# Predictive Modelling Pilot Project

## 1. Introduction

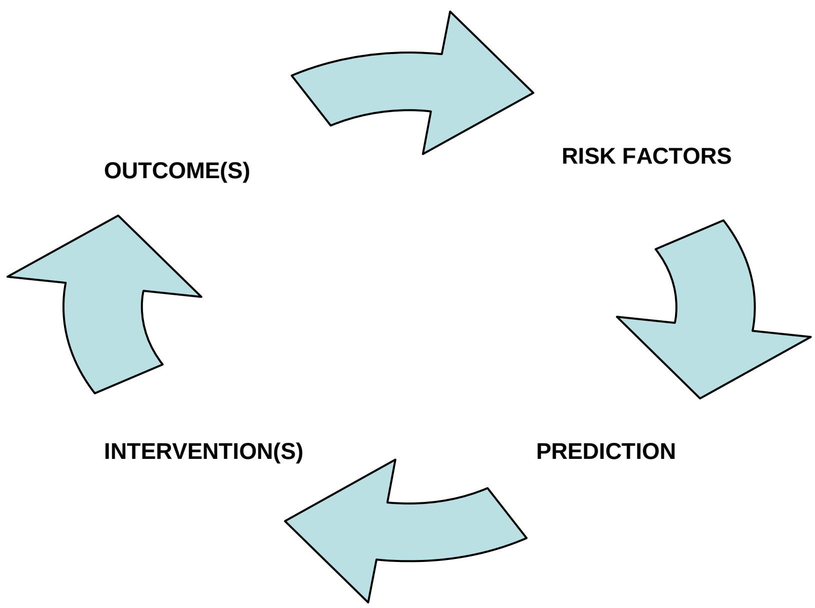
The Long Term Conditions QIPP (quality, innovation, productivity and prevention) workstream seeks to improve clinical outcomes and experience for patients with long term conditions in England. The workstream focuses on improving the quality and productivity of services for these patients and their carers so they can access higher quality, local, comprehensive community and primary care. This will in turn, aim to slow disease progression and reduce the need for unscheduled acute admissions by supporting people to understand and manage their own conditions. The model of care for Long Term Conditions QIPP is based on the following 3 key principles:

1. Risk profiling;
2. Neighbourhood teams and;
3. Self care / shared decision making.

The aim of this project is to develop and pilot an approach to risk profiling that will identify patients that are at risk of emergency admission before they are admitted, i.e. identify patients for targeted interventions to reduce admissions. Predicting the risk of an outcome is a cyclical process that incorporate patients risks, the prediction of admission, the intervention (treatment / care) they receive and the outcome that results given their risks and whether they received or did not receive an intervention (Diagram 1). Therefore, by using this approach the predictive model has the potential to:

1. Identify patients at high risk of admission;
2. Determine those patients that are likely to benefit from the interventions that are provided (impactible patient group) and;
3. Can provide an approach to ongoing evaluation of interventions and models of care.

**<u>Diagram 1.1</u> – Process of predicting and intervening to reduce risk**






Chris Harwood, Intelligence Manager
Andy Poole, Senior Analyst

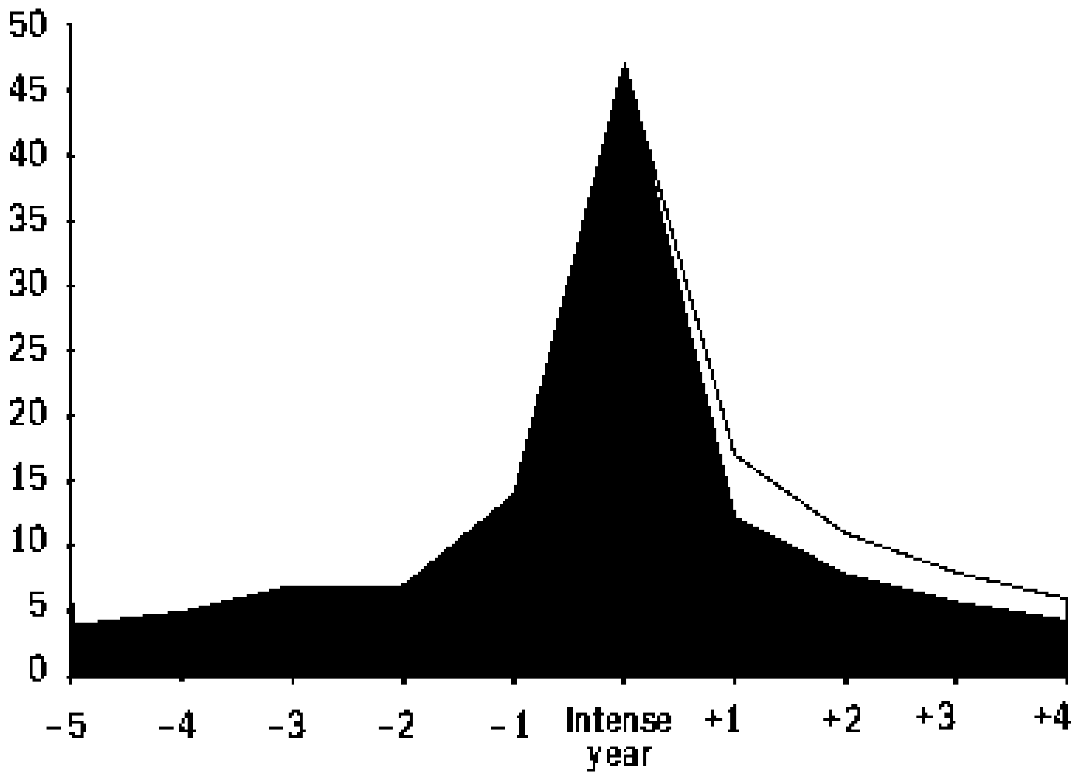
## 2. Background – Predictive models

Predictive models in healthcare use data such as patient demographics, health records and health care utilisation to find people who fall into specific categories of risk in order to 'predict' their likely future health risk. These models also produce profiles of risk for different types of people as well as model future consumption of health care resources. Models can also identify the future implications of early interventions to achieve optimal future health outcomes. The predictive model in this pilot aims to identify patients at risk of future admission.

### 2.1. Predicting risk

To predict the risk of someone being admitted to hospital in any time period, a statistical phenomenon known as "regression to the mean" must be considered. For example, patients who are currently experiencing repeated hospital admissions will, on average, have markedly fewer hospital admissions in the future, even without an intervention (Gravelle et al., 2007; Roland et al., 2005). Therefore, prevention programmes aimed at targeting patients already at risk of hospital admissions will appear to be successful, but most of the reduction would have occurred anyway (Figure 1).

**Figure 1.1: Average number of emergency bed days, targeting patients at risk (white is the improvement from targeting high risk patients)**

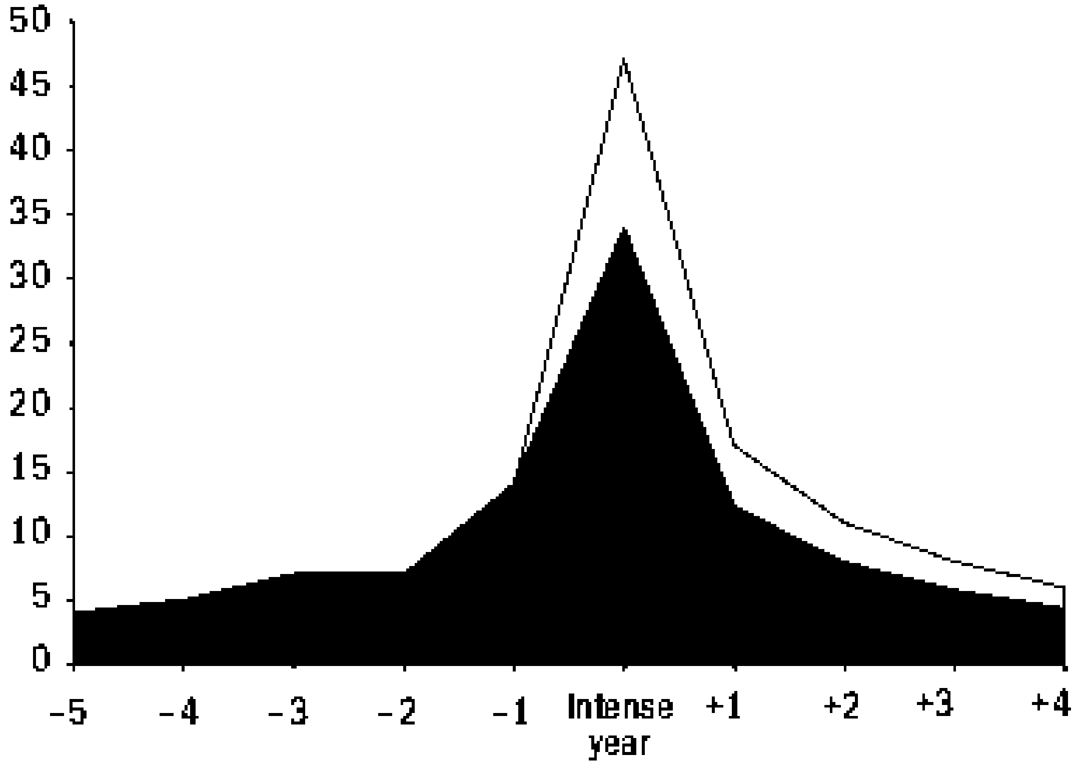



Source: Roger Halliday, Department of Health

### 2.2. Dealing with Regression to the Mean

Predictive models aim to reduce the problem of regression to the mean by identifying patients who will be at risk of admission or readmission in a period (e.g. 12 months) after prediction (Cousins et al., 2002). The use of techniques such as logistic regression or neural networks, analyse patterns in historical routine data to make predictions that reflect an individual's risk of emergency admission in a time period (e.g. 12 months) after prediction. Individuals that are identified as being at risk of emergency admission can then be targeted for interventions aimed at reducing or avoiding the need for an admission and a greater impact could be possible (Figure 2).

**Figure 1.2: Average number of emergency bed days, targeting patients before they were high risk (white is the improvement from targeting patients using predictive models)**



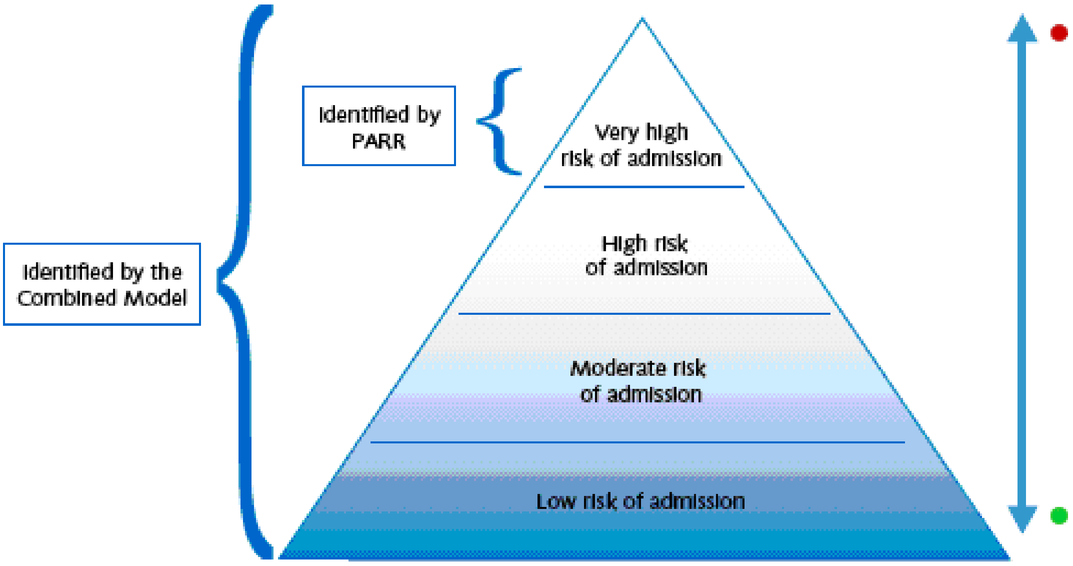

Source: Roger Halliday, Department of Health

## 2.3. PARR and Combined Predictive Model

The Department of Health commissioned two predictive models for the NHS in England, the Patients At Risk of Readmission (PARR) tool and the Combined Predictive Model (The Kings Fund, 2009). They used a broad range of variables to predict future emergency admissions, splitting them in to risk categories (Diagram 1). The data used includes demographics (age, sex and deprivation), diagnoses, frequency of admission and use of other health services (A&E and outpatients), and presence of chronic conditions. There are a number of issues with both models which mean they are not as predictive as a locally derived model. Firstly, the models were created using data from 2006/07 and 2007/08 where patterns of health and risk of admission were different from today. Secondly, they were created using data from three PCTs that were likely to have different system, pathways of care and population characteristics compared to local using data. Finally they don't allow for new data to be incorporated that will make prediction more accurate, learn new patterns or test hypotheses about the effectiveness of different treatments or interventions to reduce admissions.

This project will develop a predictive model that will initially enable primary care to target patients at high risk of admission using local data and a locally derived algorithm. The algorithm will be adapted over time to improve accuracy as well as develop "impactibility models", which identify patients who are not just at risk of admission but are most likely to benefit from targeted interventions (Lewis, 2010).

**<u>Diagram 1.1</u> – Segmentation of the population using the PARR and CPM**




Source: Kings Fund, 2009

## 2.4. <u>Aims and objectives</u>

The aim of this pilot project was to create a predictive model that identifies patients at risk of admission in a period following prediction (e.g. 12 months) for the Wirral population. The objectives were:

1. To develop a model to predict emergency admissions using primary and secondary care data from a pilot practice;
2. To develop a model using demographic and secondary care data for the Wirral population;
3. To validate the predictability (sensitivity and positive predictive value analysis) of the different models;
4. To demonstrate the value of a local solution compared to national models such as the PARR and CPM;
5. To produce a rank ordered list of patients who are most at risk of admission in the 12 months following prediction.
6. To develop data mining software within Wirral data warehouse.

**N.B.** At this stage what the model does not do is identify patients that will benefit from preventative or treatment interventions in the 12 months following prediction. The next stage of the project will start to develop **"impactibility models"**.

## 3. Method

### 3.1. Predictive models

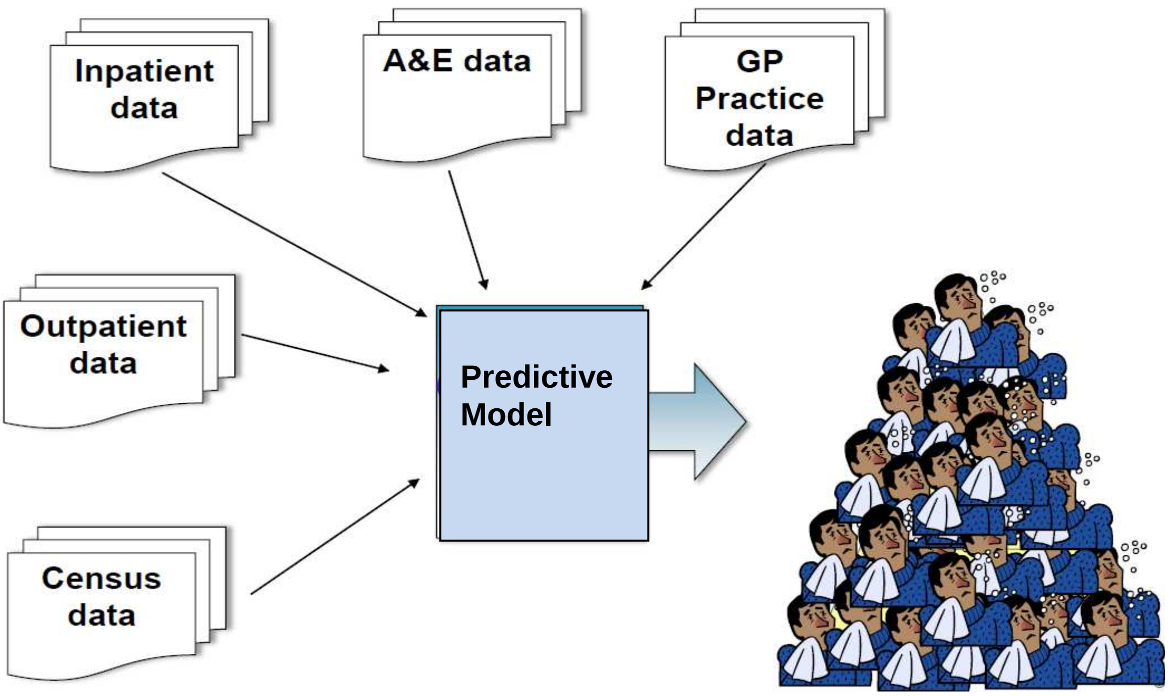
Predictive models use relationships between a set of variables in order to predict future outcomes. They use formulae which interpret past data and make forecasts about the future and map associations and statistical relationships to specific targets. They predict the risk of emergency admission based on these identified associations (Cousins et al., 2002).

The models developed in this pilot have used a number of algorithms to help predict the risk of emergency admission, including, logistic regression, neural networks, decision tree and naïve bayes. The merits or each algorithm are beyond the scope of this pilot but each algorithm has been tested to determine the validity and reliability of predicting future emergency admissions. The model takes primary and secondary care data for an entire patient population and stratifies those patients based upon their risk of emergency admission in the next 12 months.

### 3.2. Data collection

The data used in the model includes population (age, sex and deprivation), outpatient data, inpatient data, A&E data and GP practice data such as disease prevalence, prescribing and demographic data (Diagram 2). All data is required to be at the individual patient level to be able to predict the risk of emergency admission. The variables included were from the CPM with additions from a literature review and clinical advice as to the important variables related to the risk of emergency admission.

**Diagram 2.1: Data included in the predictive model**




Source: Lewis, 2010

## 3.3. Data analysis

The data analysis starts with defining the problem and involves a cyclical process with the aim of finding and using new knowledge (Diagram 2):

i. Data preparation – collecting, evaluating, and integrating patient's health status, clinical history or demographic data;

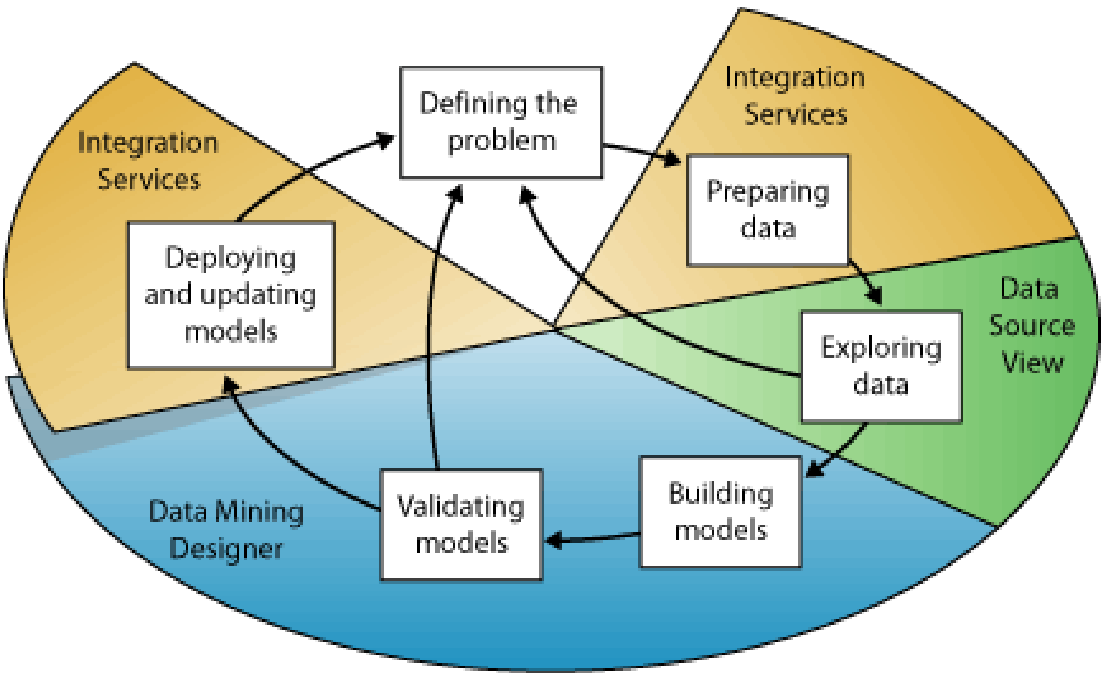
ii. Exploring the data – applying grouping and risk marker identification algorithms, episodes of care and categories of pharmaceutical treatments, e.g. utilised A&E in the past 30 days;

iii. Building Models – summarising the presence or absence of clinical risk markers and applying the algorithm to the data to identify the risk score for patients based upon the known admission profile of patients;

iv. Exploring and validating the models – Before deploying a model in a production environment it needs to be tested to see how predictive and valid it is. This will involve testing the sensitivity, specificity and the predictive power of the model (ROC curve).

v. Deploying and updating models – assessing overall risk by adding assigned risk weights of all markers to patients to identify their future risk of emergency admission. Updating a model starts at data preparation stage.

**Diagram 2.2: The process of data mining and predictive modelling**



# 4. Results

## 4.1. Pilot Practice

### 4.1.1. Baseline data

The pilot practice included a total population of 9,800. The population used in the model were aged 50 years and over, which included 3596 people. The practice has a slightly younger age profile, similar gender profile, but more deprived population compared to the Wirral as a whole (Table 1).

**Table 3.1: Practice data, demographic and chronic diseases**

| Criteria | Variable | Practice | | Wirral | |
|---|---|---|---|---|---|
| | | Persons | Percent | Persons | Percent |
| Age group (Years) | 50-59 | 1352 | 13.8% | 42727 | 13.4% |
| | 60-69 | 1108 | 11.3% | 43030 | 13.5% |
| | 70-79 | 706 | 7.2% | 26234 | 8.2% |
| | 80-89 | 356 | 3.6% | 14892 | 4.7% |
| | 90+ | 74 | 0.8% | 2921 | 0.9% |
| Gender | Men | 1689 | 47.0% | 58000 | 46.1% |
| | Women | 1909 | 53.1% | 67900 | 53.9% |
| Chronic diseases* | 0 | 2388 | 66.4% | | |
| | 1 | 909 | 25.3% | | |
| | 2 | 233 | 6.5% | | |
| | 3+ | 66 | 1.8% | | |
| IMD 2010** | 1 | 883 | 24.6% | 30717 | 25.4% |
| | 2 | 1576 | 43.8% | 20004 | 16.5% |
| | 3 | 319 | 8.9% | 18946 | 15.6% |
| | 4 | 573 | 15.9% | 23426 | 19.3% |
| | 5 | 245 | 6.8% | 28016 | 23.1% |

* Chronic disease = Coronary heart disease, cancer, hypertension, COPD, heart failure >>>
** 1 = most deprived
Source: Pilot practice and SUS data, 2011/12

### 4.1.2. Risk factors for emergency admission

#### 4.1.2.1. Demographics and practice data

Comparing some of the demographic data and emergency admissions for 2011/12, highlight a number of emerging risk factors (Table 3.2). The risk of admission for someone 90 years and over was nearly five times higher than for people age 60-69 years. People with 3 or more chronic conditions were 8 times more likely to be admitted for an emergency than someone without a chronic condition. At the univariate level there was no pattern between deprivation and emergency admissions although there may be an interaction with age and chronic conditions.

**Table 3.2 – Demographics and emergency admission, 2011/12**

| Criteria | Variable | Not admit | Admit | All | Percent |
|---|---|---|---|---|---|
| Age group (Years) | 50-59 | 1258 | 94 | 1352 | 7.0% |
| | 60-69 | 1044 | 64 | 1108 | 5.8% |
| | 70-79 | 624 | 82 | 706 | 11.6% |
| | 80-89 | 283 | 73 | 356 | 20.5% |
| | 90+ | 53 | 21 | 74 | 28.4% |
| Gender | Male | 1533 | 154 | 1687 | 9.1% |
| | Female | 1729 | 180 | 1909 | 9.4% |
| Chronic diseases* | 0 | 2267 | 121 | 2388 | 5.1% |
| | 1 | 789 | 120 | 909 | 13.2% |
| | 2 | 167 | 66 | 233 | 28.3% |
| | 3+ | 39 | 27 | 66 | 40.9% |
| IMD 2010** | 1 | 783 | 100 | 883 | 11.3% |
| | 2 | 1430 | 146 | 1576 | 9.3% |
| | 3 | 295 | 24 | 319 | 7.5% |
| | 4 | 534 | 39 | 573 | 6.8% |
| | 5 | 220 | 25 | 245 | 10.2% |

* Chronic disease = Coronary heart disease, cancer, hypertension, COPD, heart failure >>>
** 1 = most deprived
Source: Pilot practice and SUS data, 2011/12

#### 4.1.2.2. Hospital utilisation data

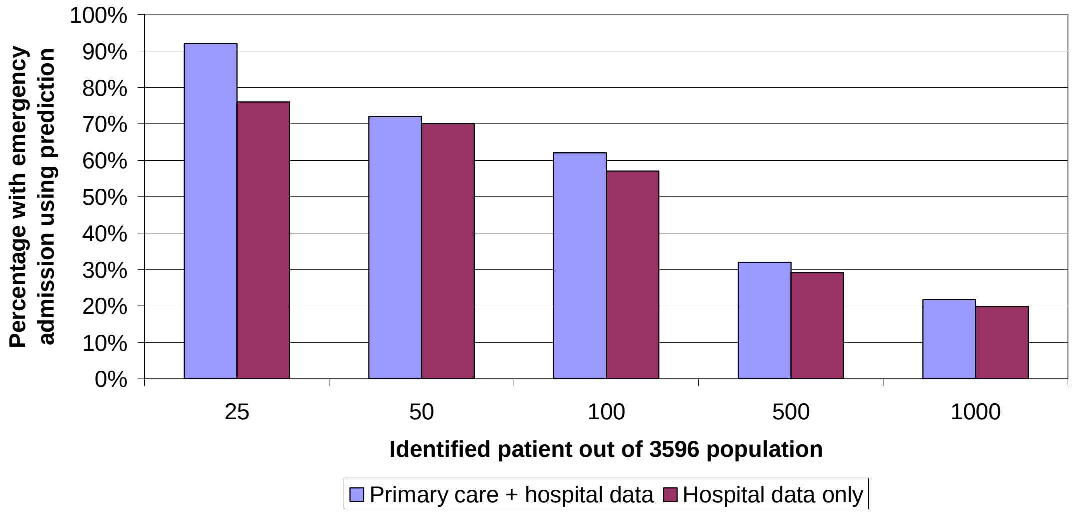
Hospital utilisation data is strongly associated with the risk of emergency admission in the next 12 months. Univariate analysis suggests that A&E attendance data were most strongly associated with the prediction of risk of emergency admission in the following 12 month (Appendix 1).

### 4.1.3. Predictive Model

The Positive Predictive Value (PPV) is a reflection of the number of patients who actually had an emergency admission in the year following prediction. Figure 3.1 shows patients that have been identified in different risk segments of the population. For example 62 out of the top 100 patients predicted by the primary care and hospital model (combined model) were admitted for an emergency.

The primary care and hospital model combined was more predictive than the hospital only model for each segment of the risk population. For the top 25 very high risk patients the combined model predicted 92% of admission correctly compared to 76% accuracy for the hospital data only model (Figure 3.1).

**Figure 3.1: Positive Predictive Value for emergency admissions**




Source: Pilot practice and SUS data, 2009/10, 2010/11 and 2011/12

#### 4.1.3.1. Measuring the statistical significance of models

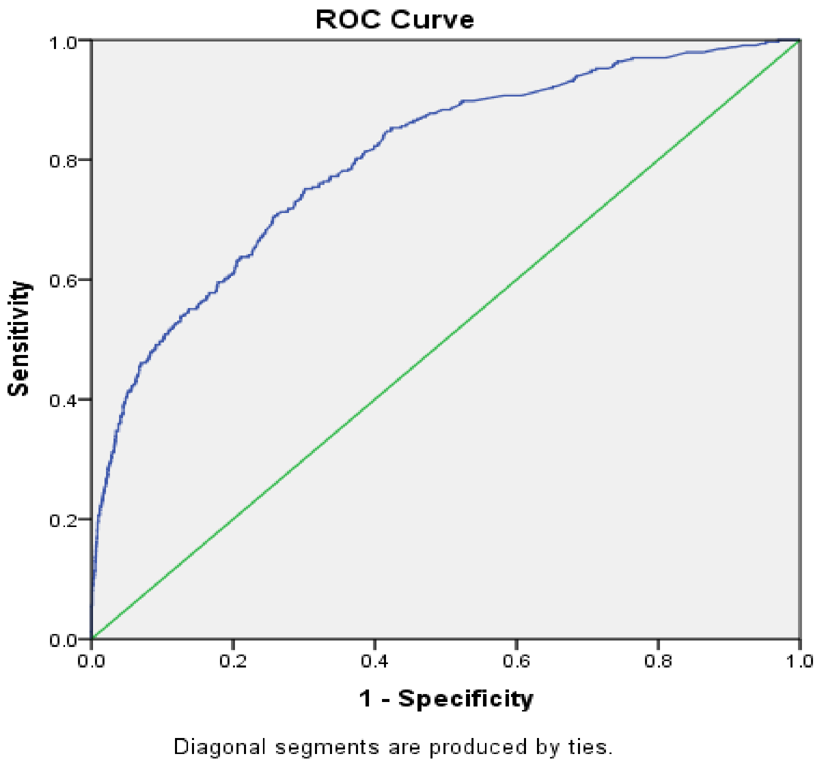
A Receiver Operating Curves (ROC) plot the true positives (sensitivity) vs. false positives (1 − specificity), for a binary classifier system as its discrimination threshold is varied (Figure 3.2). Since, a random method describes a horizontal curve through the unit interval, it has an area under the curve of 0.5. The area under the curve for the Combined model and Hospital only model are significantly more accurate than randomly guessing whether patients are at risk of emergency admission. The Combined model is significantly more accurate than the Hospital only model (Table 3.3).

**Table 3.3: Area under the curve, Combined and Hospital only models**

| Model | Area | Std. Error[a] | Sig.[b] | 95% Confidence Interval | |
|---|---|---|---|---|---|
| | | | | Lower | Upper |
| Primary care + hospital model | .800 | .013 | .000 | .774 | .826 |
| Hospital model only | .729 | .016 | .000 | .697 | .761 |

a. Under the nonparametric assumption
b. Null hypothesis: true area = 0.5

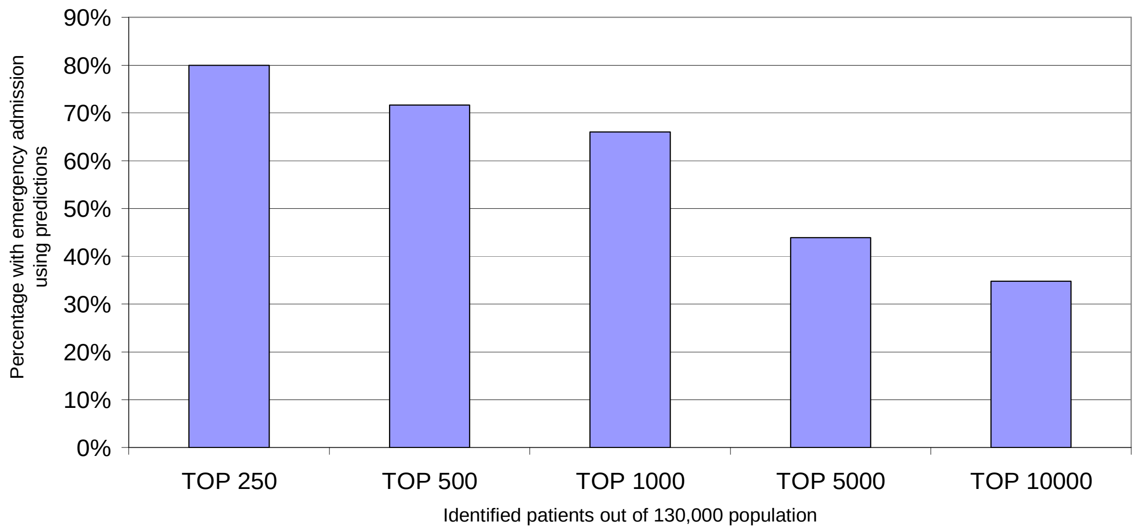
**Figure 3.2: ROC curve for the Combined model**



## 4.2. Wirral model (Hospital data only)

The Wirral model uses the same methods as the pilot practice hospital data only but includes all the Wirral population aged 50 years and over. The model identifies the population for different levels of risk of emergency admission (Figure 3.3). For the high risk patients 200 out of the top 250 patients predicted by the model were admitted for an emergency (PPV = 80%). Out of the top 1000 patients the model predicted 660 were emergency admission (PPV = 66%).

**Figure 3.3: Positive Predictive Value for emergency admissions**




### 4.2.1. Measuring the statistical significance of the Wirral model

The area under the curve for the Wirral model is significantly more accurate than randomly guessing whether patients are at risk of emergency admission (Table 3.3).

**Table 3.3: Area under the curve, Wirral model**

| Model | Area | Std. Error[a] | Sig.[b] | 95% Confidence Interval | |
|---|---|---|---|---|---|
| | | | | Lower | Upper |
| Wirral hospital model | .744 | .002 | 0.000 | .739 | .749 |

a. Under the nonparametric assumption
b. Null hypothesis: true area = 0.5

## 5. Discussion

The results of the pilot project show that the predictive models are accurate at identifying patients at high risk of emergency admission 12 months following prediction. The models have stratified the risk of admission for the whole population and produced a list of high risk patients that can be targeted for admission prevention. The pilot also illustrates that the predictions are more accurate if primary care data is included, even though the numbers of patients in the pilot are small. In addition, the use of SQL server to run the predictive model algorithms means that the data is stored in a safe environment where multiple sources of data can be used to improve the accuracy of the models.

## 6. Next Steps

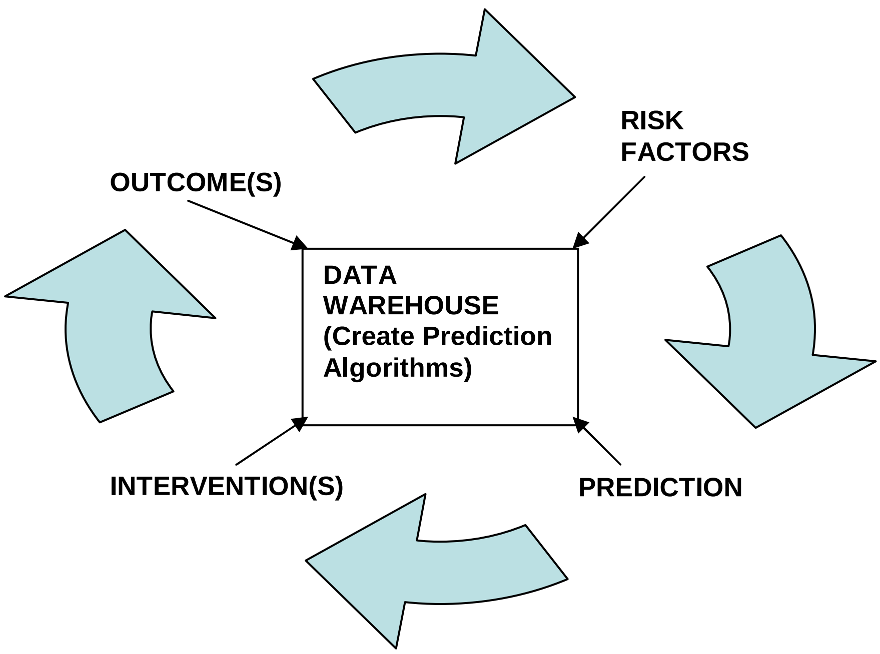
### 6.1. Validation

The next step in the project is to continue to update and deploy the models to identify high risk patients. This will involve gathering and using more data to improve the accuracy of the models as well as validate and test the algorithms against other sources of data.

### 6.2. Impactible Models

Work to produce models that can identify "impactible" patient groups (i.e. identify patients at risk of admission but are most likely to benefit from admission prevention) will need further development. To achieve this objective patient level data that links the predictive risk model to the treatment and interventions that patients receive will need to be included into the data warehouse (Diagram 6.1).

Diagram 6.1: The process of identifying "impactible" patient groups



## 7. References

Billing, J., and Mijanovich, T. (2007). Improving the Management of Care for High Cost Medicaid Patients. Health Affairs, 26(6): 1643-1654.

Cousins MS. (2002). Introduction to predictive modelling. Shickle LM and Bander JA. Disease Management 5[3].

The King's Fund, (2009): [online] http://www.kingsfund.org.uk/research/projects/predicting_and_reducing_readmission_to_hospital/

Gravelle, H., Dusheiko, M., Sheaff, R. (2007). Impact of care management (Evercare) on Frail Elderly Patients: Controlled Before and After Analysis for Quantitative Outcome Data. BMJ, 334(7583): 31.

Roland, M., Dusheiko, M., Gravelle, H. (2005). Follow up of People Aged 65 and Over with a History of Emergency Admissions: Analysis of Routine Admission Data. BMJ, 330(7486): 289-292.

Lewis, G. (2011). Predictive Modelling in Action: How 'Virtual Wards' Help High Risk Patients Receive Hospital Care at Home. The Commonwealth Fund, Volume 94.

## Appendix 1: Hospital utilisation and emergency admissions

| Criteria | Variable | Not admit | Admit | All | Percent |
|---|---|---|---|---|---|
| A&E | 1+ AEattendance030days | 24 | 20 | 44 | 45.5% |
| | 1+ AEattendance3190days | 50 | 28 | 78 | 35.9% |
| | 1+ AEattendance91180days | 82 | 38 | 120 | 31.7% |
| | 1+ AEattendance181365days | 176 | 58 | 234 | 24.8% |
| | 1+ AEattendance366730days | 304 | 101 | 405 | 24.9% |
| | 1+ AEAttAmbulance030days | 12 | 17 | 29 | 58.6% |
| | 1+ AEAttAmbulance3190days | 27 | 20 | 47 | 42.6% |
| | 1+ AEAttAmbulance91180days | 44 | 32 | 76 | 42.1% |
| | 1+ AEAttAmbulance181365days | 64 | 35 | 99 | 35.4% |
| | 1+ AEAttAmbulance366730days | 135 | 70 | 205 | 34.1% |
| | 1+ AEAttInvestigation030days | 23 | 20 | 43 | 46.5% |
| | 1+ AEAttInvestigation3190days | 40 | 23 | 63 | 36.5% |
| | 1+ AEAttInvestigation91180days | 60 | 32 | 92 | 34.8% |
| | 1+ AEAttInvestigation181365days | 129 | 45 | 174 | 25.9% |
| | 1+ AEAttInvestigation366730days | 206 | 91 | 297 | 30.6% |
| Inpatient | 1+ IPEmergencyAdm030days | 20 | 13 | 33 | 39.4% |
| | 1+ IPEmergencyAdm3190days | 35 | 28 | 63 | 44.4% |
| | 1+ IPEmergencyAdm91180days | 51 | 32 | 83 | 38.6% |
| | 1+ IPEmergencyAdm181365days | 110 | 48 | 158 | 30.4% |
| | 1+ IPEmergencyAdm366730days | 185 | 93 | 278 | 33.5% |
| | 1+ IPAdmandDischHome030days | 68 | 17 | 85 | 20.0% |
| | 1+ IPAdmandDischHome3190days | 104 | 31 | 135 | 23.0% |
| | 1+ IPAdmandDischHome91180days | 176 | 50 | 226 | 22.1% |
| | 1+ IPAdmandDischHome181365days | 340 | 76 | 416 | 18.3% |
| | 1+ IPAdmandDischHome366730days | 519 | 143 | 662 | 21.6% |
| | 1+ IPAdmandProc030days | 53 | 8 | 61 | 13.1% |
| | 1+ IPAdmandProc3190days | 85 | 16 | 101 | 15.8% |
| | 1+ IPAdmandProc91180days | 144 | 30 | 174 | 17.2% |
| | 1+ IPAdmandProc181365days | 286 | 54 | 340 | 15.9% |
| | 1+ IPAdmandProc366730days | 426 | 101 | 527 | 19.2% |
| Outpatient | 1 OutpatientVisit030days | 234 | 36 | 270 | 13.3% |
| | 2 OutpatientVisit030days | 56 | 14 | 70 | 20.0% |
| | 3 OutpatientVisit030days | 16 | 7 | 23 | 30.4% |
| | 1 OutpatientVisit3190days | 337 | 65 | 402 | 16.2% |
| | 2 OutpatientVisit3190days | 85 | 14 | 99 | 14.1% |
| | 3 OutpatientVisit3190days | 44 | 12 | 56 | 21.4% |
| | 1 OutpatientVisit91180days | 348 | 68 | 416 | 16.3% |
| | 2 OutpatientVisit91180days | 116 | 23 | 139 | 16.5% |
| | 3 OutpatientVisit91180days | 93 | 18 | 111 | 16.2% |
| | 1-5 OutpatientVisit181365days | 886 | 154 | 1040 | 14.8% |
| | 6+ OutpatientVisit181365days | 78 | 17 | 95 | 17.9% |
| | 6-10 OutpatientVisit181365days | 65 | 11 | 76 | 14.5% |
| | 11+ OutpatientVisit181365days | 13 | 6 | 19 | 31.6% |

| Variable | Not admit | Admit | All | Percent |
|---|---|---|---|---|
| 15+ OutpatientVisit366730days | 997 | 158 | 1155 | 13.7% |
| 6-10 OutpatientVisit366730days | 119 | 33 | 152 | 21.7% |
| 11+ OutpatientVisit366730days | 55 | 15 | 70 | 21.4% |
| 1+ OutpatientVisitreferAE030days | 75 | 20 | 95 | 21.1% |
| 1+ OutpatientVisitreferAE3190days | 116 | 21 | 137 | 15.3% |
| 1+ OutpatientVisitreferAE91180days | 112 | 28 | 140 | 20.0% |
| 1+ OutpatientVisitreferAE181365days | 191 | 48 | 239 | 20.1% |
| 1+ OutpatientVisitreferAE366730days | 229 | 46 | 275 | 16.7% |
| 1+ OutpatientVisitreferNotAE030days | 248 | 42 | 290 | 14.5% |
| 1+ OutpatientVisitreferNotAE3190days | 389 | 75 | 464 | 16.2% |
| 1+ OutpatientVisitreferNotAE91180days | 478 | 90 | 568 | 15.8% |
| 1+ OutpatientVisitreferNotAE181365days | 884 | 156 | 1040 | 15.0% |
| 1+ OutpatientVisitreferNotAE366730days | 1107 | 196 | 1303 | 15.0% |
| 1 OutpatientVisitandProcedure030days | 40 | 4 | 44 | 9.1% |
| 1 OutpatientVisitandProcedure3190days | 68 | 12 | 80 | 15.0% |
| 1 OutpatientVisitandProcedure91180days | 65 | 10 | 75 | 13.3% |
| 1 OutpatientVisitandProcedure181365days | 124 | 12 | 136 | 8.8% |
| 1 OutpatientVisitandProcedure366730days | 172 | 25 | 197 | 12.7% |