HealthNumerics-RISC® Predictive Models

A Successful Approach to Risk Stratification

Contents

Section 1. Introduction

Risk stratification, the measurement of the expected health care utilisation of an individual or population, enables health care organisations to understand and predict the health risks of their registered population. Whether to obtain meaningful comparisons of provider performance, or identify patients of highest risk, sound methods of risk stratification are critical tools for any health care organisation.

The Optum risk stratification tool for the NHS, HealthNumerics-RISC®, predicts the risk of patients having unplanned chronic admissions in a 12 month period. HealthNumerics-RISC® is a dynamic, easy to use tool that takes data from multiple sources including primary and secondary care, to perform a risk stratification of the entire population of a health care organisation.

HealthNumerics-RISC® supports proactive management of the health of a population, targeting resources on the patients with the greatest need, and allows for prioritisation of community based preventative care and aids in the development of strategies to reduce emergency admissions. It provides timely, comprehensive information to clinicians to support interventions for those patients most in need, ensuring better long term patient outcomes whilst driving down costs. Given current cost pressures and the shifts in patterns of disease, understanding local patterns is becoming increasingly important.

HealthNumerics-RISC® can be used to identify patients with long term conditions for potential case management and for commissioners to analyse care pathways.

Risk stratification and its practical applications are only as good as the underlying predictive models. The models in HealthNumerics-RISC® have all the features of a valid, reliable risk stratification model:

- **Accuracy** – provides accurate forecasts of health risk;
- **Transparency** – supports users' ability to validate and explain predictions at the individual and group level;
- **Interoperability** – integrates with other clinical systems;
- **Flexibility** – includes options that best meet timing and operational needs within the limits of available data; and
- **Industry Acceptance** – the broad acceptance of a predictive model provides evidence of its credibility in the marketplace.

The models in HealthNumerics-RISC® have a key feature that sets them apart from other risk stratification models – condition identification based on Episode Treatment Groups (ETG®), the market-leading disease classification software in the United States. The basic concept of ETG® is to combine activity from multiple settings across time to create a single unit of analysis with respect to a condition. For example, the diabetes ETG® includes all GP visits, medications and acute services relevant to diabetes. The clinical richness of ETG®, its reliance on information readily available in GP and hospital systems, and its potential as a tool for describing relative patient morbidity makes it a sound basis for the development of risk stratification models. The focus in the HealthNumerics-RISC® models is therefore on the key information describing a patient's underlying medical condition (and demographics), rather than the individual services provided in its treatment.

This paper describes the methodology used by Optum to re-calibrate the models in 2013, an assessment of the performance of the models, further information on the HealthNumerics-RISC® tool and the policy implications surrounding risk stratification. As part of this latest re-calibration, two new 'short-term' models were developed to run alongside the existing 12 month models to allow users to more proactively manage patients in the near term. These are also described within this paper.

Section 2. Model Development

The HealthNumerics-RISC® predictive models were developed using a sample population of 1,870,468 patients across five Primary Care Trusts (PCTs). Two years of Secondary and Primary Care data was utilised totalling more than 70 million records.

Prior to this re-calibration, HealthNumerics-RISC® has operated using two 12 month models; an 'Acute Only' model and an 'Acute + GP' model. The 'Acute Only' uses data from acute hospital providers – Inpatient, Outpatient and Accident & Emergency data. The 'Acute + GP' model uses acute hospital provider data together with GP diagnosis, procedure and prescription data. The development work involved re-building the two current predictive models and building two new 'short-term' models.

The models were developed using logistic regression to predict the likelihood of an unplanned (emergency) inpatient admission due to a chronic condition as a function of the independent variables. The models were refined by repeating the regression review process through several regression iterations with each iteration adding a new set of variables.

The regression iteration results for each model were reviewed to determine if the markers indicated a significant predictor of unplanned admissions due to a chronic condition. Markers with a strong indication of prediction were included in the model. Markers with very low frequencies or with low statistical predictive significance were removed from the next regression iteration.

The final regression iteration for each model resulted in the final weight set identified for each model.

2.1 Model Data Types

The following six datasets were utilised for each of the five PCTS:

- Patient Registry
- Admitted Patient Care Episode (APC)
- Accident & Emergency (A&E)
- Outpatient (OP)
- GP Medication (GPRX)
- GP Diagnosis and Procedures (GP)

2.2 PCT Datasets

The five PCTs in Table 1 were chosen to use as the source of the model data, since they represented the most diverse organisations either geographically or by classification that were available to us. All data was anonymised and encrypted before its use in the model development work.

Table 1 PCT Membership Distribution and ONS Area Classifications

PCT Code	PCT Name	Cluster Group	Total	%
5H8	Rotherham	Mining and Manufacturing	259,801	13.9
5LD	Lambeth	London Cosmopolitan	438,209	23.4
5LE	Southwark	London Cosmopolitan	369,134	19.7
5QM	Dorset	Coastal and Countryside	409,450	21.9
5QN	Bournemouth	Cities and Services	393,874	21.1
Total			**1,870,468**	**100%**

Table 2 shows the registered patient and activity distribution by dataset type for all 5 PCT organisations.

Table 2 Record Distribution by PCT

	Dataset					
	Patient Registry	Inpatient	A&E Attendances	Outpatient Attendances	GP Diagnosis & Procedures	GP Medications
5H8 Rotherham						
Count	259,801	215,165	159,894	986,402	13,124,272	5,674,865
Annual rate per 1000	–	414	308	1,898	25,258	10,922
%	13.89	15.97	16.95	18.02	32.65	22.48
5LD Lambeth						
Count	438,209	216,391	211,323	1,161,548	7,831,160	4,113,526
Annual rate per 1000	–	247	241	1,325	8,935	4,694
%	23.43	16.07	22.4	21.22	19.48	16.29
5LE Southwark						
Count	369,134	203,470	210,250	1,058,639	7,508,504	4,004,070
Annual rate per 1000	–	276	285	1,434	10,170	5,424
%	19.73	15.11	22.3	19.33	18.68	15.86
5QM Dorset						
Count	409,450	376,590	185,675	1,359,447	3,700,643	3,331,266
Annual rate per 1000	–	460	227	1,660	4,519	4,068
%	21.89	27.96	19.68	24.83	9.21	13.2
5QN Bournemouth						
Count	393,874	335,310	176,124	909,019	8,032,821	8,122,835
Annual rate per 1000	–	426	224	1,154	10,197	10,311
%	21.06	24.89	18.67	16.6	19.98	32.17
Total	**1,870,468**	**1,346,926**	**943,266**	**5,475,055**	**40,197,400**	**25,246,562**

Note: Rates are per annum

2.3 The Models

The following 12 month models were re-built:

- Risk of unplanned chronic admissions (12 month) – acute only model
- Risk of unplanned chronic admissions (12 month) – acute + GP model

The following new 3 month ('short-term') models were also built:

- Risk of unplanned chronic admissions (3 months) – acute only model
- Risk of unplanned chronic admissions (3 months) – acute + GP model

2.4 Model Timings

The models were based on data in the following time periods:

3 month models: allows for a 1 month lag in data due to processing intake

- Experience (Base) Year: 01/06/2010 – 31/05/2011
- Prediction (Target) Period: 01/07/2011 – 30/09/2011

12 month models:

- Experience (Base) Year: 01/06/2010 – 31/05/2011
- Prediction (Target) Year: 01/06/2011 – 31/05/2012

The discharge date in the APC extract was used to determine if episode data for a patient was included within the experience year. For example, an episode was included if a patient entered the hospital prior to the experience year but was discharged within the experience year.

2.5 Model Development Approach

The model development involves five important steps:

1. **Create ETG Episodes of Care.** Underpinning all the predictive models in HealthNumerics-RISC® is a condition classification methodology called Episode Treatment Groups (ETG®). The software identifies and combines related services – both diagnostic and procedural – into medically relevant units describing complete episodes of care. Each patient will have an ETG for each of his or her episodes of care during the review period. The clinical richness of ETG, its reliance on information readily available in GP and hospital systems, and its potential as a tool for describing relative patient morbidity makes it a sound basis for the development of risk stratification models.
2. **Map ETGs to ERGs.** Episodes are further categorised into one of 189 episode risk groups (ERGs). The ERGs combine each patient's ETGs of similar clinical and risk characteristics. The ERGs must be distinct; in the case of a patient having related ERGs, a hierarchy is applied to determine which is most clinically meaningful. A patient can be assigned zero, one, or – if s/he has multiple medical conditions – multiple ERGs. Each ERG is a marker (independent variable) in the predictive models.
3. **Create Demographic and Custom Markers.** Additional demographic and utilisation markers (independent variables) are created using patient registry, inpatient, outpatient and A&E data to enhance the models further.
4. **Develop Patient's Risk Profiles.** Age, gender and mix of ERG and custom markers provide a clinical and demographic risk profile for a patient.
5. **Create Patient Risk Scores.** A patient's risk score is computed by summing the predetermined weights attached to each ERG and to the demographic and utilisation characteristics observed in his or her profile.

Step 1. Create ETG Episodes of Care

Using routinely collected secondary care and primary care data as input, the ETG® software captures the relevant services provided during the course of a patient's treatment, and organises the data into meaningful episodes of care. The result is the accurate identification of clinically homogenous episodes of care, regardless of treatment location or duration. Because ETG® is intuitive and adjusts for clinical severity, it can be utilised in diverse applications integral to the success of clinicians, health care researchers, and administrators.

The basic concept of ETG® is to combine activity from multiple settings across time to create a single unit of analysis with respect to a condition. For example, the diabetes ETG® includes all GP visits, medications and acute services relevant to diabetes.

Each health care activity record is assigned to an ETG® which describes the condition being treated. Building episodes is a complex and sometimes subtle process, consisting of these steps:

- Determine the type of each activity record;
- Identify anchor records and create episodes from them; and
- Assign the remaining non-anchor records to the open episodes.

The foundation of an episode is an anchor record, which demonstrates that a clinician has evaluated the patient and decided which further services may be required to identify or treat a medical condition. Three types of activity are eligible to be anchor records:

- Activity related to the evaluation of a patient's condition;
- Activity for surgical or related procedures; and
- Activity for a treatment facility or Accident & Emergency services.

Ancillary records such as x-rays, pharmaceuticals, and lab tests are evaluated and "grouped to" the clinically most appropriate anchor record, creating clinically relevant clusters. Clusters are grouped into episodes based on a series of clinical rules. ETG prioritises related medical conditions, allowing focus on the condition that best describes the mix of services required for the on-going evaluation, management, and treatment of an episode of care. For incidental diagnoses, rather than indicate a separate incidence of a new condition, ETG combines the services into the episode for the primary disorder. This complex, hierarchical grouping of conditions provides a "filter" for characterising markers of patient risk.

The complete episodes are assigned to a Base ETG category, examples of which are shown in Table 3.

Table 3. Examples of Base Episode Treatment Groups

ETG	Description
163000	Diabetes
164800	Obesity
207200	Leukemia
238800	Mood disorder, depression
239000	Dementia
315200	Epilepsy
316400	Alzheimer's disease
386500	Ischemic heart disease
386600	Pulmonary heart disease
386900	Cardiomyopathy
388100	Hypertension
438800	Asthma
439300	Chronic obstructive pulmonary disease
475300	Inflammatory bowel disease
555400	Chronic renal failure
711902	Orthopedic trauma, fracture or dislocation
712000	Osteoporosis

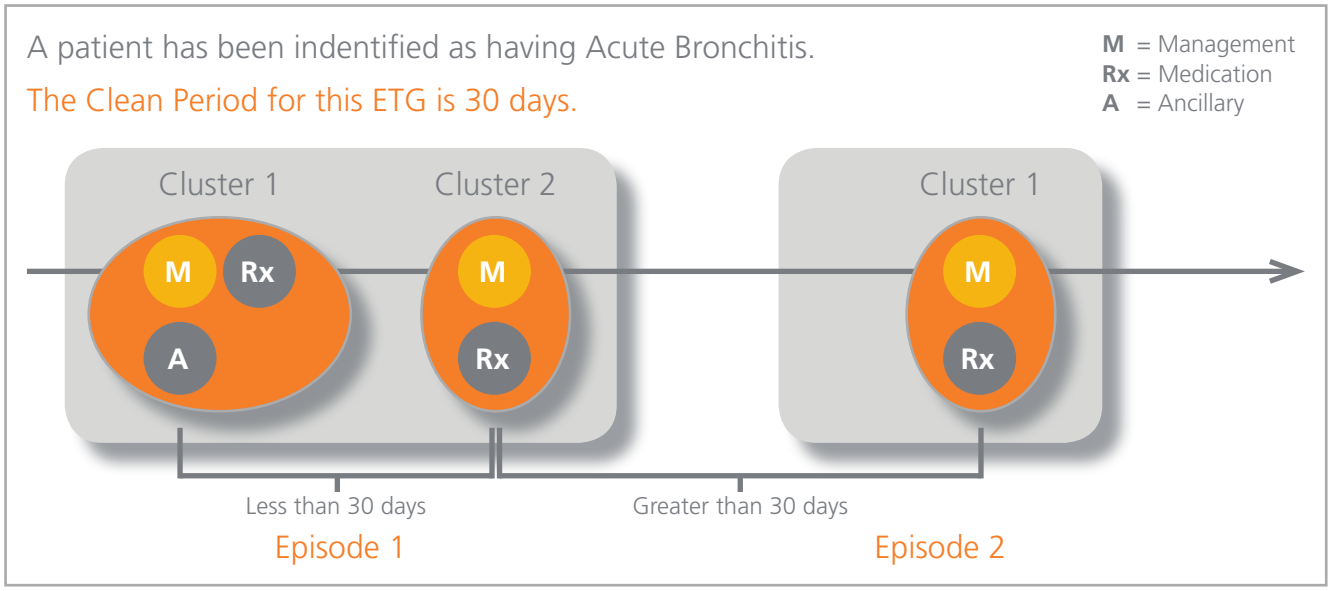
Note that Base ETGs can be created from pharmaceutical activity alone, to accommodate cases in which a GP prescribes medication for an on-going condition without requiring the patient to make a GP visit. These instances are not technically episodes of care; they are categorised into Base ETGs based on the likely indications for the drug treatment prescribed in order to retain as much information as possible from the original activity data. These pharmacy-only Base ETGs have proven valuable in assessing total risk.

Next, the ETG methodology identifies any episode-specific comorbidities and complications. Comorbidities represent on-going chronic conditions that impact treatment requirements for the episode. Complications indicate a sicker patient who may require more extensive treatment for a related condition.

A model specific to the Base ETG's unique pattern of increased severity – a model that factors in presenting comorbidities and complications – determines whether there is substantial variation in the resource costs of the episodes in that ETG to merit division of the episodes into distinct levels. If there is such variation, the episodes are split into 2, 3, or 4 severity levels. If there is minimal variation, all episodes in the Base ETG are assigned severity Level 1.

The ETG methodology continues to identify and track all clinical activity for an episode for as long as a condition is actively treated. ETG accomplishes this by the identification of discrete clean periods. A clean period is defined as the absence of treatment for a specified period of time. Each ETG has its own clean period. For example, as shown in Figure 1, the clean period for Acute Bronchitis is 30 days. Once an episode has begun for this ETG, all clinically consistent activity for acute bronchitis group to this episode until such time as 30 days passes without any corresponding clinically consistent treatment. The condition is basically life-long and all clinically consistent treatments group to an episode of benign hypertension for as long as data are available.

Figure 1 Example of Episode Building



There are 456 Base ETGs classified by ICD diagnosis codes. 146 conditions have severity levels (2, 3 or 4 levels depending upon condition) resulting in 681 condition/severity level combinations.

Step 2. Map ETGs to ERGs

The ETG output provides a record of the different episodes of care identified for an individual. A key step in developing ERGs is deciding how these episodes can best be used as markers of risk. One option is to use all of the 456 Base ETGs as separate risk markers. This approach was not chosen for several reasons. First, such a large number of risk factors would likely produce relatively small sample sizes for some markers, resulting in implausible or imprecise estimates of their contribution to risk. The level of clinical detail provided by ETGs could also produce imprecision due to the potential overlap in the impact of medically related episodes on patient risk, over- or underestimating risk for patients with different combinations of these episodes.

Episodes are therefore combined into larger groups to create ERGs. For each ETG, the severity levels described in the previous section were analysed to determine whether the pattern of increased utilisation, if any, warranted different risk levels. Both clinical input and empirical evidence guided the mapping, which involves a number of steps and assumptions:

- For selected conditions, no distinction was made between patients prescribed medication for an on-going condition without an associated GP visit and patients with the same condition who had a GP visit. For example, the migraine therapy, drug-only ETG was combined with the ETG for migraine headache to form a migraine ERG.
- ETGs with relatively low prevalence were combined with other ETGs based on clinical similarity and implications for risk assessment. The methodology for grouping of low-prevalence ETGs into ERGs should be objective, well-documented, and as consistent as possible with grouping of higher-prevalence ETGs.

- Patients are assigned an ERG risk score based on the type of ETG observed, regardless of how many episodes of the ETG are identified.
- ERG assignment is not dependent on completion status. (ETGs are considered "complete" when no related treatments are identified within a time frame that is specific to the Base ETG.)
- To enhance clinical relevance and also homogeneity in terms of risk, in each of the other steps described ETGs were only combined with other ETGs in the same Major Practice Category (MPC; see Table 4). This step permits additional focus on those episodes best describing a patient's underlying medical condition within a disease category.
- If a patient has two or more ERGs within the same MPC, a hierarchy determines which ERG is most clinically significant. Only that ERG which typically has the strongest predictive capacity is retained.
- Using this approach, a total of 189 ERGs were identified. Table 5 presents examples of the final set of risk groups.

Table 4 Major Practice Categories (MPCs)

MPC	Description	MPC	Description
1	Infectious Diseases	12	Hepatology
2	Endocrinology	13	Nephrology
3	Hematology	14	Urology
4	Psychiatry	15	Obstetrics
5	Chemical Dependency	16	Gynecology
6	Neurology	17	Dermatology
7	Ophthalmology	18	Orthopedics & Rheumatology
8	Cardiology	19	Neonatology
9	Otolaryngology	20	Preventative & Administrative
10	Pulmonology	21	Late Effects, Environmental Trauma & Poisoning
11	Gastroenterology	22	Isolated Signs & Symptoms

Table 5 Examples of Episode Risk Groups

ETG	Description	ETG	Description
MPC1	Infectious Diseases	MPC2	Endocrinology
1.001	Lower cost infectious disease	2.011	Other lower cost endorinology
1.021	Other moderate cost infectious disease	2.021	Diabetes without significant complication/comorbidity
1.031	Non HIV major infectious disease, I	2.022	Diabetes with significant complication/comorbidity, I
1.032	Non HIV major infectious disease, II	2.023	Diabetes with significant complication/comorbidity, II
1.033	Non HIV major infectious disease, III	2.031	Hyperlipidemia, excluding lipidoses

Notably, differences in treatment are not factored into the ERGs. The risk assessment should not reward or penalise treatment decisions such as the decision to admit a patient to the hospital, perform a surgery, or prescribe a medication. This is particularly important where results are used for payment purposes or assessing efficiency in providing medical care.

Step 3. Create Demographic and Custom Markers

Each of the datasets (PDS, APC, OP, A&E, GP and GPRX) were analysed to determine custom markers (independent variables) that would be useful in the prediction of an unplanned hospital admission due to a chronic condition.

Some markers are tagged based on the timing of when they occurred within the experience period. In general, markers occurring during the last 3 months of the 12-month experience period are differentiated from markers taking place during the initial 9 months of the time period. The distinction between the time frame defined as 0-3 months from the end of the experience period and the time frame defined as 4-12 months from the end of the experience period is made based on whether the

activity date on the record is on or after the date 91 days from the end of the experience period. Generally, markers occurring within the 0-3 month timeframe will indicate a higher probability of a future, unplanned, chronic admission and thus there is higher risk associated with these markers.

Demographic Markers: PDS data is used to help define patient demographic information. Every patient with a valid age and gender is assigned to a demographic marker based on their gender and age as of the last day of the reporting period used for assigning risk. In order to not over specify the model, one age group is excluded from the dataset. For this update, the Male 19-34 group was excluded due to the high volume included in it. This allows the intercept to represent this group from risk perspective.

Length of Stay (LOS) Markers: LOS is defined as the number of days a patient was present in a hospital where the stay is not for the treatment for oncology, dialysis or maternity care. The longer the length of stay, the higher the probability of an unplanned future admission. The LOS thresholds were defined as High, Medium and Low. Each of the thresholds is assigned a different risk weight. A patient triggering the variable for the "high" threshold will have a higher risk weight than a patient who triggers the marker for a "low" threshold.

IP – Unplanned Chronic Admission: An IP spell is defined as a stay administered by a non-private hospital which is not for the treatment for oncology, dialysis or maternity care. A utilisation variable was constructed to account for the number of unplanned chronic admissions a patient has.

OP: An OP attendance is defined by the patient being physically present at the facility. A utilisation variable was constructed to account for the number of attendances a patient has.

A&E: An A&E attendance is defined where a patient arrives to a 24 hour facility by ambulance for an initial encounter or an unplanned visit and the visit does not result in admittance to the facility, death or transfer to another facility. A utilisation variable was constructed to account for the number of attendances a patient has.

Specialties: Where a patient has a treatment function recorded in the OP data, the patient is assigned the specialty marker for that treatment function.

Table 6 Demographic Markers

Marker	Description	Marker	Description
f00_05	Female between 0-5	m00_05	Male between 0-5
f06_11	Female between 6-11	m06_11	Male between 6-11
f12_18	Female between 12-18	m12_18	Male between 12-18
f19_34	Female between 19-34	m35_44	Male between 35-44
f35_44	Female between 35-44	m45_54	Male between 45-54
f45_54	Female between 45-54	m55_64	Male between 55-64
f55_64	Female between 55-64	m65_74	Male between 65-74
f65_74	Female between 65-74	m75_84	Male between 75-84
f75_84	Female between 75-84	m85_UP	Male 85 or over
f85_UP	Female 85 or over		

Table 7 Inpatient Chronic Markers

Marker	Description
chronicip3_lo	If 1 Chronic Admission in last 3 month period
chronicip3_hi	If greater than 1 Chronic Admission in last 3 month period
chronicip9_lo	If 1 Chronic Admission in last 9 month period
chronicip9_med	If between 2 and 3 Chronic Admissions in last 9 month period
chronicip9_hi	If greater than 3 Chronic Admissions in last 9 month period

Table 8 Length of Stay Markers

Marker	Description
los_lo	If sum of Length of Stay less than 5 days in period
los_med	If sum of Length of Stay between 6 days and 15 days in period
los_hi	If sum of Length of Stay greater than 15 days in period

Table 9 Accident & Emergency Markers

Marker	Description
ae3_lo	If 1 A&E Attendance in last 3 month period
ae3_med	If 2 A&E Attendances in last 3 month period
ae3_hi	If 3 or more A&E Attendances in last 3 month period
ae9_lo	If 1 A&E Attendance in last 9 month period
ae9_med	If 2 A&E Attendances in last 9 month period
ae9_hi	If 3 or more A&E Attendances in last 9 month period
aeambulance	If 1 or more A&E Attendances in period Arrival Mode

Table 10 Outpatient Markers

Marker	Description
opattend3_lo	If 1 or 2 first or follow-up Outpatient Attendences in last 3 month period
opattend3_hi	If greater than 3 first or follow-up Outpatient Attendences in last 3 month period
opattend9_lo	If 1 or 2 first or follow-up Outpatient Attendences in last 9 month period
opattend9_med	If between 3 and 7 first or follow-up Outpatient Attendences in last 9 month period
opattend9_hi	If greater than 8 first or follow-up Outpatient Attendences in last 9 month period

Table 11 Outpatient Specialty Markers

Marker	Description	Marker	Description
X100	General Surgery	X310	Audiological Medicine
X101	Urology	X313	Clinical Immunology
X103	Breast Surgery	X314	Rehabilitation
X107	Vascular Surgery	X321	Paediatric Cardiology
X110	Trauma & Orthopaedics	X329	Cardiology
X120	ENT	X330	Dermatology
X130	Ophthalmology	X340	Respiratory Medicine
X140	Oral Surgery	X349	Respiratory
X150	Neurosurgery	X361	Nephrology
X170	Cardiothoracic	X370	Medical Oncology
X171	Paediatric Surgery	X400	Neurology
X180	Accident & Emergency	X410	Rheumatology
X190	Anaesthetics	X420	Paediatrics
X191	Pain Management	X421	Paediatric Neurology
X300	General Medicine	X430	Geriatric Medicine
X301	Gastroenterology	X502	Gynaecology
X302	Endocrinology	X503	Gynaecological Oncology
X303	Clinical Haematology	X719	Mental Health
X304	Clinical Physiology	X800	Clinical Oncology
X306	Hepatology	X949	Other
X307	Diabetic Medicine		

Step 4. Develop Patient's Risk Profiles

A patient's age, gender and mix of ERG and custom markers are used to create his or her risk profile. Patients without activity data will have no episodes of care and no ERGs or custom markers. For these patients, risk is based solely on age and gender.

Step 5. Create Patient Risk Scores

The next step is the assignment of a weight to each ERG, custom and demographic marker of risk. These weights describe the contribution to risk of being in a specific age-sex group or having a particular medical condition included in an ERG or having high utilisation based on the custom markers.

The markers are set to 1 if the marker is observed for an individual, 0 if not. Each patient has his or her own profile of age-sex, ERGs and utilisation. To calculate a person's risk score, the sum of these risk weights for each marker observed, including the intercept, is computed before it is exponentiated and divided by the same figure plus one i.e. if the total amount of the coefficients is X, then the risk score is: Exp(X) / (1 + Exp(X)).

Example using the Acute + GP 12 month model:

A 52 year old male has the following activity in the experience year:

- Two A&E visits within the 0-3 month timeframe
- One IP admission with a LOS of 3 days
- Four OP attendances
- Two specialties triggered: one in Endocrinology and another in Cardiology

Table 12 presents the markers triggered for the patient and the final probability calculated based on the markers. Note: The risk ratio is the likelihood (probability) divided by the overall average.

Table 12 Example Score Calculation

Variable	Description	12m
F_01_011	Lower cost infectious disease	0.1725
F_08_042	CAD, heart failure, cardiomyopathy, II	0.3932
X302	Endocrinology Specialty	0.1715
X329	Cardiology Specialty	0.2840
ae3_med	If 2 A&E Attendances in last 3 month period	0.7340
los_lo	If sum of Length of Stay less than 5 days in period	0.3645
m45_54	Male aged between 45-54	0.9491
opattend3_hi	If greater than 3 first or follow-up Outpatient Attendances in last 3 month period	0.2930

Intercept	**-5.4605**
TOTAL (-Intercept)	**-2.0987**
Exp (TOTAL)	**0.1092**
Risk Ratio	**4.1800**

Section 3. Performance

3.1 How Well do the Models Perform

The weights assigned to variables in the predictive risk models are obtained through a process called calibration. Calibration in this case is the process of weighting the linear (meaning additive) terms in the risk equation.

Logistic Regression is the statistical process that chooses the best possible set of weights for each variable, that is, it chooses the weights which produce the greatest accuracy in predicting the dependent variable. Common performance measures for predictive risk models are the C-Statistic and Max R-Squared.

Max R-Squared provides a measure of how well observed outcomes are replicated by the model, as the proportion of the total variation of outcomes explained by the model. Values range from 0 to 100%. 0% indicates that the model explains none of the variability and 100% indicates the model explains all of the variability of the response data around its mean.

C (Concordance) Statistic (C-STAT) is a standard measure of the predictive accuracy of a logistic regression model. Scores vary between 0.5 and 1.0 with higher scores indicating a better predictive model. A score of 0.5 indicates that the model is no better than chance at making a prediction. A score of 1.0 indicates a perfect model.

The statistical strength of the RISC models is shown below. The King's Fund suggest that a model with a C-STAT above 0.7 is a reasonable result and above 0.8 is a very good result. All of our models have a C-STAT above 0.8 and compare very favourably with other risk stratification models on the market.

Table 13 Regression Results

Datasets	PCT	R^2	C-stat
12m – Actute	All PCT	0.2265	0.845
3m – Actute	All PCT	0.1964	0.855
12m – Actute + GP	All PCT	0.2333	0.852
3m – Actute + GP	All PCT	0.2016	0.860

3.2 Observations on Model Performance

- Although adding GP/Meds data provides only modest improvement to model performance, its inclusion does have a large impact on the patients that actually appear at the top of the risk list i.e. those patients with the highest risk scores. This is because with the inclusion of GP/Meds data, membership in the highest risk category is driven much less by the demographic variables
- The usefulness of the reporting tool is also greatly enhanced with the addition of GP/Meds data
- Optum simulation studies show that for an LTC programme to have a high ROI it must not only have a model that performs well but it must be combined with a programme that makes effective interventions for individual patients

3.3 Usage Statistics

Predictive Ratios by Risk Marker

Predictive ratio analysis was used to ascertain the likelihood of unplanned admissions by key variables. Counts of how many patients triggered a particular marker and how many of those patients had an unplanned chronic admission in the target year were compared to counts of patients who did not trigger the marker and how many of those had an unplanned chronic admission. Results can be viewed in Table 14 and clearly indicate that patients with particular markers are more likely to have an unplanned chronic admission.

Table 14 Unplanned Admissions Predictive Analysis

	Observed			Not Observed			
Risk Marker Description	Marker Flagged	Admissions	Admission %	Marker Not Flagged	Admissions	Admission %	Predictive Ratio
High cost neurology, I	2,249	514	22.85	1,713,000	37,994	2.22	10.30
High cost neurology, II	766	232	30.29	1,714,483	38,276	2.23	13.57
Epilepsy, I	2,244	356	15.86	1,713,005	38,152	2.23	7.12
Epilepsy, II	1,183	333	28.15	1,714,066	38,175	2.23	12.64
Ischemic heart disease, Heart failure, Cardiomyopathy, I	10,729	2,155	20.09	1,704,520	36,353	2.13	9.42
Ischemic heart disease, Heart failure, Cardiomyopathy, II	1,764	468	26.53	1,713,485	38,040	2.22	11.95
Ischemic heart disease, Heart failure, Cardiomyopathy, III	1,617	428	26.47	1,713,632	38,080	2.22	11.91
Ischemic heart disease, Heart failure, Cardiomyopathy, IV	2,102	537	25.55	1,713,147	37,971	2.22	11.53
Ischemic heart disease, Heart failure, Cardiomyopathy, V	541	184	34.01	1,714,708	38,324	2.24	15.22
Ischemic heart disease, Heart failure, Cardiomyopathy, VI	189	63	33.33	1,715,060	38,445	2.24	14.87
Hypertension, without complication/comorbidity	41,865	3,289	7.86	1,673,384	35,219	2.10	3.73
Hypertension, with complication/comorbidity	608	200	32.89	1,714,641	38,308	2.23	14.72
Hypertension, with significant complication/comorbidity	230	80	34.78	1,715,019	38,428	2.24	15.52
Asthma, COPD, I	62,654	2,574	4.11	1,652,595	35,934	2.17	1.89
Asthma, COPD, II	19,785	2,479	12.53	1,695,464	36,029	2.13	5.90
Asthma, COPD, III	2,101	583	27.75	1,713,148	37,925	2.21	12.53
Asthma, COPD, IV	181	76	41.99	1,715,068	38,432	2.24	18.74
Chronic renal failure, I	834	224	26.86	1,714,415	38,284	2.23	12.03
Chronic renal failure, II	274	114	41.61	1,714,975	38,394	2.24	18.58
Chronic renal failure, III	337	138	40.95	1,714,912	38,370	2.24	18.30

Predictive Performance

In order to gauge the predictive performance of our models we looked at the entire population and then sorted in ascending order on the likelihood of having an unplanned admission. Table 15 shows percentage of patients having unplanned admissions by various cut-off percentages in terms of risk.

Table 15 Predictive Values

Acute 12 month			
Top %	Count	Unplanned Admissions	%
0.5%	8576	3786	44.15%
1.0%	17152	6150	35.86%
5.0%	34305	7766	22.64%
10.0%	51457	8713	16.93%
25.0%	137220	10434	7.60%
50.0%	222982	11319	5.08%

Acute 3 month			
Top %	Count	Unplanned Admissions	%
0.5%	8576	1455	16.97%
1.0%	17152	2194	12.79%
5.0%	34305	2577	7.51%
10.0%	51457	2793	5.43%
25.0%	137220	3228	2.35%
50.0%	222982	3391	1.52%

Acute + GP 12 month			
Top %	Count	Unplanned Admissions	%
0.5%	8576	3864	45.06%
1.0%	17152	6218	36.25%
5.0%	34305	7986	23.28%
10.0%	51457	8989	17.47%
25.0%	137220	10812	7.88%
50.0%	222982	11475	5.15%

Acute + GP 3 month			
Top %	Count	Unplanned Admissions	%
0.5%	8576	1485	17.32%
1.0%	17152	2196	12.80%
5.0%	34305	2612	7.61%
10.0%	51457	2847	5.53%
25.0%	137220	3265	2.38%
50.0%	222982	3395	1.52%

3.4 Sampling and Over Fitting

After finishing the model we carried out repeated split-sampling to verify the results. The process involves the following steps:

- Randomly divide the 'Acute + GP' 12 month patient profile table into two equal groups (group A and B);
- Run the regression using all final markers against both sets of patients;
- Review the output from this regression to ensure that there were not significant differences between groups A & B;
- Create spreadsheet containing:
  - Markers
  - Counts of how many patients triggered that marker and how many of those patients had an unplanned chronic admission in target year
  - Counts of how many patients did not trigger a given marker and how many of those patients had an unplanned chronic admission in target year

The results of this exercise yielded no significant difference between the output from Group A and Group B.

Section 4. HealthNumerics-RISC® Application

HealthNumerics-RISC® is the application that operationalises Optum's predictive risk models for the NHS.

Some of the key features of HealthNumerics-RISC® are detailed below:

- Provides a 'prioritised list' of your entire population by their level of risk – viewed as either a list or card
- Shows all key information about an individual in one place – e.g. primary and secondary care activity, prescribing information, diagnoses, etc. and includes a care delivery timeline for all activity
- Identifies patients rising up the risk ranking – "high climbers"
- Shows Mental Health, Social Care and long term condition risk factors information against each individual (if agreed/provided locally)
- Allows you to customise the list at your desktop by single or combination factors such as condition, age, risk level, % risk change and geography
- Highlights candidates being case managed and by whom, and allows you to comment on cases and review previous notes
- Provides a customisable home page – showing individuals on a 'watch' list and short cuts to your own saved reports (refreshed each month)

The risk scores in HealthNumerics-RISC® are presented as a ratio of the average risk score in the entire population. This means that:

- A typical young patient with no utilisation will have a risk ratio between 0.23 and 0.30 (specific to the organisation's population)
- A risk ratio of 1.0 indicates a person who has risk equal to the average of the entire population
- A risk ratio score of 5.0 indicates a 5 times greater risk of unplanned hospitalisation for chronic disease than the typical patient
- A risk ratio of 30 indicates a risk that is 30 times greater for a future hospitalsation than that of the typical patient
- The 'riskiest' patients typically have RISC ratios of 25-40

HealthNumerics-RISC® is:

- **User Friendly.** Designed for clinicians, commissioners and case managers, and requires little or no training;
- **Easy to Implement.** Supports CCGs and CSUs of various sizes on a single Optum-hosted platform. Client-hosted solutions can also be discussed.
- **Secure.** Hosted at a secure N3 data centre, so client data never leaves the NHS. Optum is fully compliant with the Information Governance Toolkit and has also gained ISO27001 accreditation. Includes roles based access/restriction to patient data.
- **Fully Supported.** Helpdesk serving over 8,000 GP practices. Development partnerships.

Section 5. Policy Implication

In the last year, the English NHS has seen a shift to giving clinicians powers to commission care for their localities. Alongside this we are seeing changes in the patterns of disease moving away from infection and towards non-communicable diseases and long term conditions, as the principal causes of morbidity and mortality within populations. Health spending is also being squeezed.

However, improving the quality of care the patient receives has also become an imperative both clinically and politically; and especially understanding how to intervene and by what method to try and prevent the individual from becoming a patient or by mitigating further complications. Targeting patients at risk of complications or co-morbidities helps the individual to lead a higher quality of life, whilst at the same time contributing to improved productivity within the health system. Similarly, with sections of the population at higher risk of unplanned admission to hospital, ensuring early enough interventions to prevent these occurrences improves overall quality of care.

Examples of how the model works and the return on investment:

Case Study Example: North East Lincolnshire Care Trust Plus

Optum worked with NEL CTP to create, implement and oversee an integrated health and social care case management programme – focusing particularly on individuals with complex needs and their carers. By working with the Health and Social Care Leads and Case Managers to decide what qualifies as a 'complex' need – as well as to embed the technology and tools that underpinned the programme. Since July 2010, the combination of the complex case management and rapid response programmes have improved quality and reduced service use – leading to £1.7 million in savings in the first recorded nine months, as well as 1,291 avoided visits to A&E, 504 avoided emergency hospital admissions and 346 avoided permanent care home placements. Other quality improvements and financial savings were achieved through: Reduced hospital length of stay; Avoided GP home visits; Avoided ambulance calls; and Avoided provision of, or increase to, care packages.

Key elements of the model included:

- **Risk Stratification.** We used HealthNumerics-RISC® – UnitedHealth UK's risk stratification tool – to identify and target individuals with a high risk of admission to hospital. We also developed an adult social care risk assessment tool to identify individuals with a high risk of being admitted to a care home.
- **In-reach.** Working with NEL CTP, we designed an 'in-reach' discharge process to help develop care plans for individuals with complex needs who we had identified in secondary care, intermediate care and short stay beds. Case managers now receive an alert when these individuals are admitted on to the medical wards.
- **User Empowerment to Develop an Integrated Care Plan.** We developed booklets for individuals to record their personal goals and develop contingency plans for when their care situation deteriorates rapidly – enabling them to design their own care plan.
- **Outcomes Reporting.** We designed and developed a reporting tool that enables community teams (responsible for complex case management) and the rapid responders to record and view outcomes for the individuals in their care.

Case Study Example: Erewash CCG

Optum is working with multiple CCG's across Derbyshire and in Leicestershire to develop Integrated Care models, engage stakeholders across the health & social care system, lead behavioural change, identify gaps in services, develop and recruit to new care coordinator roles, and implement a new integrated care model and pathway. We are working with each of the CCG's and their Adult Care Boards & Teams, GP's, Mental Health Providers, Community Providers, Third Sector, Secondary Care and Patient Participation groups to ensure a robust service is implemented at pace and with full engagement. In eight months, we have set up the Single Point of Access Service and Community Delivery Teams for all Erewash GP practices to utilise, and in Hardwick, we have rolled out their virtual ward model to 16 practices.

In both Erewash and Hardwick CCGs, HealthNumerics-RISC® was chosen as their risk stratification tool. Care coordinators were given responsibility to utilise the tool to identify individuals at highest risk of hospital admission, those with the most complex needs and the high climbers. Appropriate patients were then selected to take to their multi disciplinary meeting for discussion. In addition, patients managed by the community matrons were identified and noted in the HNR tool so people being case managed were easily recognised. The matrons worked closely with the care coordinators to manage this cohort of individuals.

> "Optum have provided pace and focus for our CCG, social care, community health provider and voluntary providers in Erewash to support individuals to receive appropriate care at home or as close to home as possible, and to prevent inappropriate hospital attendances and admissions through development of a model, clinical navigation and integrated community teams. Their clinical and programme management insight and rigour has been a key success factor for our integrated care programme."
>
> Rakesh Marwaha, Accountable Officer, Erewash CCG

Section 6. Summary and Conclusions

Risk stratification is increasingly important for the day-to-day operations and strategic decision making of health care organisations. It plays a key role in allocating resources effectively and in targeting opportunities for clinical and financial improvement.

Risk stratification has a number of practical applications for health care analysis and health services research. Accurate risk scoring and stratification is essential in creating valid comparisons of the efficiency and quality of the services provided to patients. By identifying higher risk patients, it contributes to assertive, effective care management.

The HealthNumerics-RISC® predictive models use powerful, cost effective, and practical risk stratification methodology. The clinical relevance of the models provides an effective tool for understanding patient profiles of medical conditions and how they impact current and future health risk. The economies gained by organisations when using consistent methods for performing both risk stratification (HealthNumerics-RISC®) and episodes of care grouping (ETG) also offer significant advantages. The ETG methodology and the architecture for mapping ETGs to ERGs employ systematic, logical, and transparent clinical approaches.

About Optum

Optum, a division of UnitedHealth Group, is a leading information and technology-enabled health services business dedicated to helping improve the health system for everyone.

Optum is leading information and technology-enabled health services business dedicated to helping improve the health system for everyone.

Our work touches virtually every segment of the health care market which means we're able to bring a big picture perspective to specific client challenges – around the world.

We draw on our diverse expertise, deep knowledge and unmatched experience in helping clients across the health spectrum become more connected, intelligent and aligned, and to take steps to increase efficiency, lower costs and raise standards of health and well-being.

We apply our capabilities and experience in data, technology, analytics, care management and well-being support programmes for employers, governments, health care professionals and individuals in all corners of the globe.

OPTUM™

T 020 7121 0560 I E ukcontact@optum.com I **www.optum-uk.com**
3rd Floor, 3 Sheldon Square, Paddington, London W2 6HY