WORK IN PROGRESS - PLEASE DO NOT QUOTE

A predictive model for use in long-term conditions management in a Primary Care Trust

Peter Yuen, Gerald Lutgen, Nancy K. Williams, Donovan Crew, Mary Anne Smalley

UnitedHealth Europe
15 Greycoat Place
London SW1P 1SB

Email: peter_yuen@uhc.com

December 2006

Aims: to describe a predictive model to identify and stratify patients at high risk of unplanned admission related to a chronic disease and how the results of the model can be used in long-term/chronic disease management programme in PCTs in England

Methods: Patient-level activity data were obtained from PCTs within two Strategic Health Authorities (SHAs) on acute hospital inpatient, outpatient, Accident and Emergency (A&E) departments as well as GP practices. Information on medical and clinical history, use of inpatient, outpatient and A&E, GP consultation and medications in a period of 12 months were used to construct a set of potential predictor variables to predict the probability of a patient to be admitted to hospital as an emergency case in the next 12 months, accompanied by a diagnosis for a chronic disease. Multiple logistic regression was used to derive weights for these predictor variables using historical data. These weights were then applied to the most recent 12 months of experience data for the whole population in order to calculate risk scores for each individual in the population. By ranking these risk scores, high risk patients can then be identified. With the availability of patient-level medical/clinical information, community matrons, nurses or clinicians will be able to pin-point those patients in need of long-term condition management.

Results: Age of the patient, use of prescription only medicine, illness score derived from the medical and clinical history of the patients, use of specialist care, number of outpatient attendances and visits to A&E departments are significant predictors for emergency admissions. The Area Under ROC curve (AUC) is 87%, showing excellent results of the model in discriminating patients correctly.

Conclusions: Stratifying patients using a predictive modelling tool provides PCT's with further insight into key factors that impact their patients' health. By reviewing patients' clinical history alongside their predicted level of risk, community matrons or case managers are able to conduct patient assessments both more efficiently through the immediate identification of patients at greatest risk, and more effectively by associating patterns of chronic health care utilisation with increased risk. Additionally, the summary report provides GPs a stratified list of their patients which is useful for practice based commissioning and programme development. This method provides opportunities for early intervention and treatment in order to improve outcomes, such as diverting avoidable hospital admissions, thereby lessening the strain on acute hospital resources.

Introduction

The mounting problem faced by the NHS in England of people suffering from long-term conditions is reflected by figures extracted from GP disease registries. The latest Quality and Outcome Framework for GP practices (QOF) [1] disease prevalence data show that of the 49 million people in England, 6.5 million people have hypertension, over 3 million have asthma, 2 million have coronary heart disease, 1.9 million are diabetic, nearly 1 million have stroke and 0.7 million have COPD. Even these figures are an under-estimation of the number of people suffering from long-term conditions such as mental health and arthritis due to under-reporting or exclusion from the disease registries. Results from health surveys also suggest that as many as one-third of the England population have one or more chronic health problems. Multiple chronic conditions are causing serious concern among the general population, especially amongst older people.

The economic and health care burden of long-term conditions is huge. The care of people with chronic conditions consumes disproportionate amounts of health care resources especially in acute hospital care. People with long-term conditions are found to be frequent users of inpatient, outpatient and accident and emergency departments. Patients with a chronic condition or complications use over 60% of hospital bed days and account for about two thirds of emergency admission [2]. It is estimated that 80% of GP consultations relate to long-term conditions. The use of POM (Prescription Only Medications) is high among older people with multiple chronic conditions [3]. Evidence from the US shows the care of people with long-term conditions consumes over 75% of the total healthcare spending.

The management of high cost patients with long-term conditions is high on the Government's health policy agenda. The White Paper "The NHS improvement plan: putting people at the heart of public services" [4] published in 2004 addressed the problem of long-term conditions and called for the support of people with long-term conditions to live healthy lives. One of the Public Service Agreement (PSA) targets on health [5] specifies the need to improve outcomes for people with chronic health conditions by reducing emergency bed days by 5% by 2008 through improved care in primary and community care. The target also requires assuring vulnerable patients at risk to be offered personalised care plan.

Against this background, the policy framework "Supporting people with Long-term conditions" [6] published in 2005 introduced a model designed to help improve care of people with chronic conditions. Among a number of key aspects of the model is to identify patients with long-term conditions locally and to stratify patients into levels of need (see Figure 1). It is believed timely and appropriate intervention would improve care and prevent future hospital admissions especially unplanned hospitalisation which results in better outcomes and saves resources in the long run. Avoidable unplanned admissions may be an indicator of inappropriate patient care. When a patient's health deteriorates in an un-managed way, avoidable unplanned admissions often occur in an acute setting, resulting in costly treatment in an expensive emergency setting. To reduce avoidable unplanned admissions the Department of Health has set targets for 3,000 community matrons to be appointed by 2007 to organise, manage and join up care through case management for those patients with highly complex needs. US evidence shows that clinical intervention through effective case and disease management programmes can improve health outcomes and control costs by focusing resources on care of patients at high risk of deteriorating health [7]. There is also evidence of cost-effectiveness of patient screening tools of case management programmes[8].

In the US, Congress mandated in the 1979 Balance Budget Act a study of best practices in case management and disease management programmes. Mathmatica Policy and Reach Group studied the success of 157 distinct programmes. They concluded that having a risk stratification method of patient selection is integral to the overall success of any programme committed to achieving impressive outcomes.

An efficient, effective and accurate method is needed to identify and stratify patients at high risk of unplanned admissions. There are three common methods:

1. Threshold modelling is a rule/criteria-based case finding method. An Example of this approach is using 2 or more hospital admissions, age only and/or two plus long term conditions. This method has been found to be less effective because is often suffers from selection bias and it can have low predictive accuracy [9].
2. The clinical referral method which is based on doctors/nurses knowledge and experience has also been found to be inefficient, highly biased, and has low levels of accuracy when used on its own.
3. The predictive model may use some combination of historical clinical, health service utilisation, demographic and socio-economic data to predict future health outcomes such as unplanned hospital admission or cost. Predictive models typically are more accurate and suffer less from selection bias. The power of the model can also be

magnified when the results are further reviewed by clinical personnel prior to making patient programme classification decisions [10].

Once established, a well constructed predictive model can then be used to derive risk scores to rank the population according to their risk for a specific outcome. PCT's, care managers, commissioners and clinicians can then use the resulting list to make decisions about the level of intervention a patient should receive. By proactive identification and stratification of patients at risk, a predictive model can help improve financial and health outcomes.

In this paper we describe how a predictive model and the subsequent screening tool RISC (Risk Identification and StratifiCation) was developed for use to identify and stratify patients at risk of unplanned admission related to chronic disease and how the results of the model can be used in a long-term/chronic disease management programme in PCT's throughout England.

Methods and Material

The development of RISC (Risk Identification and StratifiCation) involves the process of collecting patient health utilisation data from the participating PCTs, preparing the data for analysis, performing an initial fitting for the predictive model, applying the resulting weights periodically to the whole population to calculate a risk ratio for each individual and finally producing both a patient level and a population risk profile report (see Figure 2).

Data Source

We received patient-level data for a period of 24 months from participating PCTs within two former SHAs in England. These data covered activities from both the primary care and secondary acute care of patients for whom the PCTs were responsible. Data from Acute Trusts included records of admitted patient care (including inpatient and day case treatment), outpatient attendances and Accident and Emergency (A&E) visits. GP consultation and medication records were extracted from various electronic patient systems of participating GP practices within the PCTs. Patient identification information in the form of registration files were prepared by the PCTs and encrypted according to Caldicott guidelines. Details of the information collected from Acute Trusts, GP practices and PCTs are shown in Table 1.

Once received, the data are checked for quality through a data validation process. Data failing the quality threshold are returned to the PCTs for re-work and re-submission. Data extraction from GP electronic patient record system is still relatively new and has been problematic. As a result, there was significant difficulty in getting GP data from most PCTs. This has greatly affected the number of GP records available for the analysis.

Predictive model

The construction of a predictive model consists of selecting the base and target time periods, identifying a target variable, reviewing potential predictor variables from the base period and then selecting and weighting the predictors of the target variable through statistical modelling.

Patients are admitted to hospital as emergency cases for various medical reasons. In this context we model the risk of having at least one emergency hospital admission with a chronic condition in the next 12 months as the target (i.e. dependent) variable in the predictive model. These chronic conditions include conditions such as COPD, asthma, congestive heart disease, stroke and others which in our data set account for around 50% of all emergency admissions and where effective primary and community care, social services and case management would help in reducing the risk of unplanned admission.

To predict unplanned admission in the next 12 months, two years of data are required: the first 12 months for the base year and second 12 months for the target year. The base and target year data are used for model building by taking prior information from the base year to model the known outcome (unplanned admission with a chronic condition) in the target year. The resulting model can then be used to predict the outcome in a future predicted year using data from a preceding base year (See Figure 3).

The choice of variables is driven by clinical relevance and evidence but data availability must also be considered. With this in mind a set of patient-level data which allow for a list of potential predictors are collected from the participating PCTs for model building. In addition to age, this data relates to the patients clinical/medical history as well as their pattern of health services utilisation.

A conscious choice was made to exclude a series of proposed variables that relate to the patients in certain groups but do not relate to their own clinical history. Examples of these variables are

a. Patient Ethnicity

b. Unplanned admission rate of GP to which the patient is registered

c. Readmission rate of the last Acute Trust to which the patient was admitted

d. Unplanned admission rate of the locality in which the patient lives.

These types of variables are often of keen epidemiologic interest. However, these variables usually do little to improve the overall accuracy of the models, while at the same time introducing complexity to the models and actually reducing the usefulness of models as indicators of patient-level need for care management. Also, in the case of Ethnicity there is a major concern about the quality of the underlying data.

The final list of potential predictors (independent variables) is derived from the data shown below:

Patient Age:

*Prior utilisation of acute care in the last 12 month periods***:**

? Accident and Emergency department visits

? Outpatient attendances

? Planned and unplanned hospital admission

Clinical/Medical history in the last 12 Months – Burden of Illness;

Burden of Illness is a composite measure of how much a persons risk is increased due to the presence of indicators of chronic disease. Burden of Illness is built up from 3 components of data

? Diseases identified by an examination of diagnosis and procedure coding for inpatient admissions

? High level inferences regarding the presence of disease based on record of specialist visited in both inpatient and outpatient settings

? Medical conditions identified from GP consultation information, including Read Diagnosis codes and medication history

For details of Burden of Illness, see Appendix 1.

Use of prescription medicine: Unique count of prescription medicines

Once the set of potential predictor variables has been constructed, statistical modelling using multiple logistic regression [11] is carried out to evaluate the relationship between these predictors derived from the base year and the actual emergency admission related to a chronic condition from the target year. The analysis is done using SAS v.9. Through fitting and re-fitting, testing and re-testing, statistically significant predictors are identified and the final model is obtained. The accuracy of the predictive model in its ability to identify high risk patients correctly at a certain cut-off point is assessed using sensitivity, specificity and the positive predictive values (PPV). The overall performance of the model's ability in discriminating between those patients who will have unplanned admission and those who will not is given by AUC (Area Under the ROC curve) [11].

Risk Scoring

By applying the weights (the ß coefficients) of the final model to the corresponding variables in the production base year, a predicted risk score or likelihood of unplanned admission in the next 12 months for each individual person in the entire PCT population is obtained. A risk ratio for each individual is then derived by dividing the patient's risk score with the percentage of people at risk of unplanned admission with a chronic condition at the PCT level. Thus an individual with a risk ratio of 30 has a 30-fold increased chance of an unplanned admission due to chronic disease relative to the average of the PCT where the individual resides.

Reporting system

RISC produces a master listing of all registered patients within a PCT ranked by Risk Ratio along with key utilisation history summary for each patient including inpatient, outpatient, A&E as well as GP visits and medication history in the past 12 months. For each patient a summary and detailed patient profile report are available. The detail profile report presents a full picture of the patient's healthcare utilization in the past 12 months. These reports allow high level or drill down for a complete patient's picture- saving hours of clinical time reviewing medical history. For confidentiality, patient's details can only be decrypted by authorised PCT personnel such as the IM&T (Shared Services) so that patient's name and NHS number will be available to GPs and community matrons.

Results

The mean age of patients included in the predictive model was around 38. On average they had 0.88 outpatient attendance, 0.15 A&E visit and had 3.15 unique drugs in the last 12 months (Table 2). Age of the patient, use of prescription only medicine, Burden of Illness score derived from the medical and clinical history of the patients, use of specialist care, number of outpatient attendances and number of visits to A&E departments are significant factors for the prediction of unplanned admission with a chronic condition in the next 12 months. The Area Under ROC curve (AUC) is 87%, showing excellent results of the model in discriminating patients correctly (Figure 4).

One of the most important outputs of RISC is the set of reports it generates. The Master List Report provides a list of patients ranked by their risk ratio. Figure 5 shows some of the highest score patients with their identification information taken out for confidentiality. The third patient on the list, for instance, is a female of 85 years of age, had a high score of 38.48, that is, an over 38 fold of risk of having an unplanned admission to hospital relative to the average resident in her local PCT. In the last 12 months, she had six inpatient admissions, all unplanned, four A&E visits and three outpatient attendances. She also visited her GP 27 times and had 31 unique medicines during that period of time. A detail Patient Profile Report is also available for this patient detailing the hospital activities including admission information, diagnosis and operation procedures, length of hospital stay etc. There are also details of each GP visit and medicines prescribed by the GP.

Discussion

Stratifying patients using a predictive modelling tool assists PCT's to gain further insight into factors that impact their patients' health. It allows clinical teams to proactively identify, analyse and stratify patients into actionable groups. GPs are able to look at their practices as a population and community matrons or case managers are able to begin the assessment process more effectively and efficiently by reviewing summary and patient detail clinical information. This will help them understand the factors that drive the risk score and alert them to opportunities for early intervention and treatment to improve patient outcome, including diverting avoidable admissions. PCT commissioners and clinical leads are able to use the results to analyse/study the patient population and understand the drivers of risk relative to chronic diseases.

RISC is designed as a tool to allow PCT's to gain greater insight into the factors that drive patient's health. Building on a strong theoretical framework, expert analytical experience and clinical knowledge with rigorous empirical experience and by combining sophisticated data management, high quality data and reporting software, RISC can assist community matrons, GP's and Long-term condition leads to rapidly and efficiently analyse and stratify the patients into actionable groups offering proactive verse reactive care to patients. By providing a mechanism to target resources to those most at need, RISC allows PCT's to better recognise the factors that drive healthcare cost through out the pyramid and promotes an overall understanding of key drivers of cost.

In addition to being a screening tool, RISC also provides users with operational capabilities by increasing the productivity and efficiency for long-term condition programme personnel. The RISC reports increases community matrons' efficiency in researching patient medical history and provides access to complete acute utilization which is often sketchy or often not recorded in the GP practice system. Patient diagnosis and procedure code descriptions are presented in the patient profile and no cross referencing to coding tables is necessary. In this way it saves significant amount of community matrons' time in researching patient's clinical background and profile.

Detailed data from primary and secondary acute care are organized in such a way that it aids the step of screening persons for LTC or other chronic disease programmes as well as managing them once they are in the programme. Through close integration between the risk scoring methods in the model and the RISC reporting tool, uers can clearly understand why patients get the score they have. Additionally, the reporting tool details patient's medication regimes and therefore is particularly helpful in identifying often overlooked poly-pharmacy issues.

RISC can lead to greater LTC programme impact through positive health changes for individual patients in the programme. The summary and detailed patient profile report presents a full picture of the patient's healthcare utilization in the past 12 months (assuming all primary and secondary care utilisation is captured) . These reports allow high level reviews and drill downs for a complete patient picture. Since RISC produces a score for every registered person in the PCT population, it can be used for all levels of an LTC programme, while at the same time providing important insights into whole populations for commissioning purposes.

The predictive model on which RISC is based is parsimonious and is built upon two distinct dimensions that drive risk. Firstly it uses a sub-model that measures “burden-of-illness”, a measure of the degree to which a patient’s specific mix of medical and clinical history impacts their risk of unplanned admission. Secondly it incorporates variables based on the utilisation of services from primary and secondary care. These variables enter into the final equation in a straightforward and interpretable manner that links with the companion reporting system. RISC, however, does not incorporate non patient level aggregate data such as ethnicity and multiple deprivation area indexes. The model uses the risk of unplanned admission with a diagnosis for a chronic disease as a more robust dependent variable. When combined with other modelling tactics this results in models with a superior performance profile.

While RISC and the predictive model perform well, there are limitations. A significant amount of production time of RISC has been spent on improving data quality checking and validating data submitted by PCTs for missing and miscoding. The deficiency in acute hospital activity data remains a major problem, especially in A&E and in outpatient reporting. Most of the acute trusts in the participating PCTs of the project failed to provide diagnostic and procedure information in these two areas making it difficult to report a complete picture of the patients’ medical history and for the construction of “Burden of Illness” scores. The lack of good quality GP practice data and lower than expected number of practices taking part in the project has greatly reduced the number of patients available for the predictive modelling. It is hoped that as the project progresses and by working closely with those involved in the data collection the situation will improve.

References

1. Health and Social Care Information Centre. Quality Outcomes Framework Data 2005/06. 2006. London, Health and Social Care Information Centre.

2. Department of Health. Improving Chronic Disease Management. 2004. London, Department of Health.

3. Yuen P. Use of prescription only medicine in older people. 2005. London, OHE.

4. Department of Health. NHS Improvement plan 2004: Putting people at the heart of public services. 2004. London, Department of Health.

5. HM Treasury. 2004 Spending Review: Public Service Agreements 2005 - 2008. 2004. London, HM Treasury.

6. Department of Health. Supporting People with Long Term Conditions. An NHS and Social Care Model to support local innovation and integration. London: Department of Health, 2005.

7. Sidorov J, Shull R, Tomcavage J, Girolami S, Lawton N, Harris R. Does diabetes diseasse management save money and improve outcomes? *Diabetes Care* 2002;**25**:684.

8. Mukamel D, Chou CC, Zimmer JG, Rothenberg BM. The effect of accurate patient screening on the cost-effectiveness of case management programs. *Gerontology* 1997;**37**:777.

9. National Primary Care Research and Development Centre. Evercare evaluation interim report: implications for supporting people with long - term conditions. 2005. Manchester, National Primary Care Research and Development Centre.

10. Cousins MS, Shickle LM, Bander JA. Introduction to predictive modelling. *Disease Management* 2002;**5**.

11. Hosmer DW, Lemeshow S. Applied logistic regression 2nd edition. New York: John Wiley & Son Inc, 2000.

Table 1: Data files, sources, and elements.

Data file (source)	**Element**
Registry	
(PCT)	Encrypted Patient ID
	Age
	Gender
	GP Practice Code
A&E visit	
(Acute Trust)	Encrypted Patient ID
	Date Of Service
	Diagnosis Code 1-2
	Investigation Code 1-2
	Treatment Code 1-2
	Trust Code
Outpatient attendance	
(Acute Trust)	Encrypted Patient ID
	Date Of Service
	Consultant Specialty Code
	Attendance Code
	First Attendance
	Trust Code
Admitted patient care (Inpatient and day cases)	
(Acute Trust)	Encrypted Patient ID
	Date Of Service
	Discharge Date
	Episode Start & End Dates
	Consultant Specialty Code
	Primary Diagnosis
	Subsequent Diagnosis
	Secondary Diagnosis 1-2
	Procedure Code 1-4
	HRG
	Admission Method
	Discharge Destination
	Spell Number
	Length of Spell
	Trust Code
GP consultation and Medication	
(GP Practice)	Encrypted Patient ID
	Activity Code or Medication Code
	Repeat Medication Flag

Note. List of elements is not exhaustive, but rather reflects the minimum request in order to test the predictive model. Not all of the elements listed above were necessarily used in the final model. Please contact the authors for further details.

Table 2 Descriptive Statistics for Model Development Data Set

Variable	Mean	Standard deviation
Age	37.9	22.7
Number of specialist consultations in the last 12 months	0.88	2.40
Number of unplanned admission related to chronic disease in the last 12 months	0.03	0.26
Number of A&E visits in the last 12 months	0.15	0.55
Number of unique drugs in the medication data base in the last 12 months	3.15	4.72

Figure 1 Kaiser Permanente risk stratification triangle

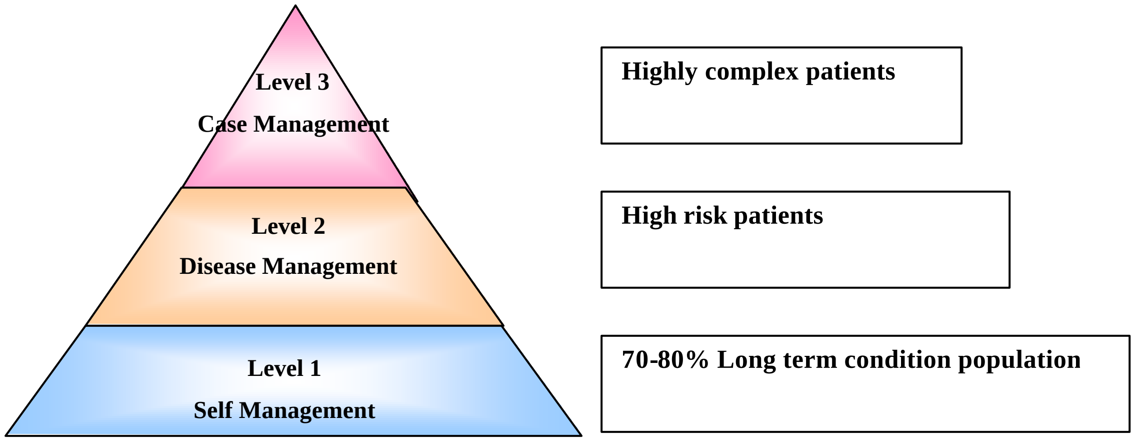


Figure 2. RISC process

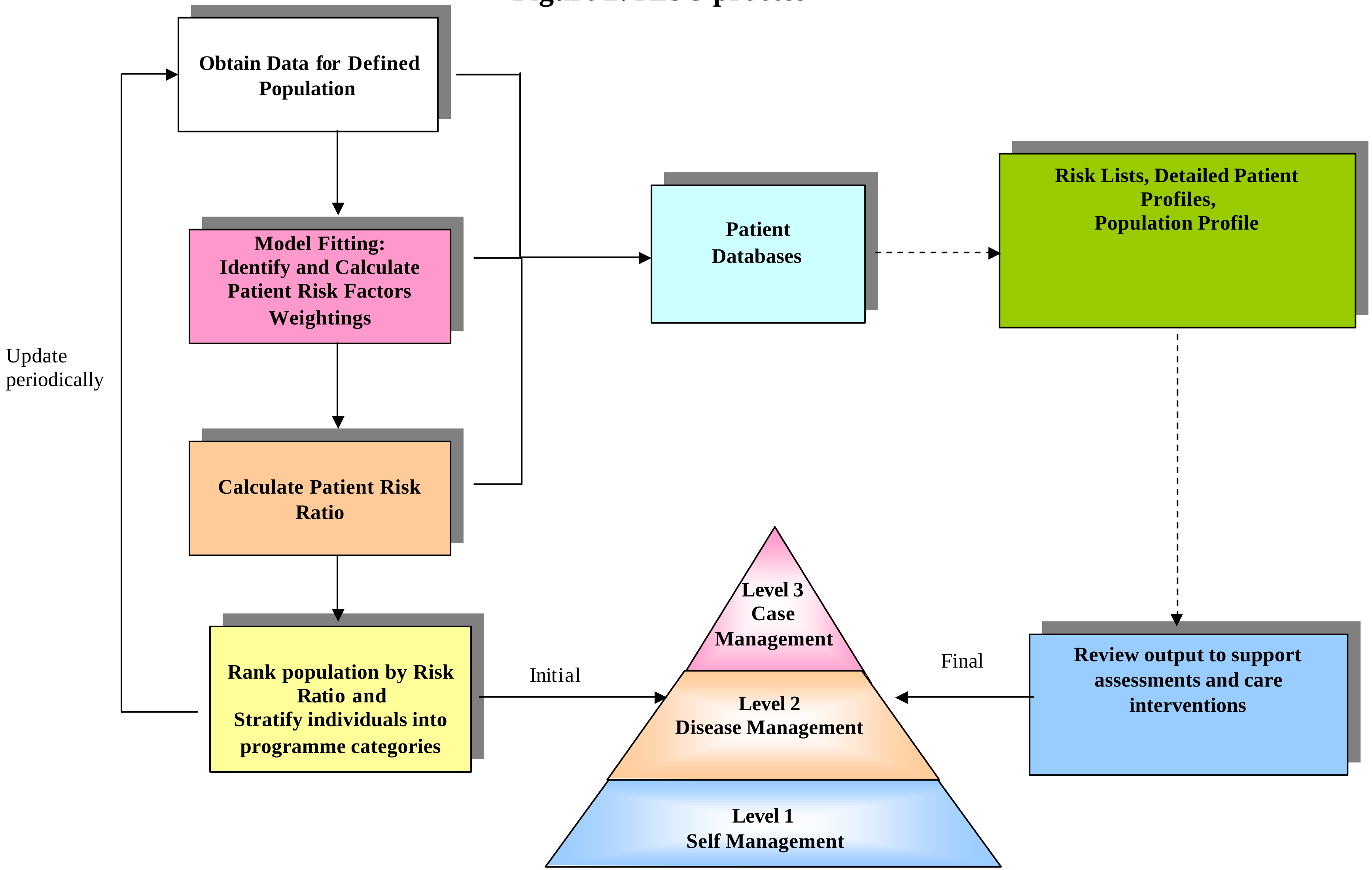


Figure 3. The Predictive Model

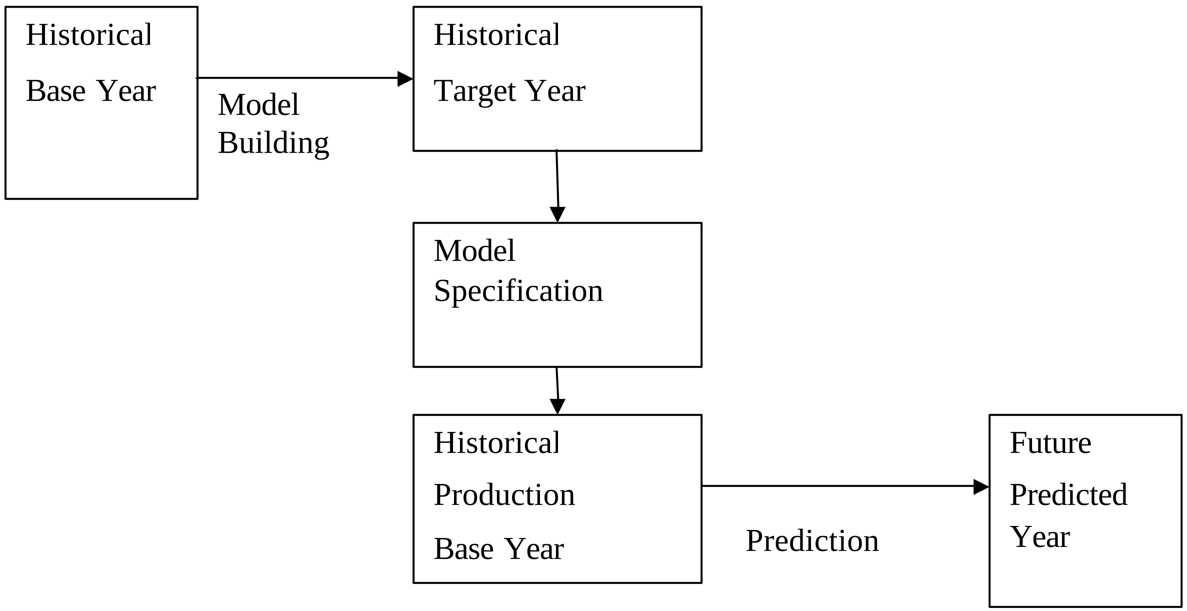


Figure 4: Area under ROC curve

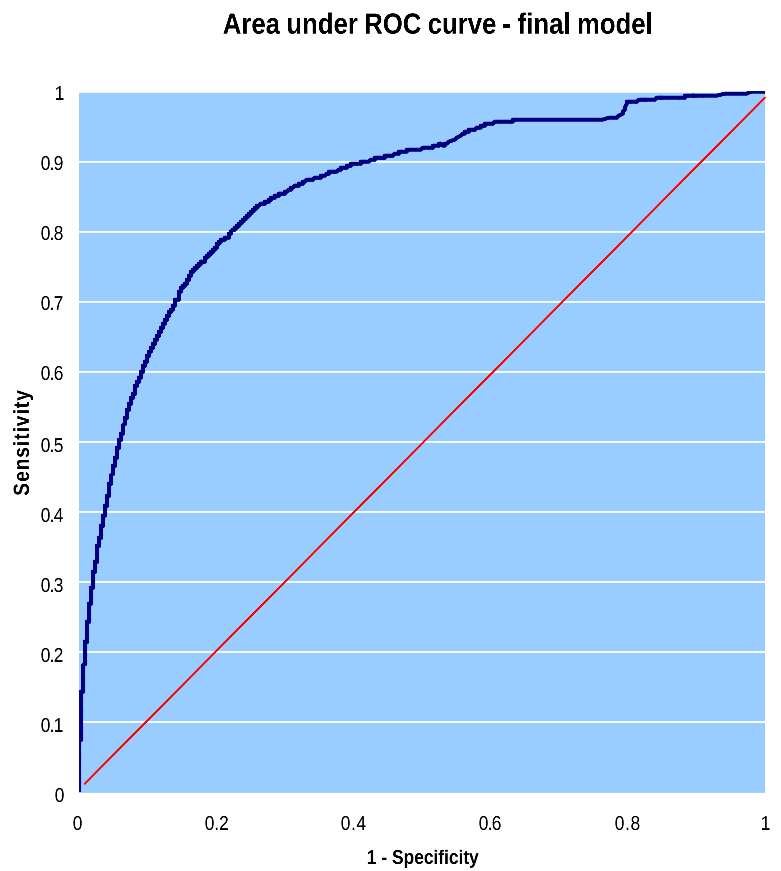
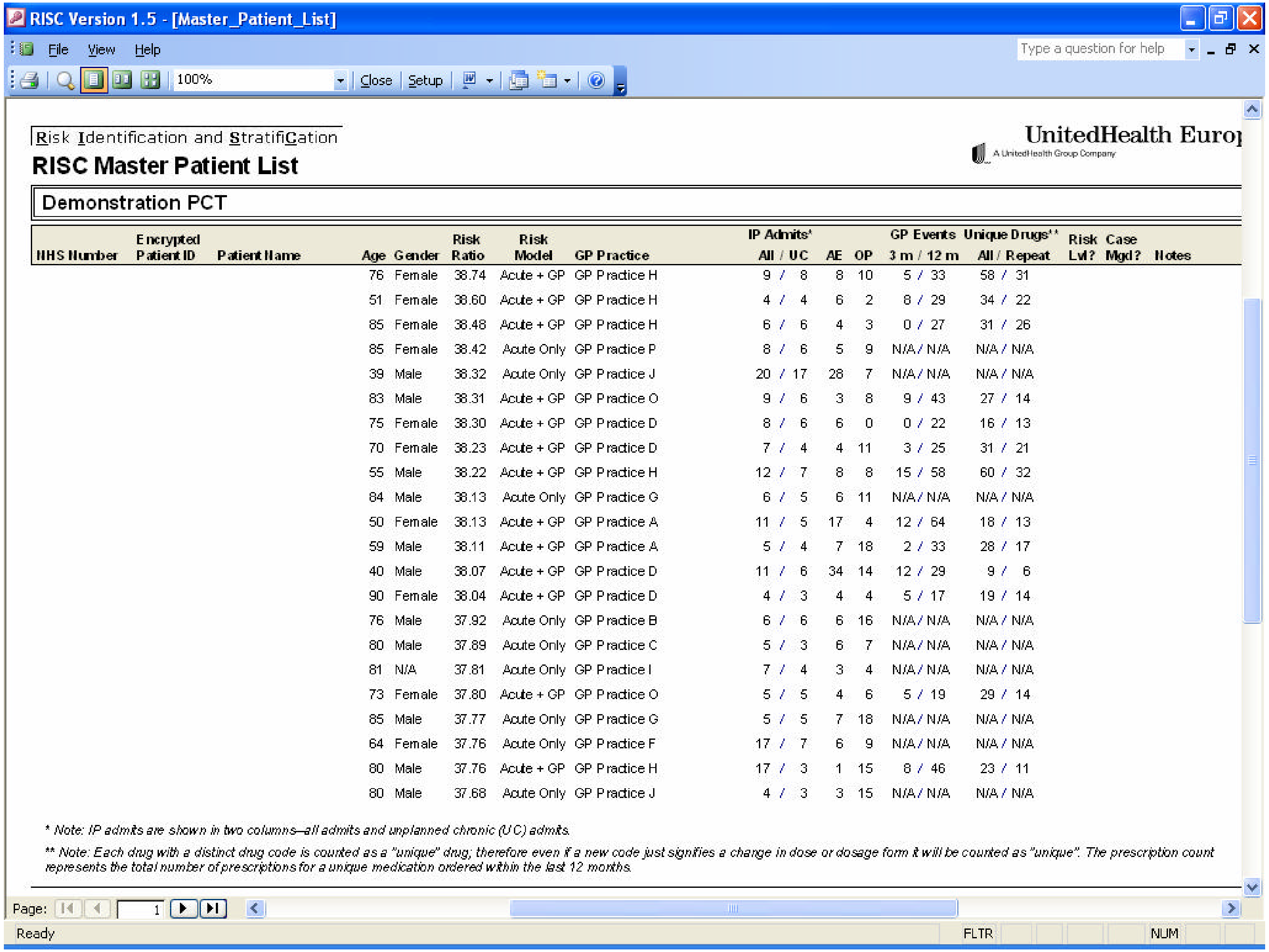


Figure 5 RISC Master Patient List



Risk Identification and StratifiCation

RISC Master Patient List

Demonstration PCT

NHS Number	Encrypted Patient ID	Patient Name	Age	Gender	Risk Ratio	Risk Model	GP Practice	IP Admits* All / UC	AE	OP	GP Events 3 m / 12 m	Unique Drugs** All / Repeat	Risk Lvl?	Case Mgd?	Notes
			76	Female	38.74	Acute + GP	GP Practice H	9 / 8	8	10	5 / 33	58 / 31			
			51	Female	38.60	Acute + GP	GP Practice H	4 / 4	6	2	8 / 29	34 / 22			
			85	Female	38.48	Acute + GP	GP Practice H	6 / 6	4	3	0 / 27	31 / 26			
			85	Female	38.42	Acute Only	GP Practice P	8 / 6	5	9	N/A / N/A	N/A / N/A			
			39	Male	38.32	Acute Only	GP Practice J	20 / 17	28	7	N/A / N/A	N/A / N/A			
			83	Male	38.31	Acute + GP	GP Practice O	9 / 6	3	8	9 / 43	27 / 14			
			75	Female	38.30	Acute + GP	GP Practice D	8 / 6	6	0	0 / 22	16 / 13			
			70	Female	38.23	Acute + GP	GP Practice D	7 / 4	4	11	3 / 25	31 / 21			
			55	Male	38.22	Acute + GP	GP Practice H	12 / 7	8	8	15 / 58	60 / 32			
			84	Male	38.13	Acute Only	GP Practice G	6 / 5	6	11	N/A / N/A	N/A / N/A			
			50	Female	38.13	Acute + GP	GP Practice A	11 / 5	17	4	12 / 64	18 / 13			
			59	Male	38.11	Acute + GP	GP Practice A	5 / 4	7	18	2 / 33	28 / 17			
			40	Male	38.07	Acute + GP	GP Practice D	11 / 6	34	14	12 / 29	9 / 6			
			90	Female	38.04	Acute + GP	GP Practice D	4 / 3	4	4	5 / 17	19 / 14			
			76	Male	37.92	Acute Only	GP Practice B	6 / 6	6	16	N/A / N/A	N/A / N/A			
			80	Male	37.89	Acute Only	GP Practice C	5 / 3	6	7	N/A / N/A	N/A / N/A			
			81	N/A	37.81	Acute Only	GP Practice I	7 / 4	3	4	N/A / N/A	N/A / N/A			
			73	Female	37.80	Acute + GP	GP Practice O	5 / 5	4	6	5 / 19	29 / 14			
			85	Male	37.77	Acute Only	GP Practice G	5 / 5	7	18	N/A / N/A	N/A / N/A			
			64	Female	37.76	Acute Only	GP Practice F	17 / 7	6	9	N/A / N/A	N/A / N/A			
			80	Male	37.76	Acute + GP	GP Practice H	17 / 3	1	15	8 / 46	23 / 11			
			80	Male	37.68	Acute Only	GP Practice J	4 / 3	3	15	N/A / N/A	N/A / N/A			

* Note: IP admits are shown in two columns—all admits and unplanned chronic (UC) admits.

** Note: Each drug with a distinct drug code is counted as a "unique" drug; therefore even if a new code just signifies a change in dose or dosage form it will be counted as "unique". The prescription count represents the total number of prescriptions for a unique medication ordered within the last 12 months.

Appendix 1

Burden of illness Method – Episode Treatment Groups(ETG)/Episode Risk Groups (ERG)™

As mentioned above in the list of independent variables, we used Acute Trust and GP Consultation information and GP medications history to form a Burden of Illness score that was formed an independent variable in our model. This was not achieved using a simple list of criteria based on clinical diagnosis and procedure codes. Rather, we used a commercial product: Episode Treatment Groups (ETG) to process clinical data and identify the diseases associated with each patient medical history. These episodes/diseases treatment groups are then rolled up into Episode Risk Groups that are in turn given weights in order to derive a total Burden of Illness score. The methodology behind the ETGs and ERGs is complex and has been discussed elsewhere, so we will only present a brief summary here.

The ETGprocessor analyses any set of clinical data and attempts to identify meaningful and discrete categories of illness that a patient's activity suggests. Data analysed by the ETG processor may include prescription medications, visits to specialists and hospital admissions, which ideally would include any diagnoses made and procedures performed. Data from multiple sources may be cross-referenced and form a single grouping of treatment related to a disease. For example, a blood glucose level lab test in GP consultation data alone would not necessarily assign an illness of diabetes to a patient, but if a diabetes related prescription was uncovered in the same patient's medications data, then diabetes would more likely be assigned. Likewise, multiple hospital admissions over the course of a specified time period are analysed and cross-checked for the presence of disease. In another hypothetical example, a patient readmitted for pneumonia within 90 days of a previous admission may be classified with either COPD, congestive heart failure, both conditions, or neither condition, depending on the totality of clinical data available.

ETG processing results in two basic types of files for our purposes: a patient-level file and an activity-level file with disease (i.e. ETG) flags. The patient-level file includes flags for every disease the processor found in a patient's data and is used mostly for reporting. These flags from the activity level file are feed into the ERG processor which uses them to form "risk groups" associated with the patient. Since some GP practices in the PCTs did not (or could not) submit patient activity data, we split patient illnesses uncovered in the GP consultation and medication data from the ones uncovered in the Acute Trust data, and tested them separately in two sub-models. We then combined the results from these two sub-models, along with the sub-model based on specialist data into the final BoI score, where we had patients from participating GP practices. For those patients without GP consultation data, we simply used the BOI constructed from their Acute Trust data and specialist data.

The activity-level file in the target period generated by the ETG processor was used in conjunction with a list of reference conditions to flag inpatient admissions as being related to a chronic disease (hereafter referred to as unplanned chronic admissions). ETG processing provided additional discrimination between hospital admissions due to uncontrollable acute events (such as sudden acute illness or accidents), and hospital admissions due to ongoing chronic conditions. This enabled us to narrow our outcome variable from any unplanned admission down to any unplanned admission related to chronic disease