Machine Learning for Health and Social Care Demographics in Scotland

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

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

***Abstract*—This paper outlines an extensive study of applying machine learning to the analysis of publicly available health and social care data within Scotland, with a focus on learning the most significant variables involved in key health care outcome factors, such as for male life expectancy and premature deaths. It uses the publicly available data set from ScotPHO Profiles and uses the important metrics from the Profiles for the training. The paper analyses 56 routinely available variables based on local authority regions within Scotland, and then uses linear regression to match them to health risks. A forest regression method is then used to find the best prediction for machine learning methods. Each training variable is then trained against three other variables, which provides 26,235 different models. These models are later assessed for their success using the complete dataset. The top models are assessed for the metrics used. A frequency analysis method is finally used to determine the most defined variables for each of the variables being trained against.**

The results outline the significant factors that match to key health care objectives using a best match machine learning method. Other variables are however more gender-specific for example crime rates in men and claiming pension credits in women for life expectancy. There is a range of success scores for the variables, with many giving a success rate of over 87%. Along with this, there are several significant findings, and a key one is that obesity at primary school has a strong relationship with deaths for those 15-44 years old. In conclusion, the method provides a way of analysing open-source data and provides new insights into contributory factors within the health and social care conditions. It provides a ranked listing of the matches of variables to health and social care factors, and also an ordered list of the most significant variables. These can be used to further focus on health population surveys.

Strengths and limitations of this study are: New methodology in the assessment of variables within health and social care and their linkages with gathered health assessment metrics, using machine learning; Processing time-efficient time for the selection of every possible model for 56 variables; New observations found within variables for health and social care conditions; Scope identifies local authority regions in Scotland, which ranges from highly populated areas, such as within cities, and less populated areas. Metrics gathered can vary across different countries, such as in England; and short-listing of key variables for health and social care related metrics.

***Index Terms*—Open source data, machine learning, demographics, health profiles, ScotPHO**

I. Introduction

The usage of open source data within health and social care increases by the day, and it can provide a key driver in understanding complex relationships with observed events. Often we look at these metrics in isolation and then correlate them to a factor. From this, we can often define a mathematical model that matches our observations to an event. The linkage between smoking rates and cancer rates is well known, and we can often estimate death rates related to cancer if we can observe smoking rates. In this way, we can consider targeted interventions based on evidence, and address primary factors within our observed events. However, health is complex and multiple factors may contribute to an individual patient or population health and well-being. Given there are multiple factors, we thus often need machine learning techniques in order to investigate the models that fit best to our demographic.

In Scotland, which is the UK's northernmost country with 5.295 million populations in an area of 80,077 km^2, health and social care statistics are gathered at the local authority level. This includes cities such as Glasgow, Edinburgh, and Dundee, and also wider areas such as Aberdeenshire, Angus, and Argyll and Bute. The health and social care issues tend to vary across the various region, and where we can move from highly affluent areas of East Dunbartonshire to less affluent areas within West Dunbartonshire. In order to target health and social care services, we thus need to understand the differing dynamics of our demographics, and recognise the key metrics we can use to understand the cause and effect agents responsible for treating within key objective areas. If we want, for example, to reduce the new cancer registrations, what are the metrics that can be used to affect this reduction?

Many health and social care studies try to identify cause and effect using simple linear regression models, such as where we have one predictor (x) for a continuous output (y) [1]:

$$\gamma = \alpha + x\beta + \epsilon \tag{1}$$

With a multi-variable regression, we can have many predictors ($x_1, x_2 \ldots x_n$):

$$\gamma = \alpha + x_1\beta_1 + x_2\beta_2 + ... + x_n\beta_n + \epsilon \tag{2}$$

In [2] the authors define that multivariate and multivariable regression are often used in an inter-changeable way, but that they are two distinct methods. Multivariate analysis uses models that have two or more dependent variables that are dependent variables on the outcome, whereas multivariable analysis has multiple independent or response variables. In the multivariable approach, we can thus identify independent relationships, while investigating confounders.

In this paper, we will use 56 health and well-being metrics and use the forest regression method [3] to understand the key relationships involved, and then define a short-list of health and well-being metrics which are associated with a condition. With random forests we use a number of learning methods in for classification and regression, and which involves the creation of many decision trees when training and then outlining a class to define the classification or mean prediction (regression) related to individual trees. The paper presents new results on the inter-relationship of the factors within health and well-being using a publicly available data set and itemises the metrics which provide the best correlation. It uses a training method of three metrics to train against a target metric and then scores these within a ranking system to give a short-list of the main metrics.

II. Literature Review

There has been an increasing interest in the usage of machine learning methods in health care, especially to understand complex inter-related relationships in health and well-being. Zhou [4] examined risk factors within electronic health records (EHRs) to predict a diagnosis of the condition in secondary care EHRs. For this, they examined 2,238,360 patients over the age of 16, and matched 20,667 to a secondary care rheumatology clinical system. Their focus was on rheumatoid arthritis from primary care and they found 900 predictors (out of a total of 43,100 variables) that can be seen in a primary care record. Using a decision tree model, they managed to reduce the variables in 37 groups of clinical codes, and defined eight significant predictors.

There is also a growing usage of machine learning in terms of understanding mental health-related conditions. Kessler [5] use machine learning to predict factors within suicide for service personnel, and used four machine learning classiers: naive Bayes; random forests; support vector regression; and elastic net penalised regression. They found that 41.5% of suicides occurred within the 12.0% of soldiers who were outpatients seen by mental health specialists. Overall they found 10-14 predictors, and that the risk was highest within 26 weeks of visits. With [5], Kessler compared machine learning into major depressive disorder (MDD) illness against linear regression methods using 1,056 respondents with lifetime issues and found that machine learning methods performed considerably better than regression methods.

Nishida [6] outlined the application of the tree regression methods in understanding lifestyle-related factors to periodontitis risk, and on the strong factors which affect these. Their study analysed 372 Japanese workers and used a questionnaire to analyse lifestyle-related factors, including smoking (in terms of packets per year) and obesity (with BMI measurement). They found that simple logical regression identified significant factors such as age, gender, alcohol consumption, smoking status, BMI, and frequency of tooth-brushing. Although the Tree Regression Method outlined the correlation between some factors (namely, pack-years, BMI and age) and periodontitis, the other factors (namely, alcohol consumption, gender, and tooth-brushing frequency) were not correlated with periodontitis.

Swedish researchers Karlsson et al. [7] used machine learning to predict adverse drug events (ADE) based on information extracted from data from electronic patient records. Physicians sometimes incorrectly assign diagnosis codes because they fail to identify a new medical event as an ADE. This may not only affect patient safety but also the number of reported ADEs, resulting in incorrect risk estimates of prescribed drugs. The researchers selected one ADE (skin eruption due to drugs and medicaments) and applied Random Forest algorithms (RF) to predict potentially missed diagnoses from the deployment set. By applying the RF model on the deployment set, patients were ranked order by the probability of having the selected ADE. The top ten ranked candidates were inspected by a medical expert, which showed that six patients' narratives supported the models' prediction indicating that the ADE was likely missed.

Qu H-Q et al. [8] used machine learning techniques particularly Support Vector Mechanism (SVM) and Bayesian Logistic Regression (BLR) to automatically identify insulin resistance (HOMA-IR) index in an adult Hispanic population of 1854. The adults have been randomly selected on the basis of 2000 Census tract data in the city of Brownsville, Cameron County. The K-means clustering was then used to classify the individual by insulin resistance. Based on the classification results, a series cutoffs of HOMA-IR was evaluated for true positive rate and true negative rate. A Receiver Operator Characteristics (ROC) analysis was then performed based on these two rates in order to identify the best cutoff value. The study presented 3.80 as the best cutoff of HOMA-IR for identifying those with insulin resistance which is dramatically different compared with the popular clinical cutoff of 2.60.

Existing risk prediction models are predominately biophysical in nature, focussing on processes such as physiology, pathology and biochemistry to anticipate harm. Aggregate [9] weighted track-and-trigger systems [10] such as the National Early Warning Score, co-morbidity indexes such as the Charlson and Elixhauser score [10], and biomarkers such as Troponin T are examples of risk prediction models based on these physical processes. Evidence of psychosocial and environmental contributions to an adverse outcome is increasing. For example, social isolation [11], broadband access, and car ownership [12] have been described as independent markers of mortality and hospital admission. There have been calls for a more integrated biopsychosocial approach [13], [14]. How to integrate these domains into a comprehensive risk model is an area of research priority that may be best served by this methodology.

III. Method

Scotland provides a good test case in analysing health and social care factor, especially it has 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 [15], they do not give enough details on the factors which could influence a more in-depth analysis of contributory factors. The research thus uses the publicly available data set from ScotPHO Profiles [1], and uses this to train towards the best machine learning models in order to predict the outcomes. Table VIII outlines the variables gathered and the areas of Scotland covered. As we have 56 variables to train against, 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 tests (Table I). A benchmark of the time to check a model and to match against predicted values gives an estimated time of 0.4 seconds. Table I 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 the work uses three variables to train against. Missing data, while rare, is replaced by the average of the overall data.

To achieve the goal of this research work, the following methodology will be taken:

- First, the variables are correlated, using linear regression, against each other to see significant linkages. For example, "Male life expectancy " and "Patients with emergency hospitalisations" gives a Pearson correlation coefficient of -0.721579 [16], which has a negative correlation meaning that as we increase the male life expectancy in an area, we are likely to reduce the patients being admitted with emergency hospitalisation
- Second, a machine is trained using 30% of the data and then predicts values from the data. In this case the machine is being trained against three parameters within the data in order to determine the most significant factors in the matching.
- Third, each of the models are assessed for the complete dataset, and are measured for a success rate. In this case, a success band within +/-7.5% of difference between the minimum and maximum value is used. For example, if male life expectancy rates varies from 70 to 80 years old, the success range will be 7.5% of 10, which is 0.75, so a predicted value of 75.5 against an actual value of 75 would be a success, but a predicted value of 76 would be a failure. The top contenders for a successful match are then outlined and checked for their usefulness in showing contributory factors.
- Fourth, once all the models have been determined, the top models, for at least 50 successful models, are then selected and the variables from each of the most successful training models are then scored for their success.

Overall for the training of three variables to be trained for one variable with 57, we get 27,720 possible models to compare. The models which tie in their success rate around the 50 rank will be used in the most success rankings.

The paper will present results as follows:

- The correlation between variables. This will do a one-to-one correlation between each of the variables, in order to show a correlation between them.
- The top-rated model for the variable in training. This will outline the best fit for a model in training against a variable. A success is defined as a variation of less than +/-7.5% of the maximum difference within the variable.
- The Top 10 variables which occur most often in the training against the variable. This will outline, in a ranked order, the most successful variables used in the training against a variable.

IV. Correlation Results

The first phase in looking for linkages is to run a correlation analysis, and which will show the variables which are linked. Normally we analyse the R-squared value (which ranges from +1 to -1), and the larger the magnitude, the stronger the possible correlation. Unfortunately, there is no guarantee that the R-squared value actually does show the correlation between variables, thus the next stage will provide a machine learning model which will match a variable to three other variables. R-squared value provides a measure of how well-observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model. In other words, measuring how close the outputs of a model and the predicted values. It should not be used to measure the correlation between two distinct variables.

These results are defined in[**?**], and a sample of the correlation for Male Life Expectancy is given in Appendix A. In this, we can see the strong correlations related to alcohol and smoking, and we can plot the variables involved and determine where there is a strong correlation. All of the linkages are examined in [17].

V. Machine Learning Results

While we can apply linear regression to the data, there are often complex interrelationships that need to be analysed with machine learning. Each of the learning elements is defined in [18], and the result of the Female Life Expectancy machine learning method is outlined in Table VIII. This shows the predicted value, the actual value, and the difference. In this case, a successful prediction is defined as a value of +/-7.5% of the difference between the minimum and maximum value.

VI. Findings

A key objective of any machine learning method is to produce a shortlist of parameters which can be used to give a good estimate of the variable. If we grade each of the variables within at least the Top 50 predictions, and then score the number of occurrences of each variable). We can now take the Top 15 related variables for each of the training targets and define a shortlist of related conditions (tables IV to VII). The full list of results is given in Appendix A. Within Table IV we can see that smoking has a significant effect on life expectancy, but in females, we see that pension credits are a considerable correlation factor. Within the table, we can also see that that

TABLE I. 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

Variable	Best match (% success)
Female Life Expectancy	96.67
Male Life Expectancy	96.67
Deaths all ages	96.67
All mortality among 15-44 year olds	86.67
Early deaths from CHD (<75)	96.67
Early deaths from cancer (<75)	90
Estimated smoking attributable deaths	96.67
Smoking prevalence (adults 16+)	80
Alcohol-related hospital stays	86.67
Deaths from alcohol conditions	86.67
Drug-related hospital stays	73.33
Active Travel to Work	73.33
New cancer registrations	90
Patients hospitalised with asthma	86.67
Patients hospitalised with coronary heart disease	76.67
Patients hospitalised with (COPD)	76.67
Patients (65+) with multiple emergency hospitalisations	86.67
Road traffic accident casualties	76.67
Deaths from suicide	73.33
Adults incapacity benefit/severe disability allow/employment allow	100

TABLE II. Machine Learning for Female Life Expectancy

Childhood obesity in primary school has the highest impact on All morality among 15-44 year olds, which perhaps shows that obesity at primary schools causes problems in later life.

In Table V we see variables which would normally be mapped to conditions, such as alcohol-related hospital stays being strongly correlated with Deaths from alcohol conditions, but the Working-age population claiming Out of Work benefits and Secondary school attendance are perhaps more surprising.

In Table VI we see variables related to mothers, children and young people being mapped to drug-related hospital stays, where Children looked after by local authority giving the strongest correlation. For new cancer registrations deprivation around income and claiming Out-of-work benefits with Child detail health in Primary 1 giving the most significant factors for predicting these rates.

In Table VII we see the normally expected variables for COPD-related to smoking and coronary heart disease, but added to this is Single adult dwellings. For suicide rates, we have one of the lowest success rates (73.3%) and we find the strongest variables are related to Babies exclusively breastfed at 6-8 weeks, and Immunisation uptake at 24 months, which give the best model pointers.

We can now determine the percentage of each variable within each of the variables defined in tables IV – VII, and sum the contribution. This gives Table VIII, which can be used to understand the variables which are most often used within the top models. With the top ranking variables, we see that there are ones related to child health, such as "Child dental health in primary 1" and "Child obesity in primary" are fairly common. For alcohol and smoking we see "Alcohol-related hospital stays", and "Smoking prevalence (adults 16+)". A pointer to deprivation is highlighted with "Patients with emergency hospitalisations", "Working age population employment deprived", and "People living in 15% most access deprived areas'. Other significant factors include "People claiming pension credits (aged 60+)", "Breast screening uptake" and "Early deaths from CHD (<75)".

VII. Conclusions

While linear regression methods are useful for mapping two variables to each other, in complex issues such as health and social, we need complex models with complex relationships. To understand the most important variables for each variable, we need to try out a range of variables to fit the best model. This paper has used three variables to match, and which provides a balance between computing resources and identification of the best variable set.

Secondary analysis of existing large datasets has benefits. It maximises utility of already collected data. It enables the ability to test and alter research hypothesis prior to prospective data collection. It answers research questions without further resource-intensive data collection or potential harm to patients. But there are limitations. The data is retrospective. Data variables and how they are captured within datasets may not be comprehensive or optimal to answer specific research queries. As the data is observational, the association cannot imply causality (though techniques to limit bias exist e.g. propensity analysis, multivariable adjustment and instrument variable analysis).

There are several significant findings, and a key one is that obesity at primary school has a strong relationship with deaths for those 15-44 years old. The benefit of machine learning is the ability to run a large number of models to explore the interaction between variables (n=x) to find the best model fit. This study suggests that there are significant social and environmental factors associated with mortality and hospitalisations. Of note, mortality appears to be strongly associated with childhood factors, with both male and female life expectancy being associated with "Child obesity in primary" and "Immunisation uptake at 24 months-MMR". This association with childhood factors appears to extend to hospitalisation for specific patient groups. For example, "primary school attendance" and "Child obesity in primary" are associated with "Patients hospitalised with (COPD)" and "Patients (65+) with multiple emergency hospitalisations" respectively. This may suggest that social and environmental stressors in childhood is associated with less resilience later in life, findings mirrored by Alvarado et al. [19]. This "life-course" approach to risk prediction, prevention and early intervention is an area of much-needed study [20].

VIII. Appendix A

The data used in the study is at:

https://asecuritysite.com/log/well.csv

and the additional tables are defined at:

https://asecuritysite.com/appendixa.pdf

References

[1] S. H. Observatory, "Scottish Health Observatory." [Online]. Available: http://www.scotpho.org.uk/

[2] E. C. Alexopoulos, "Introduction to multivariate regression analysis," *Hippokratia*, vol. 14, no. Suppl 1, p. 23, 2010.

[3] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1. IEEE, 1995, pp. 278–282.

[4] S.-M. Zhou, F. Fernandez-Gutierrez, J. Kennedy, R. Cooksey, M. Atkinson, S. Denaxas, S. Siebert, W. G. Dixon, T. W. O'Neill, E. Choy *et al.*, "Defining disease phenotypes in primary care electronic health records by a machine learning approach: a case study in identifying rheumatoid arthritis," *PloS one*, vol. 11, no. 5, p. e0154515, 2016.

[5] R. C. Kessler, M. B. Stein, M. V. Petukhova, P. Bliese, R. M. Bossarte, E. J. Bromet, C. S. Fullerton, S. E. Gilman, C. Ivany, L. Lewandowski-Romps *et al.*, "Predicting suicides after outpatient mental health visits in the army study to assess risk and resilience in servicemembers (army starrs)," *Molecular psychiatry*, vol. 22, no. 4, p. 544, 2017.

[6] N. Nishida, M. Tanaka, N. Hayashi, H. Nagata, T. Takeshita, K. Nakayama, K. Morimoto, and S. Shizukuishi, "Determination of Smoking and Obesity as Periodontitis Risks Using the Classification and Regression Tree Method," *Journal of Periodontology*, vol. 76, no. 6, pp. 923–928, 2005. [Online]. Available: http://www.joponline.org/doi/10.1902/jop.2005.76.6.923

[7] I. Karlsson, J. Zhao, L. Asker, and H. Boström, "Predicting adverse drug events by analyzing electronic patient records," in *Conference on Artificial Intelligence in Medicine in Europe*. Springer, 2013, pp. 125–129.

[8] H.-Q. Qu, Q. Li, A. R. Rentfro, S. P. Fisher-Hoch, and J. B. McCormick, "The definition of insulin resistance using homa-ir for americans of mexican descent using machine learning," *PloS one*, vol. 6, no. 6, p. e21041, 2011.

[9] G. B. Smith, D. R. Prytherch, P. E. Schmidt, and P. I. Featherstone, "Review and performance evaluation of aggregate weighted 'track and trigger'systems," *Resuscitation*, vol. 77, no. 2, pp. 170–179, 2008.

[10] G. B. Smith, D. R. Prytherch, P. Meredith, P. E. Schmidt, and P. I. Featherstone, "The ability of the national early warning score (news) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death," *Resuscitation*, vol. 84, no. 4, pp. 465–470, 2013.

[11] A. Bottle and P. Aylin, "Comorbidity scores for administrative data benefited from adaptation to local coding and diagnostic practices," *Journal of clinical epidemiology*, vol. 64, no. 12, pp. 1426–1433, 2011.

[12] M. Pantell, D. Rehkopf, D. Jutte, S. L. Syme, J. Balmes, and N. Adler, "Social isolation: a predictor of mortality comparable to traditional clinical risk factors," *American journal of public health*, vol. 103, no. 11, pp. 2056–2062, 2013.

[13] R. Conway, S. Cournane, D. Byrne, D. O'Riordan, S. Coveney, and B. Silke, "Broadband access, housing and car ownership as deprivation predictors of the emergency medical admission," *European journal of internal medicine*, vol. 35, pp. e25–e26, 2016.

[14] G. L. Engel, "The need for a new medical model: a challenge for biomedicine," *Science*, vol. 196, no. 4286, pp. 129–136, 1977.

[15] SIMD, "Scottish index of multiple deprivation," 2018. [Online]. Available: http://www.gov.scot/Topics/Statistics/SIMD

[16] Asecuritysite, "Well-being data set," accessed 2019-02-01. [Online]. Available: http://asecuritysite.com/log/well.csv

[17] ——, "Plotting data." [Online]. Available: http://asecuritysite.com/bigdata/pplot_2?file=well.csv

[18] ——, "Machine learning." [Online]. Available: http://asecuritysite.com/bigdata/ml?file=well.csv

[19] N. De Witte, R. Gobbens, L. De Donder, S. Dury, T. Buffel, J. Schols, and D. Verté, "The comprehensive frailty assessment instrument: development, validity and reliability," *Geriatric Nursing*, vol. 34, no. 4, pp. 274–281, 2013.

[20] B. E. Alvarado, M.-V. Zunzunegui, F. Béland, and J.-M. Bamvita, "Life course social and health conditions linked to frailty in latin american older men and women," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 63, no. 12, pp. 1399–1406, 2008.

TABLE III. Top matches (Female Life Expectancy)

Top	Female Life Expectancy	Male Life Expectancy	Deaths all ages	All mortality among 15-44 year olds	Early deaths from CHD (<75)
2	Deaths all ages	People claiming pension credits (aged 60+)	Estimated smoking attributable deaths	Child obesity in primary	Estimated smoking attributable deaths
3	Male life expectancy	Deaths all ages	Male life expectancy	Early deaths from CHD (<75)	Male life expectancy
4	All mortality among 15-44 year olds	Female life expectancy	Female life expectancy	Male life expectancy	Female life expectancy
5	Patients hospitalised with coronary heart disease	Working age population employment deprived	Adults rating neighbourhood as a very good place to live	Working age population claiming Out of Work benefits	Adults rating neighbourhood as a very good place to live
6	Child obesity in primary	Crime rate	Child dental health in primary 1	Adults incapacity benefit/ severe disability allow/ employment allow	Child dental health in primary 1
7	Early deaths from CHD (<75)	Immunisation uptake at 24 months-MMR	Early deaths from cancer (<75)	Prisoner population	Early deaths from cancer (<75)
8	People claiming pension credits (aged 60+)	Early deaths from cancer (<75)	Adults incapacity benefit/severe disability allow/employment allow	Immunisation uptake at 24 months-MMR	Adults incapacity benefit/severe disability allow/employment allow
9	Secondary school attendance	Violent crimes recorded	Population within 500 metres of a derelict site	Referrals Childrens Reporter-violence-related off	Population within 500 metres of a derelict site
10	Immunisation uptake at 24 months-MMR	Adults incapacity benefit/severe disability allow/employment allow	Working age adults with low/no educational qual	Female life expectancy	Working age adults with low/no educational qual

TABLE IV. Top matches (Male Life Expectancy)

Top	Early deaths from cancer (<75)	Estimated smoking attributable deaths	Smoking prevalence (adults 16+)	Alcohol-related hospital stays	Deaths from alcohol conditions
2	Deaths all ages	Deaths all ages	Estimated smoking attributable deaths	Male life expectancy	Alcohol-related hospital stays
3	Estimated smoking attributable deaths	Early deaths from CHD (<75)	Child dental health in primary 1	Patients (65+) with multiple emergency hospitalisations	Working age population claiming Out of Work benefits
4	Female life expectancy	New cancer registrations	Patients hospitalised with (COPD)	Patients hospitalised with asthma	Secondary school attendance
5	All mortality among 15-44 year olds	Smoking prevalence (adults 16+)	Child dental health in primary 7	Drug crimes recorded	Male life expectancy
6	Crime rate	Active travel to work	Patients (65+) with multiple emergency hospitalisations	Low birth weight	Working age population employment deprived
7	New cancer registrations	Patients with emergency hospitalisations	Deaths all ages	Child dental health in primary 7	Bowel screening uptake
8	Child dental health in primary 1	Male life expectancy	Patients hospitalised with asthma	Patients with emergency hospitalisations	Violent crimes recorded
9	Adults incapacity benefit/severe disability allow/employment allow	Patients hospitalised with asthma	Children looked after by local authority	Single adult dwellings	Primary school attendance
10	Secondary school attendance	Drug-related hospital stays	Drug crimes recorded	Prisoner population	Patients hospitalised with coronary heart disease

TABLE V. Top matches (Deaths all ages)

Top	Drug-related hospital stays	Active Travel to Work	New cancer registrations	Patients hospitalised with asthma	Patients hospitalised with coronary heart disease
2	Children looked after by local authority	People living in 15% most access deprived areas	Child dental health in primary 1	Patients with emergency hospitalisations	Alcohol-related hospital stays
3	Mothers smoking during pregnancy	Violent crimes recorded	Population income deprived	Patients (65+) with multiple emergency hospitalisations	Population within 500 metres of a derelict site
4	Young people not in employment education/training	Single adult dwellings	Working age population claiming Out of Work benefits	Population prescribed drugs for anxiety/depression/psychosis	Drug crimes recorded
5	Patients with a psychiatric hospitalisation	Prisoner population	Working age population employment deprived	Breast screening uptake	Estimated smoking attributable deaths
6	Population within 500 metres of a derelict site	Patients with a psychiatric hospitalisation	Children Living in Poverty	Children Living in Poverty	Smoking prevalence (adults 16+)
7	Breast screening uptake	Breast screening uptake	Working age adults with low/no educational qual	Working age population claiming Out of Work benefits	Patients (65+) with multiple emergency hospitalisations
8	Deaths from suicide	Bowel screening uptake	Child dental health in primary 7	Domestic Abuse	New cancer registrations
9	Female life expectancy	Estimated smoking attributable deaths	Drug-related hospital stays	Working age population employment deprived	Early deaths from cancer (<75)
10	Early deaths from cancer (<75)	Early deaths from cancer (<75)	Prisoner population	Adults incapacity benefit/severe disability allow/employment allow	Early deaths from CHD (<75)

TABLE VI. Top matches (All mortality among 15-44 year olds)

Top	Patients hospitalised with (COPD)	Patients (65+) with multiple emergency hospitalisations	Road traffic accident casualties	Deaths from suicide	Adults incapacity benefit/severe disability allow/employment allow
2	Smoking prevalence (adults 16+)	Patients with emergency hospitalisations	Patients hospitalised with asthma	Babies exclusively breastfed at 6-8 weeks	Working age population employment deprived
3	Patients hospitalised with coronary heart disease	Patients hospitalised with asthma	Smoking prevalence (adults 16+)	Immunisation uptake at 24 months-5 in 1	Working age population claiming Out of Work benefits
4	Single adult dwellings	Female life expectancy	Child obesity in primary	Young people not in employment education/training	Breast screening uptake
5	Patients with emergency hospitalisations	Child obesity in primary	People living in 15% most access deprived areas	Breast screening uptake	Drug crimes recorded
6	Early deaths from CHD (<75)	Bowel screening uptake	Breast screening uptake	Patients hospitalised with (COPD)	Active travel to work
7	Population within 500 metres of a derelict site	Deaths all ages	Children looked after by local authority	Immunisation uptake at 24 months-MMR	Immunisation uptake at 24 months-5 in 1
8	Adults rating neighbourhood as a very good place to live	Working age adults with low/no educational qual	Bowel screening uptake	All mortality among 15-44 year olds	Children Living in Poverty
9	Bowel screening uptake	Babies exclusively breastfed at 6-8 weeks	Drug crimes recorded	Smoking prevalence (adults 16+)	Early deaths from cancer (<75)
10	Drug-related hospital stays	People claiming pension credits (aged 60+)	Immunisation uptake at 24 months-MMR	Patients (65+) with multiple emergency hospitalisations	Child dental health in primary 1

TABLE VII. Top metrics

Deaths all ages	85
Child dental health in primary 1	75
Patients with emergency hospitalisations	74
Working age population employment deprived	58
Alcohol-related hospital stays	58
People living in 15% most access deprived areas	54
All mortality among 15-44 year olds	54
Child obesity in primary	54
People claiming pension credits (aged 60+)	53
Smoking prevalence (adults 16+)	51
Breast screening uptake	50
Early deaths from CHD (<75)	50
Children looked after by local authority	47
Estimated smoking attributable deaths	47
Working age population claiming Out of Work benefits	46
Male life expectancy	45
Drug crimes recorded	43
Population within 500 metres of a derelict site	43
Babies exclusively breastfed at 6-8 weeks	42
Population income deprived	41
Patients hospitalised with asthma	40
Patients (65+) with multiple emergency hospitalisations	37
Female life expectancy	36
New cancer registrations	35
Child dental health in primary 7	34
Immunisation uptake at 24 months-MMR	33
Violent crimes recorded	33
Bowel screening uptake	33
Children Living in Poverty	32
Patients with a psychiatric hospitalisation	31
Young people not in employment education/training	31
Adults incapacity benefit/severe disability allow/employment allow	30
Teenage pregnancies	28
Early deaths from cancer (<75)	27
Prisoner population	27
Primary school attendance	27
Patients hospitalised with (COPD)	27
Population prescribed drugs for anxiety/depression/psychosis	27
Adults rating neighbourhood as a very good place to live	26
Working age adults with low/no educational qual	26
Secondary school attendance	25
Single adult dwellings	25
Patients hospitalised with coronary heart disease	25
Immunisation uptake at 24 months-5 in 1	24
Domestic Abuse	24
Crime rate	23
Mothers smoking during pregnancy	23
Low birth weight	22
Road traffic accident casualties	20
Referrals Childrens Reporter-violence-related off	20
Active travel to work	19
Deaths from suicide	19
People aged 65+ with high care needs cared at home	16
Drug-related hospital stays	16
Pop growth (2005-2015)	12
Average tariff score of all pupils on S4 roll	11