
≤104	93.46	5.91	0.99	1.11	14.1	13.4 - 14.8	84.6	81.4 - 87.3	-0.103
≤105	93.95	5.47	0.99	1.11	14.1	13.4 - 14.8	84.6	81.3 - 87.5	-0.103
≤106	94.37	4.93	0.99	1.14	14.1	13.4 - 14.8	84.2	80.7 - 87.2	-0.104
≤107	94.87	4.62	0.99	1.11	14.1	13.4 - 14.8	84.5	80.9 - 87.7	-0.104
≤108	95.64	3.70	0.99	1.18	14.1	13.4 - 14.8	83.7	79.6 - 87.3	-0.105
≤109	96.06	3.37	0.99	1.17	14.1	13.4 - 14.8	83.8	79.5 - 87.5	-0.105
≤110	96.55	2.98	1.00	1.16	14.1	13.4 - 14.8	84.0	79.4 - 87.9	-0.105
≤111	96.62	2.63	0.99	1.28	14.1	13.4 - 14.8	82.5	77.5 - 86.8	-0.105
≤112	96.91	2.14	0.99	1.45	14.0	13.4 - 14.7	80.7	75.0 - 85.6	-0.105
≤113	97.26	1.93	0.99	1.42	14.1	13.4 - 14.8	81.0	74.9 - 86.1	-0.106
≤114	97.61	1.75	0.99	1.36	14.1	13.4 - 14.8	81.6	75.3 - 86.9	-0.106
≤115	97.68	1.59	0.99	1.46	14.1	13.4 - 14.8	80.6	73.8 - 86.2	-0.106
≤116	98.17	1.32	0.99	1.38	14.1	13.4 - 14.8	81.4	74.0 - 87.5	-0.106
≤117	98.38	1.14	1.00	1.42	14.1	13.4 - 14.8	81.0	72.8 - 87.6	-0.107
≤118	98.45	1.07	1.00	1.45	14.1	13.4 - 14.8	80.7	72.2 - 87.5	-0.107
≤119	98.73	0.96	1.00	1.31	14.1	13.4 - 14.8	82.2	73.3 - 89.1	-0.107
≤120	98.95	0.70	1.00	1.51	14.1	13.4 - 14.8	80.0	69.1 - 88.4	-0.107
≤121	99.16	0.59	1.00	1.43	14.1	13.5 - 14.8	81.0	69.1 - 89.8	-0.107
≤122	99.23	0.55	1.00	1.42	14.1	13.5 - 14.8	81.0	68.5 - 90.2	-0.107
≤123	99.44	0.48	1.00	1.18	14.2	13.5 - 14.9	83.7	70.3 - 92.7	-0.107
≤124	99.51	0.41	1.00	1.21	14.2	13.5 - 14.9	83.3	68.4 - 93.1	-0.107

Criterion SBP	Sensitivity	Specificity	+LR	-LR	+PV	95% CI	-PV	95% CI	Cost
≤125	99.58	0.32	1.00	1.30	14.2	13.5 - 14.9	82.4	65.2 - 93.4	-0.107
≤126	99.58	0.29	1.00	1.45	14.1	13.5 - 14.8	80.6	62.2 - 92.7	-0.107
≤127	99.72	0.26	1.00	1.10	14.2	13.5 - 14.9	84.6	65.1 - 95.6	-0.107
≤128	99.79	0.19	1.00	1.14	14.2	13.5 - 14.9	84.2	59.6 - 96.8	-0.108
≤129	99.79	0.17	1.00	1.21	14.2	13.5 - 14.9	83.3	57.7 - 96.6	-0.108
≤130	99.86	0.16	1.00	0.87	14.2	13.5 - 14.9	87.5	60.5 - 98.6	-0.108
≤131	99.86	0.14	1.00	1.01	14.2	13.5 - 14.9	85.7	55.8 - 98.4	-0.108
≤132	99.93	0.13	1.00	0.55	14.2	13.5 - 14.9	91.7	61.5 - 99.8	-0.108
≤133	99.93	0.12	1.00	0.61	14.2	13.5 - 14.9	90.9	58.7 - 99.8	-0.108
≤135	99.93	0.10	1.00	0.67	14.2	13.5 - 14.9	90.0	55.5 - 99.7	-0.108
≤136	99.93	0.093	1.00	0.76	14.2	13.5 - 14.9	88.9	51.8 - 99.7	-0.108
≤140	99.93	0.081	1.00	0.87	14.2	13.5 - 14.9	87.5	47.3 - 99.7	-0.108
≤142	99.93	0.070	1.00	1.01	14.2	13.5 - 14.9	85.7	42.1 - 99.6	-0.108
≤143	99.93	0.058	1.00	1.21	14.2	13.5 - 14.9	83.3	35.9 - 99.6	-0.108
≤146	100.00	0.046	1.00	0.00	14.2	13.5 - 14.9	100.0	39.8 - 100.0	-0.108
≤150	100.00	0.035	1.00	0.00	14.2	13.5 - 14.9	100.0	29.2 - 100.0	-0.108
≤155	100.00	0.023	1.00	0.00	14.2	13.5 - 14.9	100.0	15.8 - 100.0	-0.108
≤159	100.00	0.012	1.00	0.00	14.2	13.5 - 14.9	100.0	2.5 - 100.0	-0.108
≤162	100.00	0.00	1.00		14.2	13.5 - 14.9			-0.108

APPENDIX G

Criterion values - DBP

Criterion DBP	Sensitivity	Specificity	+LR	-LR	+PV	95% CI	-PV	95% CI
<-544	0.00	100.00		1.00			86.6	85.9 - 87.3
≤-544	0.074	100.00		1.00	100.0	0.0 - 100.0	86.6	85.9 - 87.3
≤2	0.074	99.99	6.47	1.00	50.0	0.04 - 100.0	86.6	85.9 - 87.3
≤10	0.074	99.97	2.16	1.00	25.0	0.2 - 86.8	86.6	85.9 - 87.3
≤11	0.074	99.95	1.62	1.00	20.0	0.5 - 71.6	86.6	85.9 - 87.3
≤17	0.074	99.92	0.92	1.00	12.5	0.2 - 56.0	86.6	85.9 - 87.3
≤18	0.074	99.90	0.72	1.00	10.0	0.2 - 47.0	86.6	85.9 - 87.3
≤20	0.074	99.88	0.65	1.00	9.1	0.2 - 43.4	86.6	85.9 - 87.3
≤22	0.074	99.87	0.59	1.00	8.3	0.2 - 38.5	86.6	85.9 - 87.3
≤24	0.074	99.85	0.50	1.00	7.1	0.2 - 33.9	86.6	85.9 - 87.3
≤26	0.15	99.83	0.86	1.00	11.8	1.5 - 36.4	86.6	85.9 - 87.3
≤27	0.22	99.80	1.14	1.00	15.0	3.2 - 37.9	86.6	85.9 - 87.3
≤28	0.22	99.79	1.08	1.00	14.3	2.9 - 37.0	86.6	85.9 - 87.3
≤30	0.22	99.78	1.02	1.00	13.6	2.9 - 34.9	86.6	85.9 - 87.3
≤31	0.22	99.77	0.97	1.00	13.0	2.8 - 33.6	86.6	85.9 - 87.3
≤34	0.22	99.76	0.92	1.00	12.5	2.5 - 32.9	86.6	85.9 - 87.3
≤36	0.30	99.72	1.08	1.00	14.3	4.0 - 32.7	86.6	85.9 - 87.3
≤37	0.37	99.71	1.29	1.00	16.7	5.5 - 35.1	86.6	85.9 - 87.3
≤39	0.37	99.69	1.20	1.00	15.6	5.2 - 33.1	86.6	85.9 - 87.3
≤40	0.37	99.68	1.15	1.00	15.2	5.1 - 31.9	86.6	85.9 - 87.3
≤41	0.37	99.67	1.11	1.00	14.7	4.8 - 31.4	86.6	85.9 - 87.3
≤42	0.37	99.64	1.04	1.00	13.9	4.7 - 29.5	86.6	85.9 - 87.3
≤43	0.37	99.63	1.01	1.00	13.5	4.4 - 29.0	86.6	85.9 - 87.3
≤45	0.37	99.62	0.98	1.00	13.2	4.4 - 28.1	86.6	85.9 - 87.3
≤46	0.37	99.60	0.92	1.00	12.5	4.2 - 26.8	86.6	85.9 - 87.3
≤47	0.37	99.57	0.87	1.00	11.9	4.0 - 25.6	86.6	85.9 - 87.3
≤48	0.45	99.54	0.97	1.00	13.0	4.9 - 26.3	86.6	85.9 - 87.3
≤50	0.45	99.53	0.95	1.00	12.8	4.8 - 25.9	86.6	85.9 - 87.3
≤51	0.45	99.48	0.86	1.00	11.8	4.4 - 24.0	86.6	85.9 - 87.3
≤52	0.52	99.42	0.91	1.00	12.3	5.0 - 23.8	86.6	85.9 - 87.3
≤54	0.52	99.41	0.89	1.00	12.1	5.0 - 23.3	86.6	85.9 - 87.3
≤55	0.52	99.40	0.87	1.00	11.9	4.9 - 23.0	86.6	85.9 - 87.3
≤56	0.60	99.33	0.89	1.00	12.1	5.3 - 22.6	86.6	85.9 - 87.3
≤57	0.60	99.31	0.86	1.00	11.8	5.2 - 21.9	86.6	85.9 - 87.3
≤58	0.60	99.29	0.83	1.00	11.4	5.1 - 21.3	86.6	85.9 - 87.3
≤59	0.60	99.26	0.81	1.00	11.1	4.9 - 20.7	86.6	85.9 - 87.3
≤60	0.67	99.26	0.91	1.00	12.3	5.8 - 22.1	86.6	85.9 - 87.3
≤61	0.74	99.24	0.98	1.00	13.2	6.5 - 22.9	86.6	85.9 - 87.3
≤63	0.82	99.21	1.03	1.00	13.8	7.1 - 23.3	86.6	85.9 - 87.3
≤64	0.82	99.17	0.99	1.00	13.3	6.8 - 22.5	86.6	85.9 - 87.3
≤65	0.89	99.14	1.03	1.00	13.8	7.3 - 22.9	86.6	85.9 - 87.3
≤66	0.97	99.13	1.11	1.00	14.6	8.0 - 23.7	86.6	85.9 - 87.3
≤67	0.97	99.10	1.08	1.00	14.3	7.8 - 23.2	86.6	85.9 - 87.3
≤68	0.97	99.04	1.01	1.00	13.5	7.4 - 22.1	86.6	85.9 - 87.3

Criterion DBP	Sensitivity	Specificity	+LR	-LR	+PV	95% CI	-PV	95% CI
≤69	1.04	99.02	1.06	1.00	14.1	7.9 - 22.6	86.6	85.9 - 87.3
≤70	1.04	98.99	1.03	1.00	13.7	7.7 - 22.0	86.6	85.9 - 87.3
≤71	1.19	98.95	1.14	1.00	15.0	8.8 - 23.1	86.6	85.9 - 87.3
≤72	1.19	98.88	1.07	1.00	14.2	8.3 - 22.0	86.6	85.9 - 87.3
≤74	1.34	98.83	1.14	1.00	15.0	9.1 - 22.7	86.6	85.9 - 87.3
≤75	1.34	98.80	1.12	1.00	14.8	9.0 - 22.3	86.6	85.9 - 87.3
≤76	1.64	98.76	1.32	1.00	16.9	10.9 - 24.5	86.7	86.0 - 87.3
≤77	1.71	98.71	1.33	1.00	17.0	11.1 - 24.5	86.7	86.0 - 87.3
≤78	1.86	98.64	1.37	0.99	17.5	11.6 - 24.7	86.7	86.0 - 87.3
≤79	1.93	98.62	1.40	0.99	17.8	12.0 - 25.0	86.7	86.0 - 87.3
≤80	2.08	98.49	1.38	0.99	17.6	12.0 - 24.4	86.7	86.0 - 87.3
≤81	2.08	98.45	1.34	0.99	17.2	11.7 - 23.9	86.7	86.0 - 87.3
≤82	2.08	98.40	1.30	1.00	16.8	11.4 - 23.3	86.7	86.0 - 87.3
≤83	2.23	98.34	1.35	0.99	17.2	11.9 - 23.7	86.7	86.0 - 87.3
≤84	2.38	98.25	1.36	0.99	17.4	12.2 - 23.7	86.7	86.0 - 87.3
≤85	2.46	98.19	1.36	0.99	17.4	12.3 - 23.5	86.7	86.0 - 87.3
≤86	2.46	98.15	1.33	0.99	17.0	12.0 - 23.1	86.7	86.0 - 87.3
≤87	2.68	98.08	1.39	0.99	17.7	12.7 - 23.7	86.7	86.0 - 87.4
≤88	2.75	97.92	1.32	0.99	17.0	12.2 - 22.6	86.7	86.0 - 87.4
≤89	2.83	97.85	1.31	0.99	16.9	12.2 - 22.4	86.7	86.0 - 87.4
≤90	2.98	97.74	1.32	0.99	16.9	12.4 - 22.4	86.7	86.0 - 87.4
≤91	3.12	97.62	1.31	0.99	16.9	12.4 - 22.1	86.7	86.0 - 87.4
≤92	3.57	97.26	1.30	0.99	16.8	12.6 - 21.6	86.7	86.0 - 87.4
≤93	3.79	97.15	1.33	0.99	17.1	13.0 - 21.8	86.7	86.0 - 87.4
≤94	4.02	96.93	1.31	0.99	16.8	12.9 - 21.4	86.7	86.0 - 87.4
≤95	4.24	96.70	1.28	0.99	16.6	12.8 - 20.9	86.7	86.0 - 87.4
≤96	5.43	95.94	1.34	0.99	17.1	13.7 - 21.1	86.8	86.1 - 87.4
≤97	6.32	95.64	1.45	0.98	18.3	14.9 - 22.1	86.8	86.2 - 87.5
≤98	6.62	95.36	1.43	0.98	18.1	14.8 - 21.8	86.8	86.2 - 87.5
≤99	7.22	94.93	1.42	0.98	18.0	14.9 - 21.5	86.9	86.2 - 87.5
≤100	7.89	94.19	1.36	0.98	17.3	14.4 - 20.6	86.9	86.2 - 87.5
≤101	8.26	93.76	1.32	0.98	17.0	14.2 - 20.1	86.9	86.2 - 87.5
≤102	8.71	93.28	1.30	0.98	16.7	14.0 - 19.7	86.9	86.2 - 87.5
≤103	9.45	92.76	1.31	0.98	16.8	14.2 - 19.7	86.9	86.2 - 87.6
≤104	10.79	91.71	1.30	0.97	16.8	14.3 - 19.4	86.9	86.2 - 87.6
≤105	11.38	91.06	1.27	0.97	16.5	14.1 - 19.0	86.9	86.2 - 87.6
≤106	12.35	90.38	1.28	0.97	16.6	14.3 - 19.0	87.0	86.2 - 87.6
≤107	13.62	89.78	1.33	0.96	17.1	14.9 - 19.5	87.0	86.3 - 87.7
≤108	15.55	88.43	1.34	0.95	17.2	15.1 - 19.5	87.1	86.4 - 87.8
≤109	16.67	87.70	1.35	0.95	17.3	15.3 - 19.5	87.2	86.5 - 87.9
≤110	18.08	86.86	1.38	0.94	17.5	15.6 - 19.7	87.3	86.5 - 88.0
≤111	19.42	85.81	1.37	0.94	17.5	15.6 - 19.5	87.3	86.6 - 88.0
≤112	22.47	83.80	1.39	0.93	17.7	15.9 - 19.6	87.5	86.8 - 88.2
≤113	23.36	82.54	1.34	0.93	17.1	15.4 - 19.0	87.4	86.7 - 88.2
≤114	24.63	81.23	1.31	0.93	16.9	15.2 - 18.6	87.5	86.7 - 88.2
≤115	26.04	79.86	1.29	0.93	16.7	15.1 - 18.3	87.5	86.7 - 88.2
≤116	29.02	77.18	1.27	0.92	16.4	15.0 - 18.0	87.5	86.8 - 88.3

Criterion DBP	Sensitivity	Specificity	+LR	-LR	+PV	95% CI	-PV	95% CI
≤117	29.99	75.56	1.23	0.93	15.9	14.5 - 17.4	87.5	86.7 - 88.2
≤118	31.32	74.03	1.21	0.93	15.7	14.4 - 17.2	87.5	86.7 - 88.2
≤119	33.04	72.11	1.18	0.93	15.5	14.2 - 16.9	87.4	86.7 - 88.2
≤120	37.72	68.35	1.19	0.91	15.6	14.3 - 16.9	87.6	86.8 - 88.4
≤121	40.10	66.29	1.19	0.90	15.5	14.4 - 16.8	87.7	86.9 - 88.5
≤122	42.71	64.09	1.19	0.89	15.5	14.4 - 16.7	87.9	87.0 - 88.6
≤123	44.27	62.07	1.17	0.90	15.3	14.2 - 16.5	87.8	87.0 - 88.6
≤124	47.77	58.09	1.14	0.90	15.0	13.9 - 16.1	87.8	86.9 - 88.6
≤125	49.40	56.08	1.12	0.90	14.8	13.8 - 15.9	87.8	86.9 - 88.6
≤126	51.49	53.65	1.11	0.90	14.7	13.7 - 15.7	87.7	86.8 - 88.6
≤127	53.13	51.45	1.09	0.91	14.5	13.5 - 15.5	87.6	86.7 - 88.5
≤128	56.03	47.08	1.06	0.93	14.1	13.1 - 15.0	87.4	86.4 - 88.3
≤129	57.74	44.94	1.05	0.94	14.0	13.1 - 14.9	87.3	86.3 - 88.3
≤130	59.30	42.90	1.04	0.95	13.8	13.0 - 14.8	87.2	86.2 - 88.2
≤131	61.01	40.76	1.03	0.96	13.7	12.9 - 14.6	87.1	86.0 - 88.1
≤132	63.91	37.00	1.01	0.98	13.6	12.7 - 14.4	86.9	85.8 - 88.0
≤133	65.62	34.82	1.01	0.99	13.5	12.7 - 14.3	86.8	85.6 - 87.9
≤134	67.63	32.92	1.01	0.98	13.5	12.7 - 14.3	86.8	85.6 - 87.9
≤135	69.12	30.96	1.00	1.00	13.4	12.6 - 14.2	86.6	85.4 - 87.8
≤136	73.51	26.88	1.01	0.99	13.5	12.7 - 14.3	86.8	85.4 - 88.0
≤137	75.15	25.22	1.00	0.99	13.5	12.7 - 14.2	86.8	85.4 - 88.1
≤138	77.01	23.15	1.00	0.99	13.4	12.7 - 14.2	86.7	85.2 - 88.0
≤139	79.02	21.06	1.00	1.00	13.4	12.7 - 14.2	86.6	85.1 - 88.1
≤140	82.89	17.07	1.00	1.00	13.4	12.7 - 14.1	86.6	84.9 - 88.2
≤141	84.90	15.44	1.00	0.98	13.4	12.7 - 14.2	86.9	85.1 - 88.5
≤142	86.68	13.81	1.01	0.96	13.5	12.7 - 14.2	87.0	85.1 - 88.7
≤143	87.95	12.23	1.00	0.99	13.4	12.7 - 14.1	86.8	84.7 - 88.6
≤144	90.18	9.47	1.00	1.04	13.3	12.7 - 14.1	86.2	83.8 - 88.3
≤145	91.00	8.35	0.99	1.08	13.3	12.6 - 14.0	85.7	83.2 - 88.0
≤146	91.89	7.39	0.99	1.10	13.3	12.6 - 14.0	85.5	82.8 - 87.9
≤147	92.71	6.21	0.99	1.17	13.3	12.6 - 14.0	84.6	81.6 - 87.4
≤148	94.49	4.65	0.99	1.18	13.3	12.6 - 14.0	84.5	81.0 - 87.6
≤149	95.39	3.86	0.99	1.20	13.3	12.6 - 14.0	84.4	80.4 - 87.8
≤150	96.13	3.16	0.99	1.22	13.3	12.6 - 14.0	84.1	79.7 - 87.9
≤151	96.95	2.57	1.00	1.19	13.3	12.7 - 14.0	84.5	79.5 - 88.6
≤152	97.92	1.74	1.00	1.20	13.4	12.7 - 14.0	84.4	78.2 - 89.3
≤153	98.36	1.36	1.00	1.21	13.4	12.7 - 14.0	84.3	77.2 - 89.9
≤154	98.59	1.07	1.00	1.32	13.4	12.7 - 14.0	83.0	74.8 - 89.5
≤155	98.74	0.82	1.00	1.55	13.3	12.7 - 14.0	80.7	70.9 - 88.3
≤156	99.11	0.51	1.00	1.76	13.3	12.7 - 14.0	78.6	65.4 - 88.5
≤157	99.18	0.40	1.00	2.03	13.3	12.7 - 14.0	76.1	61.2 - 87.4
≤158	99.26	0.36	1.00	2.09	13.3	12.7 - 14.0	75.6	59.5 - 87.8
≤159	99.48	0.29	1.00	1.81	13.4	12.7 - 14.0	78.1	60.0 - 90.7
≤160	99.55	0.20	1.00	2.28	13.4	12.7 - 14.0	73.9	51.0 - 90.1
≤161	99.63	0.17	1.00	2.16	13.4	12.7 - 14.1	75.0	50.9 - 91.3
≤162	99.63	0.15	1.00	2.49	13.4	12.7 - 14.1	72.2	46.5 - 90.3
≤164	99.70	0.12	1.00	2.59	13.4	12.7 - 14.1	71.4	40.7 - 92.1

Criterion DBP	Sensitivity	Specificity	+LR	-LR	+PV	95% CI	-PV	95% CI
≤166	99.85	0.092	1.00	1.62	13.4	12.7 - 14.1	80.0	42.2 - 97.9
≤167	99.85	0.081	1.00	1.85	13.4	12.7 - 14.1	77.8	37.5 - 97.7
≤168	99.85	0.069	1.00	2.16	13.4	12.7 - 14.1	75.0	32.1 - 97.5
≤170	99.85	0.058	1.00	2.59	13.4	12.7 - 14.1	71.4	25.8 - 97.2
≤171	99.93	0.046	1.00	1.62	13.4	12.7 - 14.1	80.0	28.4 - 99.5
≤174	99.93	0.035	1.00	2.16	13.4	12.7 - 14.1	75.0	19.4 - 99.4
≤175	99.93	0.023	1.00	3.23	13.4	12.7 - 14.1	66.7	9.4 - 99.2
≤176	100.00	0.023	1.00	0.00	13.4	12.7 - 14.1	100.0	2.5 - 100.0
≤178	100.00	0.012	1.00	0.00	13.4	12.7 - 14.1	100.0	0.0 - 100.0
≤180	100.00	0.00	1.00		13.4	12.7 - 14.1		

APPENDIX H

Criterion values – pulse rate

Criterion Pulse	Sensitivity	Specificity	+LR	-LR	+PV	95% CI	-PV	95% CI	Cost
<8	0.00	100.00		1.00			87.1	86.2 - 87.9	-0.0330
≤8	0.12	99.96	3.37	1.00	33.3	0.8 - 90.6	87.1	86.2 - 87.9	-0.0330
≤18	0.36	99.96	10.10	1.00	60.0	14.7 - 94.7	87.1	86.3 - 87.9	-0.0332
≤26	0.36	99.95	6.73	1.00	50.0	11.8 - 88.2	87.1	86.3 - 87.9	-0.0332
≤34	0.48	99.95	8.98	1.00	57.1	18.4 - 90.1	87.1	86.3 - 87.9	-0.0333
≤40	1.08	99.95	20.20	0.99	75.0	42.8 - 94.5	87.2	86.3 - 88.0	-0.0337
≤47	1.20	99.93	16.84	0.99	71.4	41.9 - 91.6	87.2	86.4 - 88.0	-0.0337
≤53	2.17	99.89	20.20	0.98	75.0	52.8 - 90.5	87.3	86.5 - 88.1	-0.0343
≤55	2.29	99.89	21.32	0.98	76.0	54.4 - 90.9	87.3	86.5 - 88.1	-0.0344
≤58	3.13	99.86	21.89	0.97	76.5	58.5 - 89.4	87.4	86.6 - 88.2	-0.0349
≤63	3.49	99.82	19.53	0.97	74.4	57.6 - 87.1	87.4	86.6 - 88.2	-0.0352
≤67	3.61	99.82	20.20	0.97	75.0	58.6 - 87.4	87.5	86.6 - 88.3	-0.0352
≤68	5.17	99.77	22.27	0.95	76.8	63.6 - 87.0	87.6	86.8 - 88.4	-0.0362
≤71	5.29	99.77	22.79	0.95	77.2	64.2 - 87.3	87.6	86.8 - 88.4	-0.0363
≤72	7.10	99.61	18.06	0.93	72.8	61.8 - 82.1	87.8	87.0 - 88.6	-0.0375
≤75	7.34	99.61	18.67	0.93	73.5	62.7 - 82.6	87.9	87.0 - 88.7	-0.0376
≤77	10.35	99.30	14.85	0.90	68.8	59.9 - 76.8	88.2	87.4 - 89.0	-0.0395
≤78	10.47	99.30	15.02	0.90	69.0	60.2 - 77.0	88.2	87.4 - 89.0	-0.0396
≤80	12.52	98.39	7.78	0.89	53.6	46.3 - 60.8	88.3	87.5 - 89.1	-0.0409
≤82	12.64	98.39	7.86	0.89	53.8	46.6 - 61.0	88.4	87.5 - 89.1	-0.0410
≤84	13.48	97.02	4.52	0.89	40.1	34.3 - 46.2	88.3	87.5 - 89.1	-0.0415
≤85	13.84	97.02	4.64	0.89	40.8	35.0 - 46.8	88.3	87.5 - 89.1	-0.0418
≤87	13.96	97.02	4.68	0.89	41.0	35.2 - 47.0	88.4	87.5 - 89.2	-0.0418
≤88	16.13	95.78	3.82	0.88	36.2	31.3 - 41.4	88.5	87.7 - 89.3	-0.0432
≤90	16.37	95.78	3.88	0.87	36.6	31.6 - 41.7	88.5	87.7 - 89.3	-0.0434
≤91	18.05	93.85	2.94	0.87	30.4	26.3 - 34.6	88.5	87.7 - 89.3	-0.0445
≤92	18.17	93.85	2.96	0.87	30.5	26.5 - 34.8	88.5	87.7 - 89.3	-0.0446
≤93	18.41	93.85	3.00	0.87	30.8	26.8 - 35.0	88.6	87.7 - 89.4	-0.0447
≤94	20.34	91.33	2.35	0.87	25.8	22.5 - 29.4	88.5	87.7 - 89.3	-0.0460
≤95	20.70	91.33	2.39	0.87	26.2	22.9 - 29.7	88.6	87.7 - 89.4	-0.0462
≤96	20.82	91.33	2.40	0.87	26.3	23.0 - 29.8	88.6	87.7 - 89.4	-0.0463
≤97	22.98	87.96	1.91	0.88	22.1	19.4 - 25.0	88.5	87.6 - 89.3	-0.0478
≤98	22.98	87.94	1.91	0.88	22.1	19.3 - 25.0	88.5	87.6 - 89.3	-0.0478
≤99	24.91	83.52	1.51	0.90	18.3	16.1 - 20.7	88.2	87.3 - 89.1	-0.0491
≤100	25.03	83.51	1.52	0.90	18.4	16.2 - 20.8	88.2	87.3 - 89.1	-0.0492
≤101	25.03	83.49	1.52	0.90	18.4	16.2 - 20.8	88.2	87.3 - 89.1	-0.0492
≤102	29.24	78.43	1.36	0.90	16.8	14.9 - 18.8	88.2	87.3 - 89.1	-0.0519
≤103	29.36	78.43	1.36	0.90	16.8	14.9 - 18.8	88.2	87.3 - 89.1	-0.0520
≤104	33.33	73.18	1.24	0.91	15.6	13.9 - 17.3	88.1	87.1 - 89.0	-0.0547
≤105	33.81	73.14	1.26	0.90	15.8	14.1 - 17.5	88.2	87.2 - 89.1	-0.0550
≤106	34.54	73.14	1.29	0.90	16.0	14.4 - 17.8	88.3	87.3 - 89.2	-0.0554
≤107	41.28	67.76	1.28	0.87	16.0	14.4 - 17.6	88.6	87.6 - 89.5	-0.0598
≤108	41.76	67.76	1.30	0.86	16.1	14.6 - 17.8	88.7	87.7 - 89.6	-0.0601
≤109	44.89	62.76	1.21	0.88	15.2	13.8 - 16.7	88.5	87.4 - 89.4	-0.0622

Criterion Pulse	Sensitivity	Specificity	+LR	-LR	+PV	95% CI	-PV	95% CI	Cost
≤110	45.01	62.74	1.21	0.88	15.2	13.8 - 16.7	88.5	87.4 - 89.5	-0.0623
≤111	49.58	57.22	1.16	0.88	14.7	13.4 - 16.0	88.4	87.3 - 89.5	-0.0653
≤112	50.06	57.22	1.17	0.87	14.8	13.5 - 16.2	88.5	87.4 - 89.5	-0.0656
≤113	54.99	51.88	1.14	0.87	14.5	13.3 - 15.8	88.6	87.4 - 89.7	-0.0689
≤115	59.21	46.84	1.11	0.87	14.2	13.0 - 15.4	88.5	87.3 - 89.7	-0.0716
≤116	59.57	46.82	1.12	0.86	14.3	13.1 - 15.5	88.6	87.4 - 89.8	-0.0719
≤117	64.14	41.74	1.10	0.86	14.1	13.0 - 15.2	88.7	87.4 - 89.9	-0.0749
≤118	64.26	41.74	1.10	0.86	14.1	13.0 - 15.2	88.7	87.4 - 89.9	-0.0750
≤119	67.99	36.40	1.07	0.88	13.7	12.7 - 14.8	88.4	87.1 - 89.7	-0.0774
≤120	71.60	31.58	1.05	0.90	13.4	12.5 - 14.5	88.2	86.7 - 89.6	-0.0798
≤121	71.96	31.58	1.05	0.89	13.5	12.5 - 14.6	88.3	86.9 - 89.7	-0.0801
≤122	75.33	27.20	1.03	0.91	13.3	12.4 - 14.3	88.1	86.5 - 89.6	-0.0823
≤123	75.57	27.20	1.04	0.90	13.4	12.4 - 14.4	88.2	86.6 - 89.7	-0.0825
≤124	78.82	23.14	1.03	0.92	13.2	12.3 - 14.2	88.0	86.3 - 89.7	-0.0846
≤125	81.59	19.12	1.01	0.96	13.0	12.1 - 14.0	87.5	85.5 - 89.3	-0.0865
≤126	81.83	19.12	1.01	0.95	13.1	12.2 - 14.0	87.6	85.7 - 89.4	-0.0866
≤127	84.48	15.53	1.00	1.00	12.9	12.0 - 13.9	87.1	84.8 - 89.1	-0.0884
≤128	86.64	12.28	0.99	1.09	12.8	11.9 - 13.7	86.1	83.5 - 88.4	-0.0898
≤129	88.81	9.56	0.98	1.17	12.7	11.9 - 13.6	85.2	82.2 - 87.9	-0.0913
≤130	88.93	9.51	0.98	1.16	12.7	11.9 - 13.6	85.3	82.2 - 87.9	-0.0913
≤131	90.37	7.20	0.97	1.34	12.6	11.8 - 13.5	83.4	79.8 - 86.6	-0.0923
≤132	91.70	5.75	0.97	1.44	12.6	11.8 - 13.5	82.4	78.2 - 86.0	-0.0932
≤133	93.26	4.40	0.98	1.53	12.7	11.8 - 13.5	81.5	76.6 - 85.7	-0.0942
≤134	93.74	3.40	0.97	1.84	12.6	11.8 - 13.4	78.5	72.8 - 83.5	-0.0945
≤135	94.58	2.61	0.97	2.08	12.6	11.8 - 13.5	76.4	69.8 - 82.3	-0.0951
≤136	94.95	2.04	0.97	2.48	12.6	11.8 - 13.4	73.1	65.4 - 79.9	-0.0953
≤137	95.07	2.02	0.97	2.44	12.6	11.8 - 13.4	73.4	65.6 - 80.2	-0.0954
≤138	95.43	1.61	0.97	2.84	12.6	11.8 - 13.4	70.3	61.6 - 78.1	-0.0956
≤139	95.55	1.30	0.97	3.41	12.6	11.8 - 13.4	66.4	56.7 - 75.1	-0.0957
≤140	95.91	0.86	0.97	4.77	12.6	11.8 - 13.4	58.5	47.1 - 69.3	-0.0960
≤141	96.27	0.79	0.97	4.74	12.6	11.8 - 13.4	58.7	46.7 - 69.9	-0.0962
≤142	96.39	0.73	0.97	4.93	12.6	11.8 - 13.4	57.7	45.3 - 69.5	-0.0963
≤143	96.63	0.66	0.97	5.10	12.6	11.8 - 13.5	56.9	44.0 - 69.2	-0.0964
≤144	96.63	0.61	0.97	5.55	12.6	11.8 - 13.5	54.8	41.7 - 67.5	-0.0964
≤145	96.75	0.57	0.97	5.68	12.6	11.8 - 13.5	54.2	40.8 - 67.3	-0.0965
≤146	96.87	0.50	0.97	6.25	12.6	11.8 - 13.5	51.9	37.8 - 65.7	-0.0966
≤147	96.87	0.43	0.97	7.30	12.6	11.8 - 13.5	48.0	33.7 - 62.6	-0.0966
≤149	96.99	0.38	0.97	8.02	12.6	11.8 - 13.5	45.7	30.7 - 61.2	-0.0966
≤150	97.35	0.36	0.98	7.41	12.7	11.9 - 13.5	47.6	32.0 - 63.6	-0.0969
≤151	97.83	0.36	0.98	6.06	12.7	11.9 - 13.6	52.6	35.6 - 69.2	-0.0972
≤152	98.19	0.29	0.98	6.31	12.8	11.9 - 13.6	51.6	33.1 - 69.8	-0.0974
≤153	98.19	0.23	0.98	7.77	12.8	11.9 - 13.6	46.4	27.5 - 66.1	-0.0974
≤154	98.32	0.20	0.99	8.57	12.8	12.0 - 13.6	44.0	24.4 - 65.1	-0.0975
≤155	98.68	0.14	0.99	9.26	12.8	12.0 - 13.6	42.1	20.3 - 66.5	-0.0977
≤156	99.16	0.14	0.99	5.89	12.9	12.0 - 13.7	53.3	25.7 - 79.5	-0.0980
≤157	99.28	0.13	0.99	5.77	12.9	12.1 - 13.7	53.8	24.0 - 81.7	-0.0981
≤158	99.52	0.13	1.00	3.85	12.9	12.1 - 13.7	63.6	30.8 - 89.1	-0.0982

Criterion Pulse	Sensitivity	Specificity	+LR	-LR	+PV	95% CI	-PV	95% CI	Cost
≤159	99.64	0.13	1.00	2.89	12.9	12.1 - 13.7	70.0	34.8 - 93.3	-0.0983
≤160	99.76	0.13	1.00	1.92	12.9	12.1 - 13.8	77.8	37.5 - 97.7	-0.0984
≤200	100.00	0.00	1.00		12.9	12.1 - 13.8			-0.0986