Open-source Data Analysis and Machine Learning for Asthma Hospitalisation Rates

Laura Rooney, Chaloner Chute
Digital Health and Care Institute,
University of Strathclyde,
Glasgow, UK.
laura.rooney@dhi-scotland.com
chalonerchute@dhi-scotland.com

William J Buchanan, Adrian Smales
The Cyber Academy,
Edinburgh Napier University,
Edinburgh, UK.
w.buchanan@napier.ac.uk
a.smales@napier.ac.uk

Leigh-Anne Hepburn
Digital Health and Care Institute,
Glasgow School of Art,
Glasgow, UK.
leigh-anne.hepburn@dhi-scotland.com

***Abstract*—Long-term conditions in Scotland account for 80% of all GP consultations; they also account for 60% of all deaths in Scotland. Asthma and Chronic Obstructive Pulmonary Disease (COPD) are common long-term respiratory diseases [1]. Asthma is a heterogeneous disease, usually characterized by chronic airway inflammation. It is defined by the history of respiratory symptoms such as wheeze, shortness of breath, chest tightness and cough that vary over time and in intensity, together with variable expiratory airflow limitation [2]. So far, we know that there are many different things – such as viruses, allergens, and pollution – that cause asthma or trigger attacks but not why or how they do it. This paper outlines how an open source dataset can be used to estimate asthma hospitalisation rates and uses machine learning to predict these rates, within $\pm$7.5%, and for an 86.67% success rate.**

I. Introduction

The prevalence of asthma continues to increase worldwide (it affects 1-18% of the global population), although non-communicable diseases such as Asthma are still not seen as a health care priority in many countries despite their impact. Asthma is a chronic inflammatory disease that presents as a long-term condition with severity varying from person to person. Globally, 300 million people are affected by asthma on a daily basis [3]. In 2015, the UK had the highest number of Asthma sufferers in Europe with 4.67 million people [4].

In the UK now, 1 in 11 people have asthma, including 1.1 million children and 4.3 million adults. Shockingly, every 10 seconds, someone in the UK has a potentially life-threatening asthma attack and each day three people in the world die from asthma [5]. The landmark International Study of Asthma and Allergies in Childhood (ISAAC) and the European Community Respiratory Health Survey (ECRHS) studies demonstrated substantial national variations in asthma prevalence, with evidence suggesting that the UK ranks as being one of the countries with the highest prevalence in the world [6]. Bringing these statistics down to a more regional level, Scotland has one of the highest prevalence of asthma in Europe with 9.51% of the population being affected by the condition (554,306 people) [3]. It is also predicted that by 2025, there will be almost 400 million asthma patients worldwide.

Scotland provides a good test case in analysing health and social care factors, especially where its local authorities and health boards often face different challenges in regard to health and well-being. While national statistics, such as SIMD (Scottish Index of Multiple Deprivation) exist [7], they do not give enough details on the factors which could influence a more in-depth analysis of contributory factors. This paper thus uses the publicly available data set from ScotPHO Profiles [8], and uses this to train towards the best machine learning models in order to predict the outcomes.

The methodology used in this paper to analyse asthma hospitalisation rates are to take an open source data set with a range of metrics and then:

- The metrics are correlated, using linear regression (with the Pearson correlation coefficient) against each other to see significant linkages.
- A machine is trained using 30% of the data and then to predicts from hospitalisation for asthma. In this way, the machine is being trained against three metrics within the data in order to determine the most significant factors in the matching.
- Each of the models are then assessed for the complete dataset, along with a success threshold. In this way, a success band within +/-7.5% between the minimum and maximum value is used. The top contenders for a successful match are then outlined and checked for their usefulness in showing contributory factors.
- Once all the models have been determined, the top machine learning models for the top 100 successful models are then selected and the variables from each on the most successful training models are then scored for their success. Those metrics which appear most often within the most successful models can then be defined as the most useful in predicting asthma hospitalisation rates.

II. Factors Influencing asthma prevalence

A. Environmental factors

Despite the considerable genetic contribution, for example, the presence of asthma and atopy in the family, most importantly the maternal history of asthma, there are likely to be

 ISBN: 978-1-61208-682-8

several additional social and environmental factors involved in the exacerbation of asthma [9]. With most of the world's population living in urban areas, the environmental conditions whereby pollution continues to increase, have and will continue to have an influence on the rise in asthma prevalence. Air pollution – whether it's traffic fumes, smoke or dust particles – is an asthma trigger which is difficult to avoid in city environments. According to a survey carried out by Asthma UK, two-thirds of people reported that poor air quality makes their asthma worse.

B. Passive smoke

A study carried out to investigate the relationship between air quality and school absences found that exposure to cigarette smoke had a significant impact on current wheeze, use of services, and interference with physical activity. Among adolescents, smoke exposure was found to be a more important factor than deprivation in relation to symptoms, use of services, and impact on activities [10]. Passive smoking is a major global problem, causing 12,000 deaths each year in the UK and causing 2% of the current annual total deaths. These figures alone are enough to provide sufficient evidence to promote the prevention of passive smoking in the public domain [11].

C. Childhood obesity

It is also well documented that asthma presents its peak incidence in childhood. In the US, the age group with the highest percentage of persons with asthma from 2001 to 2015 was consistently age 5-14 years [4]. Obesity rates in children have increased dramatically across most English-speaking countries over the past decades and the impact on obesity on health has become a major global burden.

D. Poverty and asthma

A retrospective study carried out in West Midlands (England) to investigate the relationship between asthma admission rates, routes of admission and socioeconomic deprivation found that asthma admissions are strongly associated with deprivation in the community [12]. Asthma admission rates were higher in all age groups (except over 65s) for those from poorer districts. The age group with the most significant relationship between asthma admission rates and Townsend deprivation index was 0-4 years. Another study stipulates that higher hospital admission rates of asthmatics who are poor or belong to ethnic minorities may be due to the fact that these groups rely mainly on crisis management of the condition, [13] are under medicated [14], are under-users of primary care facilities, lack a planned crisis management, live in adverse environmental conditions in terms of asthma triggers such as smoking and cockroach exposure, or are exposed more frequently than other groups to psychosocial problems within the family and their community [15].

E. Asthma in the elderly

Although asthma has a high burden in children, the relative importance of asthma impact increases with old age and is also particularly apparent in the elderly, especially in women. Asthma reportedly affects 10% of the over 65 population who in general have lower lung function and greater symptom severity than young asthmatic patients. "late onset" asthma is first diagnosed after the age of 65 and is often a more severe phenotype, with less symptom-free days, and a higher requirement for oral corticosteroids [16].

III. Methodology

A. Open source data analysis

Population studies have been used within smoking and alcohol studies for many decades, such as in 1939 when Hermann Müller at Cologne Hospital in 1939 published his work on the linkage between smoking and cancer [17], and which was confirmed by Eberhard Schairer and Eric Schöniger [18]. Their work was then confirmed in the 1950s by a number of epidemiological studies, including Ernst Wynder and Evarts Graham [19]. Doll and Hill confirmed the effect by observing that smokers of 35 or more cigarettes per day had a 40.6 times increase in the odds of dying from lung cancer [20]. Since then smoking has been increasingly pinpointed within a number of ailments, including for COPD [21].

Smoking prevalence has also been studied and matched to differing demographics. This includes with [22] where researchers analysed 4,411 respondents aged 15 to 54 years and found that the smoking rates for no mental illness was 22.5%, while those with a lifetime mental illness was 34.8%. Along with this smoking is seen to be an increasing problem with poverty [23], especially as those who are wealthy are more likely to quit smoking than those in poverty [24].

B. Machine learning

In terms of asthma, machines learning is now playing an increasing role in creating expert systems for diagnosis. [25] used questionnaire data and clinical data to train a machine learning methods and used Context sensitive auto-associative memory neural network model (AMNN), a Backpropagation (BP) model, the C4.5 algorithm, a Bayesian Network (BN), and Particle Swarm Optimization (PSO). They found the accuracy of the methods varied between 81.17% and 84.16%, and that the accuracy of the AMNN and PSO methods were the most accurate, along with having excellent learning and diagnostic abilities.

[26] reviewed over five years of data and applied latent class analysis to distinguish asthma and wheezing subtypes in childhood. The data sets analysed included ALSPAC [27]; AMICS [28]; AMICS-Menorca [29]; CAPS [30]; CCCEH [31]; DARC [32]; ECRHSII [33]; EGEA2 [34]; GINIplus [35]; IoW; ISAAC phase II; Cohort; LISA; MAAS; MACS; MAS; MCS; PARIS; PASTURE; PIAMA; PIPO; SLAM; and WHEALS. Of the 34 data sets analysed, 16 included wheezing and coughing, 10 included atopic status, eight included growth patterns, and two included eczema. In the data sets, 22 related to children and only two related to adults, with the rest defined as unspecified. The cohort size was typically between 20 and 66.

 ISBN: 978-1-61208-682-8

In [36], the authors used a machine learning approach to analyse the pattern of IgE response (over time or to specific allergens) in order to identify atopic vulnerabilities related to the presence of asthma. They used data from the skin and IgE tests from childhood onwards and clustered the population into multiple atopic classes using unsupervised learning. These were trained against asthma-related data, such as for symptoms, hospitalizations, lung function and airway reactivity. Their results indicate four main classes: Multiple Early (10.6%); Multiple Late (16.2%); Dust Mite (4.5%); and Non-dust Mite (9.5%), along with a fifth class of No Latent Vulnerability (59.2%). The association for asthma was most strong for Multiple Early class, and which also showed a considerable link to lung function and airway reactivity. Along with this the work identified a highly significant increase in the risk of hospital admissions for wheeze/asthma after three years old, but only among children in the Multiple Early class.

IV. Results

The results use a publicly available data set from ScotPHO Profiles [8], and where 56 metrics are used to train against (as outlined in Table III) and using the local authority region as the index value. If we select three variables to train against, one variable we will have 26,235 triplets to test, while four variables will give us 341,055 machine learning assessments (Table II). A benchmark of the time to check a model and to match against predicted values gives an estimated time of 0.4 seconds. Table II thus outlines estimations for orders of run times. As we see the total run time for four variables is fairly large and costly in computation time, while two variables are not likely to give us enough variation in the variables in the models, thus this paper uses three variables to train against.

While we can apply linear regression to the data, there are often complex interrelationships that need to be analysed with machine learning. In the evaluation, the data was analysed using the Python RandomForestRegressor method, using a success rate of $\pm 7.5\%$ for asthma hospitalisation rates. In each case, 30% of the data is taken to train the machine model, and then all of the data is used to test for success.

Table I outlines patients hospitalised with asthma for local authority areas within Scotland. In terms of correlation with hospitalised due to asthma, Table IV provides the strong correlation factors which were greater than a magnitude of 0.5 from within the dataset. This shows that COPD, smoking, and metrics related to those hospitalised over 65 were the strongest in correlation. There are thus generally strong positive correlations with COPD, but a negative correlation between male and female life expectancy.

In Table V, we see the results of running random forest linear regression against patients hospitalised with asthma against three other metrics. The determination of success in predicting the hospitalisation rate is defined as $\pm 7.5\%$. The best success level is 86.67%, and which had eight best solutions (Table VI shows one of the best models). One of these is New cancer registrations, People aged 65+ with high care needs cared at home, and Children in Poverty. If we now analyse the Top 100 models from machine learning, we generate Table III. In this way, we see that Patients with emergency hospitalisation appear within 71.7% of the top models, and is a strong predictor for estimating hospitialisation rates for asthma.

TABLE I
Patients hospitalised with asthma (per 100,000 population. Yearly averages for three year period from 2014/2015 to 2016/2017.)

Area	Patients hospitalised with asthma
Aberdeen City	73.6
Aberdeenshire	57.7
Angus	67.7
Argyll and Bute	71.8
Clackmannanshire	77.8
Dumfries and Galloway	80.3
Dundee City	82.9
East Ayrshire	109
East Dunbartonshire	78.4
East Renfrewshire	80.7
Edinburgh	86.8
Falkirk	91
Fife	77.4
Glasgow City	119.5
Highland	92.2
Inverclyde	103.5
Lanarkshire	109.6
Mid and East Lothian	110.3
Moray	60.2
North Ayrshire	133.1
Orkney Islands	56.6
Outer Hebrides	68.6
Perth and Kinross	63.3
Renfrewshire	104.3
Scottish Borders	86.1
Shetland Islands	39
South Ayrshire	100.9
Stirling	66
West Dunbartonshire	115.8
West Lothian	87.6

V. Conclusions

Increasingly open source data can be used to make predictions based on populations, and this paper shows that key metrics in predicting asthma rates, such as COPD, smoking rates and emergency hospitalisation rates. Within machine learning, a key factor is defining the required training features within a model. This paper has outlined some of the key features which could be used to predict asthma hospitalisation rates from open source data, and the ones which can be ignored.

VI. Data set

The data set is available at https://asecuritysite.com/log/well.csv and machine learning models at https://asecuritysite.com/bigdata/ml?file=well.csv.

References

[1] L. Rutherford, S. Hinchliffe, and C. Sharp, "The Scottish Health Survey 2012," *The Scottish Health Survey*, vol. 1, 2013.

 ISBN: 978-1-61208-682-8

[2] E. D. Bateman, S. S. Hurd, P. J. Barnes, J. Bousquet, J. M. Drazen, M. FitzGeralde, P. Gibson, K. Ohta, P. O'Byrne, S. E. Pedersen, E. Pizzichini, S. D. Sullivanee, S. E. Wenzel, and H. J. Zar, "Global strategy for asthma management and prevention: GINA executive summary," *European Respiratory Journal*, vol. 31, no. 1, pp. 143–178, 2008.

[3] A. UK, "Connected asthma : how technology will transform care," p. 25, 2016.

[4] Statista, "Number of individuals with asthma in Europe in 2015 , by county (in million)," Tech. Rep., 2015.

[5] Asthma UK, "Asthma UK Strategy 2014 – 17 Reduce risk of asthma attacks Our long-term mission is to :," Tech. Rep., 2014.

[6] A. Sheikh, M. F. Steiner, G. Cezard, N. Bansal, C. Fischbacher, C. R. Simpson, A. Douglas, and R. Bhopal, "Ethnic variations in asthma hospital admission, readmission and death: A retrospective, national cohort study of 4.62 million people in Scotland," *BMC Medicine*, vol. 14, no. 1, pp. 1–9, 2016.

[7] Scottish Government Health Department, "Scottish Index of Multiple Deprivation," pp. –, 2017. [Online]. Available: http://www.scotland.gov.uk/Topics/Statistics/SIMD

[8] "Home - ScotPho." [Online]. Available: http://www.scotpho.org.uk/

[9] M. Martel, E. Rey, J. L. Malo, S. Perreault, M. F. Beauchesne, A. Forget, and L. Blais, "Determinants of the incidence of childhood asthma: A two-stage case-control study," *American Journal of Epidemiology*, vol. 169, no. 2, pp. 195–205, 2009.

[10] J. B. Austin, S. Selvaraj, D. Godden, and G. Russell, "Deprivation, smoking, and quality of life in asthma," *Archives of Disease in Childhood*, vol. 90, no. 3, pp. 253–257, 2005.

[11] L. Forbes, "Asthma and atopy: Endocrine or metabolic conditions?" *Thorax*, vol. 60, no. 10, pp. 793–794, 2005.

[12] J. P. Watson, P. Cowen, and R. A. Lewis, "The relationship between asthma admission rates, routes of admission, and socioeconomic deprivation," *European Respiratory Journal*, vol. 9, no. 10, pp. 2087–2093, 1996.

[13] N. Halfon and P. W. Newacheck, "Childhood asthma and poverty: differential impacts and utilization of health services," *Pediatrics*, vol. 91, no. 1, pp. 56–61, 1993.

[14] C. L. Joseph, S. L. Havstad, D. R. Ownby, C. C. Johnson, and B. C. Tilley, "Racial differences in emergency department use persist despite allergist visits and prescriptions filled for antiinflammatory medications," *Journal of Allergy and Clinical Immunology*, vol. 101, no. 4 I, pp. 484–490, 1998.

[15] R. J. Rona, "Asthma and poverty," *Thorax*, vol. 55, no. 3, pp. 239–244, 2000.

[16] S. S. Braman, "Asthma in the Elderly," *Clinics in Geriatric Medicine*, vol. 33, no. 4, pp. 523–537, 2017.

[17] F. H. Muller, "Tabakmißbrauch und Lungencarcinom," *Zeitschrift für Krebsforschung*, vol. 49, no. 1, pp. 57–85, 1940.

[18] E. Schairer and E. Schöniger, "Lungenkrebs und Tabakverbrauch," *Zeitschrift für Krebsforschung*, vol. 54, no. 4, pp. 261–269, 1944.

[19] G. E. A. Wydner EL, "Tobacco smoking as a possible etiologic factor in bronchogenic carcinoma. a study of six hundred and eighty four proved cases," *J Am Med Assoc.*, vol. 143, no. 143, pp. 329–336, 1950.

[20] R. Doll and A. B. Hill, "Lung Cancer And Other Causes Of Death In Relation To Smoking A Second Report On The Mortality Of British Doctors," *British Medical Journal*, vol. 2, no. 5001, p. 1071, 1956.

[21] S. Gompertz, D. L. Bayley, S. L. Hill, and R. A. Stockley, "Relationship between airway inflammation and the frequency of exacerbations in patients with smoking related COPD," *Thorax*, vol. 56, no. 1, pp. 36–41, 2001.

[22] K. Lasser, J. W. Boyd, S. Woolhandler, D. U. Himmelstein, D. McCormick, and D. H. Bor, "Smoking and mental illness," *Australasian psychiatry : bulletin of Royal Australian and New Zealand College of Psychiatrists*, vol. 23, no. 1, p. 2606, 2015.

[23] Y. Liu, K. Rao, T. w. Hu, Q. Sun, and Z. Mao, "Cigarette smoking and poverty in China," *Social Science and Medicine*, vol. 63, no. 11, pp. 2784–2790, 2006.

[24] K. Humphreys, "Why the wealthy stopped smoking, but the poor didn't," 2015.

[25] B. Prasadl, P. E. S. N. Krishna Prasad, and y. Sagar, "An Approach to Develop Expert Systems in Medical Diagnosis Using Machine Learning Algorithms (Asthma) and A Performance Study," *International Journal on Soft Computing*, vol. 2, no. 1, pp. 26–33, 2011. [Online]. Available: http://www.airccse.org/journal/ijsc/papers/2111ijsc03.pdf

[26] R. Howard, M. Rattray, M. Prosperi, and A. Custovic, "Distinguishing Asthma Phenotypes Using Machine Learning Approaches," *Current Allergy and Asthma Reports*, vol. 15, no. 7, p. 38, 7 2015.

[27] J. Henderson, R. Granell, J. Heron, A. Sheriff, A. Simpson, A. Woodcock, D. P. Strachan, S. O. Shaheen, and J. A. Sterne, "Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood," *Thorax*, vol. 63, no. 11, pp. 974–980, 2008. [Online]. Available: http://thorax.bmj.com/

[28] J. Sunyer, J. M. Antó, J. Harris, M. Torrent, O. Vall, P. Cullinan, and A. Newman-Taylor, "Maternal atopy and parity," *Clinical and Experimental Allergy*, vol. 31, no. 9, pp. 1352–1355, 2001.

[29] P. Rzehak, A. H. Wijga, T. Keil, E. Eller, C. Bindslev-Jensen, H. A. Smit, J. Weyler, S. Dom, J. Sunyer, M. Mendez, M. Torrent, O. Vall, C. P. Bauer, D. Berdel, B. Schaaf, C. M. Chen, A. Bergström, M. P. Fantini, M. Mommers, U. Wahn, S. Lau, and J. Heinrich, "Body mass index trajectory classes and incident asthma in childhood: Results from 8 European Birth Cohorts - A Global Allergy and Asthma European Network initiative," *Journal of Allergy and Clinical Immunology*, vol. 131, no. 6, pp. 1528–1536, 6 2013.

[30] G. B. Marks, S. Mihrshahi, A. S. Kemp, E. R. Tovey, K. Webb, C. Almqvist, R. D. Ampon, D. Crisafulli, E. G. Belousova, C. M. Mellis, J. K. Peat, and S. R. Leeder, "Prevention of asthma during the first 5 years of life: A randomized controlled trial," *Journal of Allergy and Clinical Immunology*, vol. 118, no. 1, pp. 53–61, 7 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16815138

[31] F. P. Perera, V. Rauh, W. Y. Tsai, P. Kinney, D. Camann, D. Barr, T. Bernert, R. Garfinkel, Y. H. Tu, D. Diaz, J. Dietrich, and R. M. Whyatt, "Effects of transplacental exposure to environmental pollutants on birth outcomes in a multiethnic population," *Environmental Health Perspectives*, vol. 111, no. 2, pp. 201–205, 2003. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1241351/pdf/ehp0111-000201.pdf

[32] A. Host, S. Halken, H. P. Jacobsen, A. E. Christensen, A. M. Herskind, and K. Plesner, "Clinical course of cow's milk protein allergy/intolerance and atopic diseases in childhood," *Pediatric Allergy and Immunology*, vol. 13, no. s15, pp. 23–28, 12 2002. [Online]. Available: http://doi.wiley.com/10.1034/j.1399-3038.13.s.15.7.x

[33] S. H. Downs, "Continued increase in the prevalence of asthma and atopy," *Archives of Disease in Childhood*, vol. 84, no. 1, pp. 20–23, 2001.

[34] F. Kauffmann and M.-H. Dizier, "EGEA (Epidemiological study on the Genetics and Environment of Asthma, bronchial hyperresponsiveness and atopy) - design issues," *Clinical and Experimental Allergy*, vol. 25, no. s2, pp. 19–22, 11 1995.

[35] P. Rzehak, S. Sausenthaler, S. Koletzko, C. P. Bauer, B. Schaaf, A. Von Berg, D. Berdel, M. Borte, O. Herbarth, U. Krämer, N. Fenske, H. E. Wichmann, and J. Heinrich, "Period-specific growth, overweight and modification by breastfeeding in the GINI and LISA birth cohorts up to age 6 years," pp. 449–467, 2009. [Online]. Available: http://www.jstor.org/stable/40284152

[36] A. Simpson, V. Y. Tan, J. Winn, M. Svensén, C. M. Bishop, D. E. Heckerman, I. Buchan, and A. Custovic, "Beyond atopy: Multiple patterns of sensitization in relation to asthma in a birth cohort study," *American Journal of Respiratory and Critical Care Medicine*, vol. 181, no. 11, pp. 1200–1206, 6 2010. [Online]. Available: http://www.atsjournals.org/doi/abs/10.1164/rccm.200907-1101OC

 ISBN: 978-1-61208-682-8

TABLE II
ESTIMATIONS OF RUN TIME FOR A RANGE OF VARIABLES TO TRAIN AGAINST

Variables to match	Combinations	Seconds per variable	Minutes per variable	Hours per variable	Total time (hours)
2	1485	594	9.9	0.165	9.24
3	26,235	10,494	174.9	2.915	163.24
4	341,055	136422	2,273.7	37.895	2,122.12
5	3478,761	1,391,504.4	23,191.74	386.529	21,645.624
6	28,989,675	11,595,870	193,264.5	3,221.075	180,380.2
7	202,927,725	81,171,090	1,352,851.5	22,547.525	1262661.4

TABLE III
ESTIMATIONS OF RUN TIME FOR A RANGE OF VARIABLES TO TRAIN AGAINST

Metric	Occurrence in machine learning model (%)
Patients with emergency hospitalisations	71.7
Patients (65+) with multiple emergency hospitalisations	26
Population prescribed drugs for anxiety/depression/psychosis	20.9
Breast screening uptake	14.9
Children Living in Poverty	13.7
Working age population claiming Out of Work benefits	10.2
Domestic Abuse	9
Working age population employment deprived	8.1
Crime rate	6.8
Adults rating neighbourhood as a very good place to live	6.8
Adults incapacity benefit/severe disability allow/employment allow	6.8
Road traffic accident casualties	6
Child obesity in primary	6
New cancer registrations	5.5
People aged 65+ with high care needs cared at home	5.1
Teenage pregnancies	4.7
Patients with a psychiatric hospitalisation	4.3
All mortality among 15-44 year olds	4.3
Deaths from suicide	4.3
Active travel to work	3.8
Child dental health in primary 1	3.8
Secondary school attendance	3.8
Immunisation uptake at 24 months-5 in 1	3
Mothers smoking during pregnancy	3
Violent crimes recorded	3
Population income deprived	3
Immunisation uptake at 24 months-MMR	2.6
Average tariff score of all pupils on S4 roll	2.6
Primary school attendance	2.6
Working age adults with low/no educational qual	2.6
Patients hospitalised with (COPD)	2.6
Early deaths from cancer (<75)	2.1
Young people not in employment education/training	2.1
Smoking prevalence (adults 16+)	2.1
Drug-related hospital stays	2.1
Prisoner population	1.7
Low birth weight	1.7
Single adult dwellings	1.7
Child dental health in primary 7	1.7
Female life expectancy	1.7
Alcohol-related hospital stays	1.3
Children looked after by local authority	1.3
People living in 15% most access deprived areas	1.3
Bowel screening uptake	1.3
Population within 500 metres of a derelict site	1.3
Patients hospitalised with coronary heart disease	1.3
Deaths all ages	0.9
Referrals Childrens Reporter-violence-related off	0.9
Pop growth (2005-2015)	0.9
Drug crimes recorded	0.4
Babies exclusively breastfed at 6-8 weeks	0.4
Estimated smoking attributable deaths	0.4
Male life expectancy	0
Early deaths from CHD (<75)	0
People claiming pension credits (aged 60+)	0

 ISBN: 978-1-61208-682-8

TABLE IV
PEARSON CORRELATION COEFFICIENT RELATED TO HOSPITALISED WITH ASTHMA

Pos	Metric	Correlation
1	Patients hospitalised with COPD	0.859453
2	Patients (65+) with multiple emergency hospital...	0.822299
3	Population 65+ years at risk of hospital admiss...	0.79177
4	Multiple admission patients 65+	0.790819
5	Patients with emergency hospitalisations	0.786936
6	Adults incapacity benefit/severe disability all...	0.786007
7	Estimated smoking attributable deaths	0.740913
8	Population prescribed drugs for anxiety/depress...	0.733842
9	Patients hospitalised with (COPD)	0.693275
10	Adults 65+ years claiming Attendance Allowance	0.683372
11	Children looked after by local authority	-0.683088
12	Deaths all ages	0.671886
13	People Claiming Pension Credits (aged 60+ years)	0.658914
14	Patients registered with cancer - Females	0.653211
15	Alcohol-related hospital stays	0.64417
16	Adults 60+ years claiming incapacity/severe dis...	0.62582
17	Cerebrovascular disease patients	0.624988
18	Early deaths from CHD (<75)	0.622994
19	Male life expectancy	-0.622139
20	Working age adults with low/no educational qual	0.621369
21	Single adult dwellings	0.619495
22	Drug-related hospital stays	0.615702
23	Early deaths from cancer (<75)	0.614767
24	New cancer registrations	0.602819
25	Population (65+) in 15% most access deprived areas	-0.597966
26	Patients hospitalised with coronary heart disease	0.590585
27	Percent of low birthweight (less than 2500g) ba...	0.553096
28	Female life expectancy	-0.547162
29	Deaths from alcohol conditions	-0.524465

TABLE V
MACHINE LEARNING METRICS FOR PATIENTS HOSPITALISED WITH ASTHMA WITH A 86.67% SUCCESS RATE OF PREDICTION

New cancer registrations	People aged 65+ with high care needs cared at home	Children Living in Poverty
Patients (65+) with multiple emergency hospitalisations	Deaths from suicide	Breast screening uptake
Patients (65+) with multiple emergency hospitalisations	Children looked after by local authority	Breast screening uptake
Patients (65+) with multiple emergency hospitalisations	Single adult dwellings	Breast screening uptake
Patients (65+) with multiple emergency hospitalisations	Secondary school attendance	Breast screening uptake
Patients (65+) with multiple emergency hospitalisations	Children Living in Poverty	Adults rating neighbourhood as a very good place to live
Patients (65+) with multiple emergency hospitalisations	Children Living in Poverty	Breast screening uptake
Population prescribed drugs for anxiety/depression/psychosis	People aged 65+ with high care needs cared at home	Working age adults with low/no educ qual

 ISBN: 978-1-61208-682-8

TABLE VI
TRAINING DATA: NEW CANCER REGISTRATIONS, PEOPLE AGED 65+ WITH HIGH CARE NEEDS CARED AT HOME AND CHILDREN LIVING IN POVERTY.
TRAINED AGAINST: PATIENTS HOSPITALISED WITH ASTHMA WITH 30% TRAINING DATA

Index	Predicted	Actual	Diff	Success
Aberdeen City	73.64	73.60	0.04	Success
Aberdeenshire	56.67	57.70	-1.03	Success
Angus	65.69	67.70	-2.01	Success
Argyll and Bute	75.11	71.80	3.31	Success
Clackmannanshire	83.33	77.80	5.53	Success
Dumfries and Galloway	83.83	80.30	3.53	Success
Dundee City	85.45	82.90	2.55	Success
East Ayrshire	104.56	109.00	-4.44	Success
East Dunbartonshire	75.25	78.40	-3.15	Success
East Renfrewshire	75.78	80.70	-4.92	Success
Edinburgh City	94.06	86.80	7.26	Failed!
Falkirk	90.35	91.00	-0.65	Success
Fife	80.98	77.40	3.58	Success
Glasgow City	116.11	119.50	-3.39	Success
Highland	65.61	92.20	-26.59	Failed!
Inverclyde	104.67	103.50	1.17	Success
Lanarkshire	105.53	109.60	-4.07	Success
Mid and East Lothian	103.27	110.30	-7.03	Success
Moray	61.28	60.20	1.08	Success
North Ayrshire	106.91	133.10	-26.19	Failed!
Orkney Islands	52.13	56.60	-4.47	Success
Outer Hebrides	65.66	68.60	-2.94	Success
Perth and Kinross	64.72	63.30	1.42	Success
Renfrewshire	96.24	104.30	-8.06	Failed!
Scottish Borders	69.97	86.10	-16.13	Failed!
Shetland Islands	57.29	39.00	18.29	Failed!
South Ayrshire	96.40	100.90	-4.50	Success
Stirling	68.98	66.00	2.98	Success
West Dunbartonshire	113.28	115.80	-2.52	Success
West Lothian	87.35	87.60	-0.25	Success

 ISBN: 978-1-61208-682-8