

Insights into Antidepressant Prescribing Using Open Health Data

Brian Cleland¹, Jonathan Wallace¹, Raymond Bond¹, Michaela Black², Maurice Mulvenna¹, Deborah Rankin² and Austin Tanney³

¹ Ulster University, Jordanstown Campus, Northern Ireland, UK

² Ulster University, Magee Campus, Northern Ireland, UK
{b.cleland, jg.wallace, rb.bond, mm.black, md.mulvenna}@ulster.ac.uk

³ Analytics Engines, 1 Chlorine Gardens, Belfast, Northern Ireland, UK
a.tanney@analyticsengines.com

Abstract. The growth of big data is transforming many economic sectors, including the medical and healthcare sector. Despite this, research into the practical application of data analytics to the development of health policy is still limited. In this study we examine how data science and machine learning methods can be applied to a variety of open health datasets, including GP prescribing data, disease prevalence data and economic deprivation data. This paper discusses the context of mental health and antidepressant prescribing in Northern Ireland and highlights its importance as a public policy issue. A hypothesis is proposed, suggesting that the link between antidepressant usage and economic deprivation is mediated by depression prevalence. An analysis of various heterogeneous open datasets is used to test this hypothesis. A description of the methodology is provided, including the open health datasets under investigation and an explanation of the data processing pipeline. Correlations between key variables and several different clustering analyses are presented. Evidence is provided which suggests that the depression prevalence hypothesis is flawed. Clusters of GP practices based on prescribing behaviour and disease prevalence are described and key characteristics are identified and discussed. Possible policy implications are explored and opportunities for future research are identified.

Keywords: Health Policy, Data Analytics, Big Data, Prescribing, Prevalence, Machine learning, Deprivation.

1 Introduction

As the influence of data science has grown across industry and society, so the term “big data” has become widely used to describe the phenomenon. Every economic sector has been affected by this trend, and healthcare is no exception [1, 30, 17]. Governments worldwide have begun to include the impact of big data in their policy statements. The UK government recently stated that big data had “huge unrealised potential, both as a driver of productivity and as a way of offering better products and services to citizens” [15]. Despite this, research on the implications of big data to medical policymaking and service delivery remains relatively limited. With the exception of some public health surveillance [31, 5] and pharmacovigilance systems [33], the opportunities for better policymaking using big data are still largely unexplored.

In this study we provide an analysis of the potential application of big data and machine learning methods to the development of public policy and service delivery. In particular, we focus on the use of heterogeneous open data from a variety of online sources, including disease prevalence data, GP prescribing data and economic deprivation data. We examine how such datasets can be brought together and analysed in such a way as to generate usable, actionable insights for clinicians, policymakers and the general public.

2 Related Work

2.1 Mental Health Policy in Northern Ireland

In Northern Ireland, as in other parts of the UK [27], mental health has been identified as a priority policy area for service provision. In a report produced for the Northern Ireland Assembly, Betts and Thompson [4] state that mental illness is the single largest cause of ill health and disability. They note that 318 suicides were registered in NI during 2015, the highest since records began in 1970 and a 19% increase on the suicides recorded in 2014. The report refers to calls for a ten-year regional mental health strategy and a mental health champion to lead work across government departments. Specific policy challenges identified by the authors include the need for more personalised models of care, the need to address stigmas around mental health, improved access to services and more GP training. The study also notes that Northern Ireland is also lagging in the provision of psychological therapies such as psychotherapy, cognitive behavioural therapy (CBT), and trauma therapy.

Various factors particularly impact mental health policy in Northern Ireland. According to the government figures, Northern Ireland has a 25% higher prevalence of mental illness compared to England. It also has lower levels of public spending on health services, with health services accounting for 19.7% of the public budget, in comparison with 22% in England, 20.4% in Scotland, and 20.3% in Wales [24]. The region experiences higher levels of suicide - according to Office of National Statistics figures, suicide rates are 16.4 per 100,000 population, whereas the equivalent rates in England, Wales and Scotland are 10.3, 9.2 and 15.4 respectively. Mental health-related issues in Northern Ireland may be due at least in part to the historical conflict [3, 6]. Finally, the number of economically inactive adults is 28.4% - 5% above the UK average [28].

2.2 Antidepressant Prescribing in Northern Ireland

In a report on mental health in Northern Ireland, the Mental Health Foundation [24] states that "according to prescribing trends, Northern Ireland has significantly higher

levels of depression than the rest of the UK.” This statement assumes that prescribing data can be used as an indicator for underlying health phenomena, and more specifically, that antidepressant prescribing in particular reflects levels of depression in the wider population. This use of prescribing data as a proxy variable for the prevalence of mental illness is not limited to this particular report.

Another example of this assumption being adopted is a pair of studies looking at mental health impacts and burdens in Northern Ireland using administrative data [20, 21]. In these analyses, prescribing and other forms of administrative data, such as social and economic indicators and life event data, are used to determine the effect on mental health of factors such as deprivation, bereavement, care-giving and transition into care. Findings from these studies suggest that the impact on the mental health of individuals of such circumstances are very significant. The authors also found that prescribing rates varied widely between GP practices, although they speculate that this might be explained by differences in practice population composition or levels of deprivation in the practice area.

Another interesting, non-academic, analysis of the same subject using similar data sources was published by the Detail Data [22]. This data journalism piece points out that compared with a major international study by the OECD, antidepressant usage in Northern Ireland is significantly higher than any of the 23 countries surveyed [22]. Antidepressant prescription rates in Northern Ireland stand at 129 daily doses per thousand, compared with the overall UK figure of 72 daily doses per thousand. The authors also demonstrate that there was a strong correlation between economic deprivation and levels of antidepressant usage. Interestingly, their analysis shows that depression prevalence is not correlated with either economic deprivation or antidepressant prescribing. When asked what factors might be behind increasing levels of antidepressant prescribing, GPs point to growing public awareness and patient demand as a driver.

2.3 Summary of Policy Context

Looking at mental health in Northern Ireland, it is clear that the region has important challenges in this area, and there are some very specific political, social and economic factors that must be taken into account. The burden of mental health issues in Northern Ireland is significantly greater than in other parts of the UK. While the region experiences higher rates of suicide and illnesses such as depression and anxiety, it is also faced with lower levels of public spending on health compared to other parts of the UK. There has been some use of administrative data to try to understand the policy implications of issues in Northern Ireland. Such studies have attempted to illustrate the impact of factors such as economic deprivation and bereavement on illnesses such as depression and anxiety. In some of these studies prescribing rates have been used as a proxy variable for mental health issues. There

is some evidence, however, that the link between prescribing and prevalence is not straightforward.

3 Materials and Methods

3.1 Hypothesis: Depression Prevalence as a Mediating Factor Between Economic Deprivation and Antidepressant Prescribing

Much of the literature on the use of prescribing data as a proxy for public health suggests that the correlation between disease prevalence and prescribing is sufficiently strong to make such analyses useful in the development of public policy [12, 14]. Based on this assumption, a number of studies have examined the correlation between economic deprivation and prescribing and used this to propose that economic factors have a measurable impact on mental health. We argue that these studies contain an implicit assumption that *depression prevalence is a mediating factor in the relationship between economic deprivation and antidepressant prescribing* (see Fig. 1). There is, however, some evidence that the link between disease prevalence and prescribing levels is not straightforward or reliable [19].

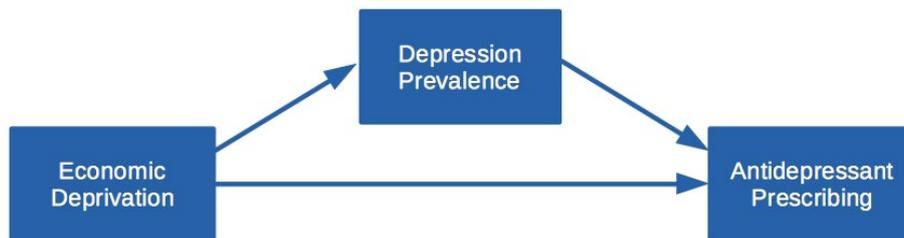


Fig. 1. Hypothesis: mental health is a mediating factor linking economic deprivation and antidepressant prescribing (arrows indicate mediating relationships).

In this study, we tested the hypothesis presented in Fig. 1 by using open data drawn from multiple publicly available sources. Specifically, we examined the links between three major variables - economic deprivation, depression prevalence and antidepressant prescribing - in order to explore the correlations between them. We also examined correlations between these variables and other disease prevalence data, and with GP prescribing data for other drug groups. Finally, we applied a k-means clustering algorithm to determine if meaningful GP practice sub-groups could be identified from the overall dataset.

3.2 Open Health Datasets

GP Prescribing Data

Antidepressant prescribing data was downloaded from the Open Data NI portal [37], operated by the Northern Ireland Department of Finance. For the purposes of this study, data for the twelve months of 2016 was used. In order to provide a direct comparison with international figures, it was necessary to classify the data according to the international Anatomical Therapeutic Chemical (ATC) Classification System [45]. Since UK prescribing data is encoded using the British National Formulary (BNF) standard, each drug had to be re-classified. This was done using a dataset provided by NHS Dictionary of Medicines and Devices (dm+d) [33]. Classifying the data according to the ATC system also allowed prescribing levels to be compared across different drugs. This was done using a “defined daily dose” (DDD), defined by the World Health Organisation as “the assumed average maintenance dose per day for a drug used for its main indication in adults” [45]. Due to some gaps in the NHS dm+d data, only chapters 1-10 of the BNF system could be meaningfully analysed in this study.

Economic Deprivation Data

The metric for economic deprivation used in this study was the Multiple Deprivation Measure (MDM), which is published by the Northern Ireland Statistics and Research Agency (NISRA) [34]. Unlike in other parts of the UK, individual GP practice data does not include deprivation measures. In order to link GP practices to the MDM data, the postcode for each practice was obtained from the Detail Data portal [10], an open data resource provided by the Northern Ireland Council for Voluntary Action. The online MySociety Mapit service [31] was then used to convert the postcode data into super output areas (SOA). By linking each practice to a specific SOA we could then assign a deprivation measure for each prescriber, based on the data supplied by NISRA [34]. While deprivation measures are available for other geographic boundaries, super output areas were chosen due to their relatively high granularity and stability compared to other boundaries such as electoral wards.

Disease Prevalence Data

Disease prevalence data, that is the number of patients per 1000 diagnosed as suffering from different diseases, is available across most of the UK under the Quality Outcomes Framework (QOF), a collection of data which is designed to measure GP performance in order to support GP payments. In Northern Ireland the QOF data no longer includes disease prevalence figures, but fortunately these are still published separately by the Department of Health [15]. Since the prevalence data is linked directly to GP practice identifiers, it allows for a direct comparison between the number of patients being diagnosed with depression and the amount of antidepressants being prescribed.

3.3 Analytics Tools and Pipeline

Jupyter Notebooks [20] were used to algorithmically restructure, transform and merge datasets were required. Pandas [38] and NumPy [35] were used for the statistical analysis. Visualisation of the data was also done through Jupyter Notebooks using Matplotlib [41] and Seaborn [42] for charts and graphs and iPyLeaflet [19] for maps. Correlations between the key variables were explored using the Pearson correlations and p-values. The scikit-learn library [41] was employed to perform K-means clustering on the data.

Two major challenges were addressed in the creation of the data pipeline. The first was identifying the defined daily dosage (DDD) for each drug in the GP prescribing datasets. This was required so that prescribing patterns could be meaningfully compared across GP practices. In order to do this, the BNF codes used by the UK authorities had to be linked to the World Health Organisation's ATC classification system, since only the latter provides standardised DDD data. After some searching it was discovered that the NHS Business Services Authority had created a table or correspondences between BNF and ATC codes for its own use, and that this table was available as open data from their website.

The second challenge was assigning economic deprivation measures to the individual GP practices. Unlike in the rest of the UK, GP practice data does not include this information. This problem was resolved by using practice data to interrogate the online Mapit service (provided by the not-for-profit MySociety organisation), which allowed a Special Output Area (SOA) to be assigned to each practice. These SOAs were then matched to the economic deprivation data from NISRA to allocate a multiple deprivation measure (mdm) to each practice.

The resulting data pipeline is outlined in Fig. 2.

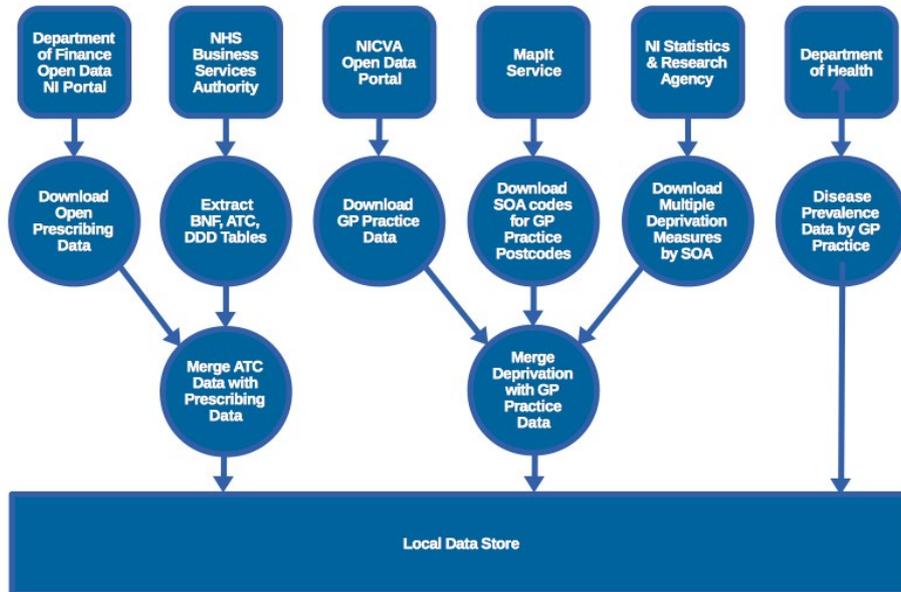


Fig. 2. Data flow diagram illustrating the main data sources and data processing activities.

3.4 Results

Variation in Prescribing Patterns Across GP Practices

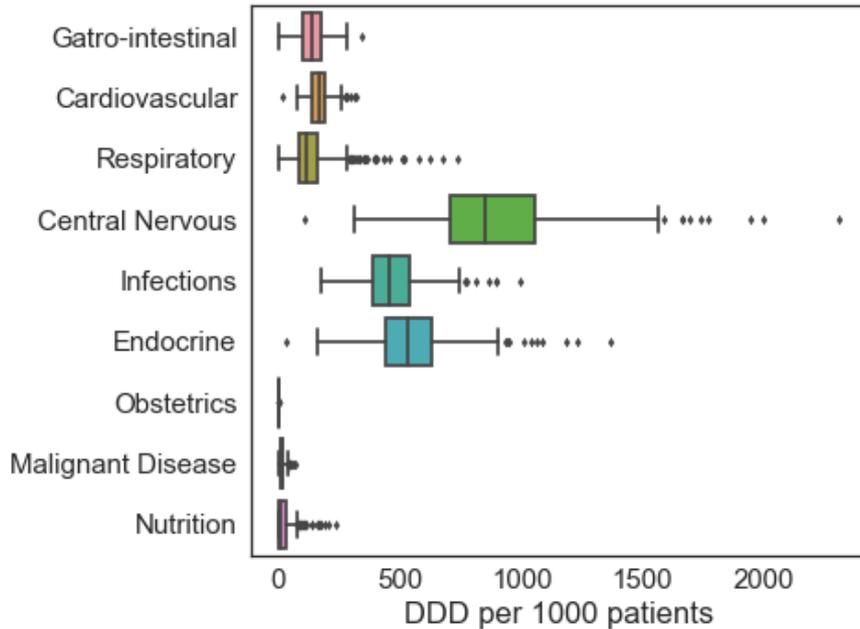


Fig. 3. Box plot illustrating variations in prescribing behaviour across GP practices.

Prescribing behaviour for all GP Practices was analysed according to how many defined daily doses were prescribed by each practice per patient per month, grouped by BNF chapter. Prescribing levels for each drug type were then visualised on a box plot (Fig. 3). As can be seen from the diagram, the most heavily prescribed drugs are those for central nervous system disorders (including antidepressants), infections, and endocrine systems disorders (including insulin). Large variations in prescribing are visible across almost all drug categories. Central nervous system prescribing (including antidepressants) has a noticeable right-skew, suggesting that a relatively small proportion of GPs are prescribing unusually high levels of these medicines.

Variation in Disease Prevalence Across GP Practices

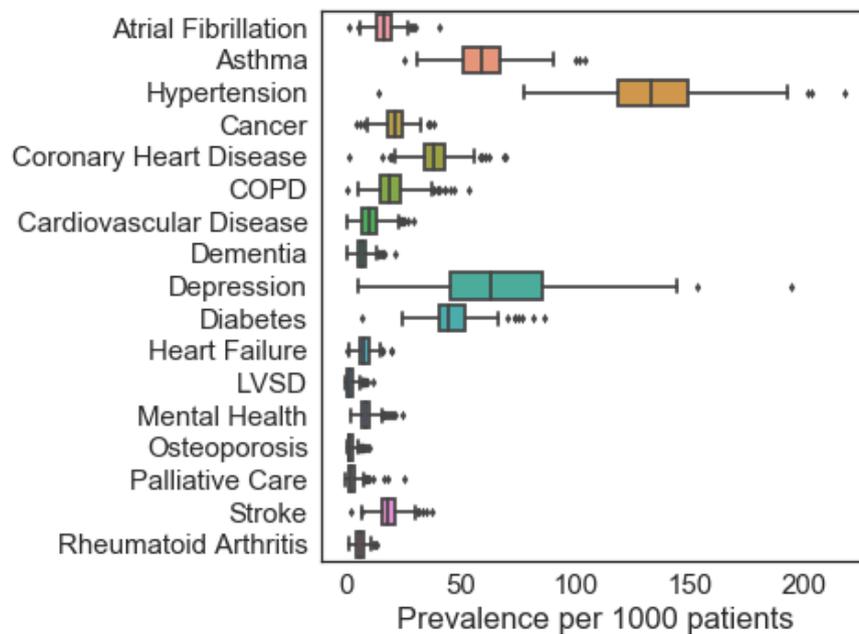


Fig. 4. Box plot illustrating variations in disease prevalence across GP practices.

A boxplot (Fig. 4) was also used to illustrate the variation in disease prevalence across all GP practices in Northern Ireland. As might be expected, the most visible variations appear in the diagnosis of more prevalent disease categories, such as depression, hypertension, asthma and diabetes. Both depression and diabetes have a noticeable right-skew, suggesting that there are a relatively small number of practices with particularly high prevalences of those diseases.

Correlation Between Antidepressants, Deprivation and Other Prescribing

A matrix was calculated showing the Pearson correlations between multiple measures of deprivation (mdm), the defined daily doses per patient of antidepressants (ddd_per_patient) and prescribing levels for other drug groups based on BNF chapter. These values were then visualised in a heatmap (Fig. 5). Key correlations with economic deprivation and antidepressant prescribing are described briefly below.

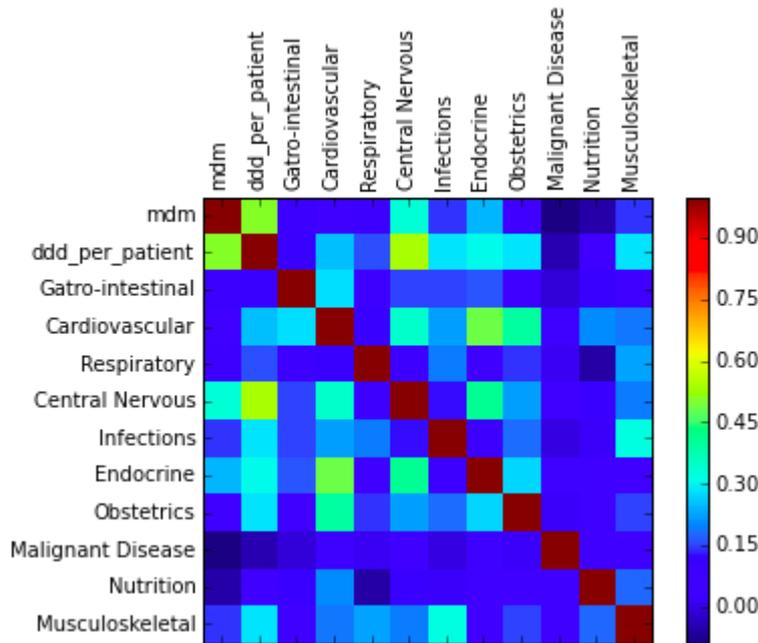


Fig. 5. Heat map showing correlations between economic deprivation (mdm), antidepressant prescribing (ddd_per_patient), and other drug prescribing according to BNF chapter.

Economic deprivation and other prescribing: Economic deprivation was strongly correlated with central nervous system prescribing ($r=0.34$) and moderately correlated endocrine system prescribing ($r=0.24$).

Antidepressant prescribing and other prescribing: Antidepressant prescribing was strongly correlated with central nervous system prescribing ($r=0.54$) and endocrine prescribing ($r=0.31$). It was also moderately correlated with prescribing for cardiovascular ($r=0.25$), infections ($r=0.29$), obstetrics ($r=0.29$) and musculoskeletal (0.29).

Correlation Between Antidepressants, Deprivation and Other Diseases

A second matrix was generated showing the relationships between multiple measures of deprivation (mdm), the defined daily doses per patient of antidepressants (ddd_per_patient) and disease prevalences for all practices. Once again, a heatmap was used to visualise the results (Fig. 6). Brief descriptions of the findings are given below.

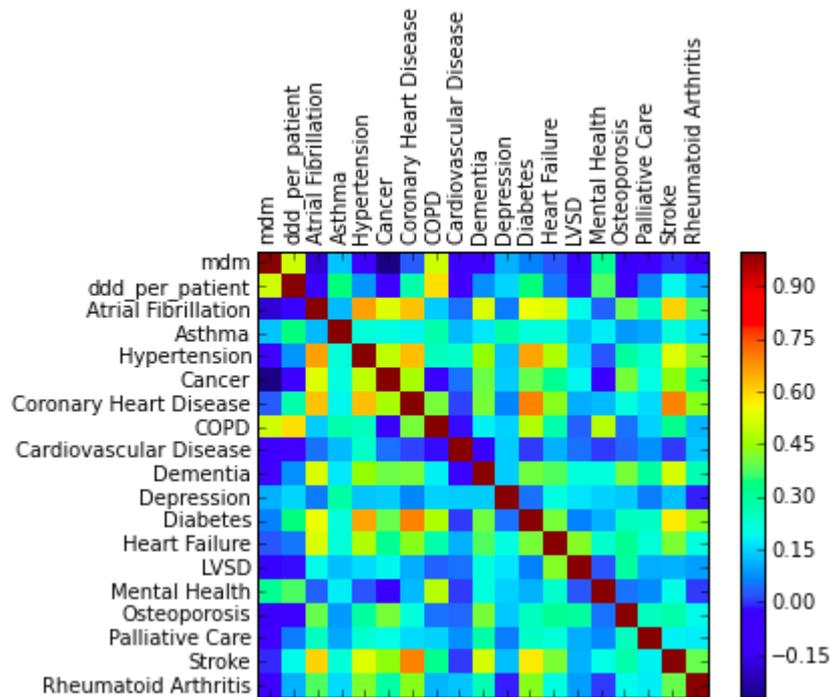


Fig. 6. Heat map showing correlations between economic deprivation (mdm), antidepressant prescribing (ddd_per_patient), and other disease prevalences.

Antidepressant prescribing and economic deprivation: There was strong correlation ($r=0.51$) between levels of antidepressant prescribing and the multiple deprivation measure for each practice.

Depression prevalence and economic deprivation: There was a weak correlation ($r=0.12$) between the prevalence of depression among patients of a given GP practice and the multiple deprivation measure for the area in which the practice resides.

Depression prevalence and antidepressant prescribing: There was a weak correlation ($r=0.15$) between the prevalence of depression among patients of a given GP practice and the multiple deprivation measure for the area in which the practice resides.

Economic deprivation and other disease prevalences: Economic deprivation was particularly strongly correlated with chronic obstructive pulmonary disease ($r=0.51$) and mental health ($r=0.31$) prevalence.

Antidepressant prescribing and other diseases: Antidepressant prescribing was particularly strongly correlated with asthma ($r=0.34$), chronic obstructive pulmonary disease ($r=0.59$), diabetes ($r=0.34$) and mental health prevalence ($r=0.37$).

Clustering GP Practices Based on Prescribing Patterns

GP practices were clustered using a k-means algorithm [9], based on the prescribing patterns according to the defined daily dosage (DDD) per patient per month for each BNF chapter group. The default stopping criterion for the scikit-learn library [41] was used, which tests for convergence within a predetermined tolerance. The whitening algorithm from the scikit-learn library was also used to pre-process the data before clustering [9]. The “elbow method” [9] was applied to determine a suitable number of clusters for the analysis (see Fig. 7). According to this criterion, a two-cluster and three-cluster analysis of the data were undertaken.

The mean feature values for the two-cluster analysis can be viewed in Fig. 8, where each line represents the features of a cluster centroid. It can be seen from the chart that the cluster A is differentiated by a higher level of economic deprivation (mdm) and higher levels of prescribing across all BNF chapters, with the exception of “Malignant Disease”. The results of the three-cluster analysis are visible in Fig. 9. Cluster C (the “deprived practices”), perhaps unsurprisingly, shows higher levels of prescribing across all categories except “Malignant Disease”. Clusters A and B (the “non-deprived practices”) show significant differences in prescribing levels, with cluster A being higher in every category.

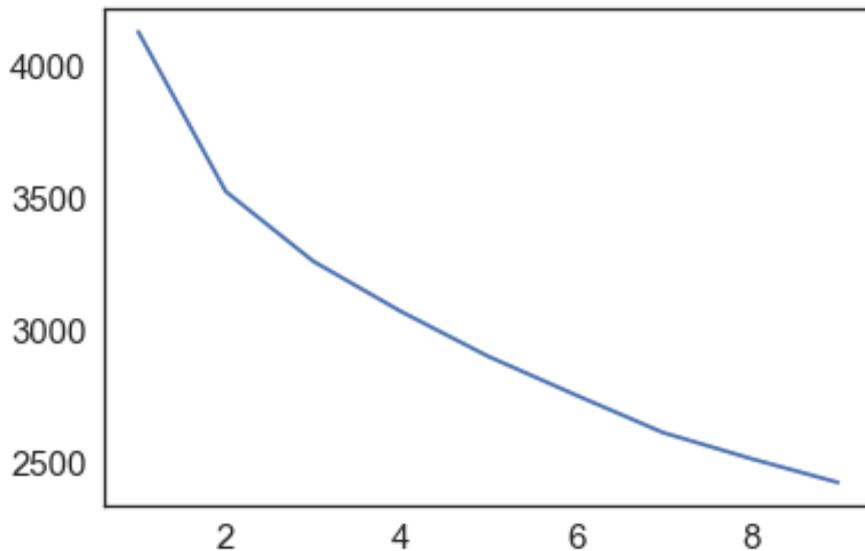


Fig. 7. Elbow criterion plot for clusters based on GP prescribing patterns.

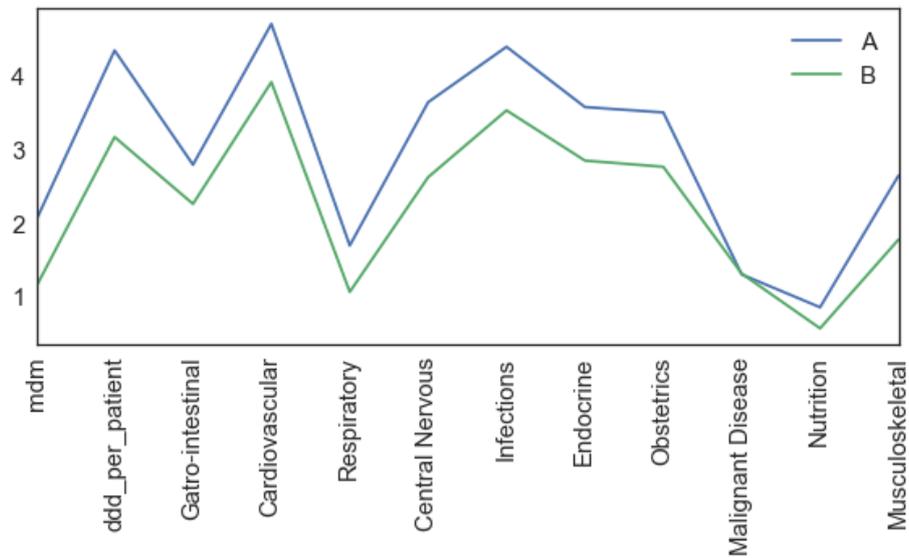


Fig. 8. Line chart illustrating the mean feature values for each cluster based on a two-cluster analysis of GP prescribing patterns.

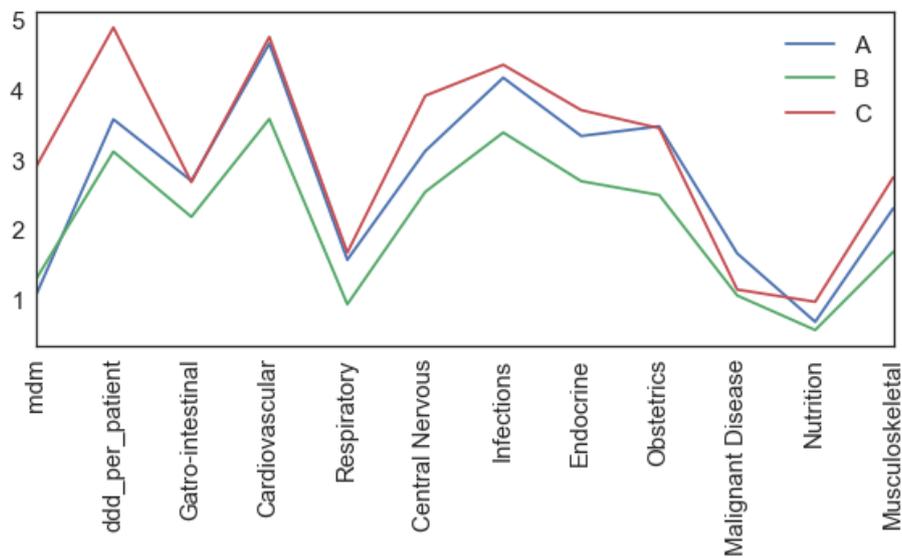


Fig. 9. Line chart illustrating the mean feature values for each cluster based on a three-cluster analysis of GP prescribing patterns.

Clustering GP Practices Based on Disease Prevalence

GP practices were also clustered based on disease prevalence according to data from the NI Department of Health Quality Outcomes Framework. The data was once again whitened, and the “elbow method” was used to determine a suitable number of clusters for the analysis (see Fig. 10). This time, a two-cluster and four-cluster analysis of the data were undertaken.

Fig. 11 shows the mean feature values for each cluster for the two-cluster analysis, with each line illustrating the features of a cluster centroid. It can be seen from this visualisation that the two clusters have similar levels of deprivation (mdm), but that cluster A has significantly higher disease prevalence levels across all categories compared to cluster B. The difference is particularly pronounced in both dementia and palliative care, perhaps suggesting an age difference among patients between the clusters.

The four-cluster analysis (Fig. 12) reveals some other interesting characteristics. Cluster C (the “deprived practices”) has higher economic deprivation levels and medium levels of prevalence for most diseases. Clusters A, C and D (the “non-deprived practices”) show a range of disease prevalence across all the categories. The most interesting divergence is between cluster B, which has low deprivation and relatively low disease prevalence across the board, and cluster D, which also has a low deprivation score, but much higher prevalence for almost all disease categories.

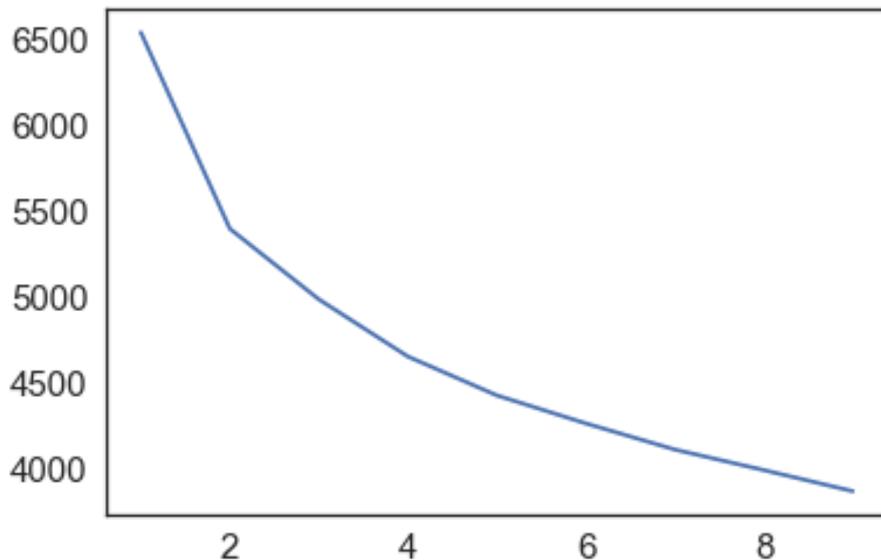


Fig. 10. Elbow criterion plot for clusters based on disease prevalences.

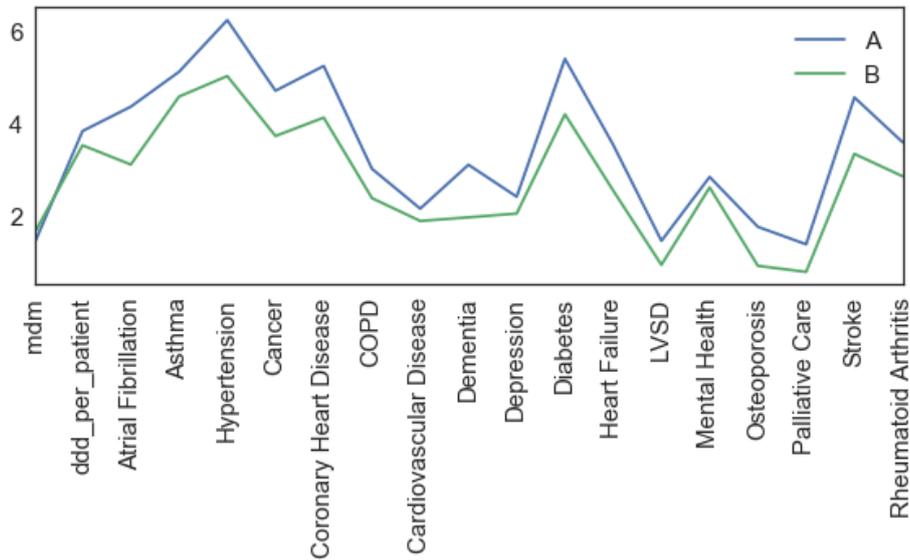


Fig. 11. Line chart illustrating the mean feature values for each cluster based on a two-cluster analysis of disease prevalences.

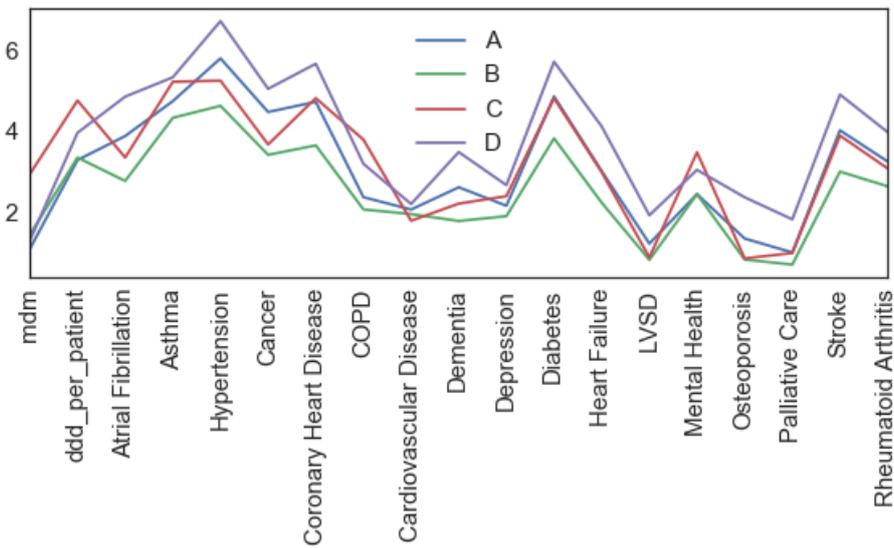


Fig. 12. Line chart illustrating the mean feature values for each cluster based on a four-cluster analysis of disease prevalences.

4 Discussion

4.1 Summary of the Main Findings

The issue of relatively greater incidence of mental health issues in Northern Ireland compared to other parts of the UK remains an important public health challenge in the region. Numerous studies have pointed to a range of factors that might be impacting the prevalence of such illnesses. Some studies have noted the higher levels of antidepressant prescribing compared to other countries and have suggested that there is a link between such prescribing rates and underlying mental health issues. These links are not currently well proven or understood.

This study explored correlations between three main variables - economic deprivation, depression prevalence and antidepressant prescribing - based on open GP prescribing data. The results showed that while there was a strong correlation between economic deprivation and antidepressant prescribing, the correlations between deprivation and prevalence and between prevalence and prescribing were weak. We therefore propose that the hypothesis that depression is a mediating factor between deprivation and prescribing is not supported by the available data. Moreover, the lack of a clear link between depression prevalence and prescribing rates suggest that the clinical basis for increased antidepressant prescribing requires further investigation.

Correlations were also explored between deprivation, antidepressant prescribing and the prescribing of other drug groups according to BNF Chapter. It can be seen from the resulting chart (Fig. 5) that deprivation was strongly correlated with both central nervous system (including antidepressants) and endocrine system (including insulin) prescribing. Higher levels of antidepressants were correlated with higher prescribing levels across multiple drug categories, including cardiovascular system, infections, endocrine system and musculoskeletal and joint diseases.

The study also examined the correlation between deprivation, antidepressant prescribing and other disease prevalence. This highlighted the link between deprivation and prevalence of mental health disorders, as well as chronic obstructive pulmonary disease (COPD). It also showed that higher levels of antidepressant prescribing tended to occur alongside higher levels of COPD, asthma, diabetes and mental health disorders. While the link between antidepressants and mental health seems unsurprising, the correlations with COPD, asthma and diabetes may suggest opportunities for further research.

One of the things that emerges from the disease prevalence correlation chart (Fig. 6), is that diabetes seems to be connected to a range of other illnesses, including atrial fibrillation, hypertension, coronary heart disease and stroke. In terms of antidepressant prescribing, the link between diabetes and depression has been well

documented in the literature, with some research suggesting that the relationship may be bi-directional [10]. This has policy implications in terms of both how diabetes might impact depression in the general population, and how the effect of depression on self-care, health care usage and expenditure might be addressed [11, 13].

4.2 Challenges and limitations

The major challenges for the study were in identifying, integrating and analysing open datasets from diverse sources. Data provided by the public sector in Northern Ireland differs in a number of ways from that provided in other parts of the UK, or by other nations, which makes valid comparisons more difficult. Specific challenges included the use of BNF as opposed to ATC identifiers in prescribing datasets, the lack of a formally supported method for linking these two classification systems, differences between how prevalence data is collected from other parts of the UK, the absence of a standardised method for assigning deprivation measures to GP practices, and difficulties in linking GP practice data to geographic boundaries.

The use of British National Foundry (BNF) drug identifiers in the UK caused particular challenges due to the lack of an official and comprehensive way of linking these codes to the World Health Organisation's Anatomical Therapeutic Chemical (ATC) classification system. This is an important consideration as it is from the ATC system that the defined daily doses (DDD) are taken. These DDD values are in turn used to compare relative prescription levels across different drug categories. While a NHS-derived table of correspondences between the two systems is available, it is not yet entirely comprehensive, meaning that a small percentage of drugs are not fully accounted for in this study.

Finally, this study has only explored data within Northern Ireland, although it reflects earlier studies that increases in antidepressant prescribing cannot be explained by increases in the incidence of prevalence of depression [19]. Further research is required before the findings may be extended to other drug categories, types of illness, or other regions. Moreover, since it relied purely on open data, factors such as GP gender, which previous research has indicated might be significant [26], have not been examined in this study. Generally speaking, there are a range of variables which cannot be fully explored in an open data context due to privacy or confidentiality considerations.

4.3 Antidepressant Prescribing and Depression Prevalence

There has been some exploration within the literature of the validity of using pharmaceutical data as a proxy for measuring clinical conditions in a given population [8, 9, 16]. The arguments in favour of such an approach include that fact that disease prevalence data is often difficult to capture accurately [32], may be inconsistently reported [8], and in some cases may not be available at all [9]. In such

situations, prescribing data, if available, may offer an attractive alternative for studying public health issues. Indeed, some studies have suggested a strong correlation between some antidepressants and clinical diagnoses. Gardarsdottir et al. [12] for example observed that 73% of patients on an SSRI had a diagnosis of depression or anxiety, while Henriksson et al. [14] found that 82% using SSRIs had been diagnosed with depression.

While the case for using pharmaceutical data as a tool for public health surveillance may seem fairly strong, there are a number of important caveats that should be taken into consideration. Firstly, as Henriksson et al. [12] show, not all drugs are equally strongly correlated with specific illnesses. Many drugs have more than one use, and many antidepressants are used for a wide range of disorders including fibromyalgia, chronic pain, eating disorders, insomnia and migraine [12]. An analysis by Mercier et al. [23] of antidepressant prescribing among French GPs found that 20% of prescriptions are potentially unrelated to any psychiatric condition. The perceived “safety” of SSRIs may exacerbate the tendency towards prescribing for a wider range of conditions [19].

Our results show that the link between depression prevalence and antidepressant prescribing is weak. On the other hand, there does appear to be a strong correlation between antidepressant prescribing and prevalence of mental health disorders. This suggests that there may be an issue with how patients who receive antidepressants are classified - in particular whether they are classified as suffering from either depression or from mental health issues. In any case, it is clear that caution must be exercised when using prescribing data as an indicator of depression prevalence in the wider population.

4.4 Factors Driving Antidepressant Prescribing

If, as the data suggests, dramatic increases in the levels of antidepressant prescribing are not being motivated by corresponding increases in depression prevalence, the question arises as to what other factors might be behind these trends. A number of studies have attempted to answer this question by interviewing GPs in order to get their perspective on the problem. McDonald et al. [19] propose that few GPs believe that depression levels have actually risen, and that there were some concerns about whether current prescribing levels were appropriate. When asked to identify factors that were driving the increase in prescribing rates, suggestions included: successful awareness-raising campaigns on depression, the perceived safety of selective serotonin reuptake inhibitors (SSRIs), and a willingness among patients to ask for help. Some clinicians believed that normal human “unhappiness” was being inappropriately interpreted as a medical condition.

A key finding of our study is that antidepressant prescribing is not strongly correlated with depression prevalence, although it is strongly correlated with both

mental health issues and diabetes prevalence. This suggests that antidepressants are being prescribed to patients diagnosed with other mental health issues. It also implies that antidepressant usage is strongly linked to diabetes, which reflects other studies on mental health comorbidities in diabetic patients [10]. On the other hand, the lack of correlation between antidepressants and depression prevalence once raises questions about how patients receiving antidepressants are classified in terms of depression and other disease categories. Ultimately, an improved understanding of the relationships between diagnosis and prescribing may lead to a better explanation of the factors driving antidepressant usage, and of the wider healthcare implications of this trend.

4.5 Clustering Practices Based on Prescribing Patterns and Disease Prevalence

The clustering of GP practices based on prescribing patterns according to BNF drug classifications shows two quite distinct groups based on the overall level of prescribing. The group of practices in more deprived areas tended to show higher levels of prescribing in all categories except “Malignant Disease”. A three-cluster analysis of the same data showed a similar pattern, but effectively partitioned the low-deprivation practices into two clusters – a higher prescribing group and a lower prescribing group. None of these clusters were characterised by strong differences in prescribing for a particular type of drug, but rather by higher or lower prescribing levels across all categories.

The second clustering analysis, which used disease prevalence to classify GP practices, similarly grouped practices into higher and lower levels of prevalence. Interestingly, in the two cluster analysis, there was no clear distinction between the clusters in terms of deprivation scores. When the prevalence data was analysed in terms of three clusters, the practices were once again split along deprivation lines. The GP practices with higher deprivation scores were divided into those with high, medium or low prevalence measures across all disease types.

5 Conclusion

Mental health has been clearly identified as a priority area for public investment, both within Northern Ireland and across the UK as a whole. Policymakers have also recognised the need to address effectiveness and efficiency in how these services are delivered. The impact of big data on other sectors suggests that data science can help to meet some of these needs, although this is an area that is currently under-explored within the literature. In this study we investigated data analytics might be applied to open prescribing data, disease prevalence data and economic deprivation data in order to shine a light on public policy and service delivery issues.

Our findings highlight both the limitations and opportunities of such an approach. Some widespread assumptions about the value of prescribing data as a proxy for

disease prevalence have been called into question. On the other hand, our analysis suggests that interesting correlations between disease prevalences and GP prescribing do exist, and may have useful implications for future policymaking. Moreover, the clustering of GP practices based on depression prevalence data implies that different populations respond to economic deprivation factors in different ways - an insight that merits further investigation by both researchers and policymakers.

The seeming weakness of the correlation between antidepressant prescribing and depression prevalence calls into question the medical basis for increasing antidepressant usage. Within Northern Ireland, where levels of antidepressant prescribing greatly exceed other countries, this is a particularly urgent concern. In order to address this problem, further data analysis might be used to identify anomalous prescribing patterns and thereby enable targeted interventions by the Department of Health. Such analysis could also help to identify hidden environmental or clinical factors, and to provide GPs with information to support clinical decision-making.

Future research in this area might examine why depression prevalence and antidepressant prescribing are so weakly correlated and what the potential implications are for public policy. The existence of distinct depression prevalence clusters also merits a more detailed investigation. Interviews with GPs and other experts might be helpful in order to gain information that is not apparent from a purely quantitative analysis. Finally, while this study has focused on the use of open datasets, combining these with non-open data about individual GPs and patients might allow more granular exploration and provide further policy insights.

References

- [1] Andreu-Perez J, Poon CCY, Merrifield RD, Wong STC, Yang GZ (2015) Big Data for Health. *IEEE Journal of Biomedical and Health Informatics* 19:1193–1208. doi: 10.1109/JBHI.2015.2450362
- [2] Akoglu L, Tong H, Koutra D (2014) Graph-based anomaly detection and description: A survey.
- [3] Bamford Review of Mental Health and Learning Disability (2006). *A Vision of a Comprehensive Child and Adolescent Mental Health Service (CAMHs)*. Belfast: DHSSPS.
- [4] Betts J, Thompson J (2017) *Mental Health in Northern Ireland: Overview, Strategies, Policies, Care Pathways, CAMHS and Barriers to Accessing Services*.
- [5] Birkhead GS, Klompas M, Shah NR (2015) Uses of Electronic Health Records for Public Health Surveillance to Advance Public Health. *Annual Review of Public Health* 36:345–359. doi: 10.1146/annurev-publhealth-031914-122747

- [6] Bunting, BP, Murphy, SD, O'Neill, SM, Ferry, FR (2011) Lifetime prevalence of mental health disorders and delay in treatment following initial onset: evidence from the Northern Ireland Study of Health and Stress. *Psychol Med* 42 (8), 1727–1739. doi: 10.1017/S0033291711002510.
- [7] Chandola V, Banerjee A, Kumar V (2009) Anomaly detection: A survey. *ACM Computing Surveys* 41:1–58. doi: 10.1145/1541880.1541882
- [8] Chini F, Pezzotti P, Orzella L, Borgia P, Guasticchi G (2011) Can we use the pharmacy data to estimate the prevalence of chronic conditions? a comparison of multiple data sources. *BMC Public Health* 11:688.
- [9] Coates, A. Ng, A.Y., (2012) Learning feature representations with k-means. In *Neural networks: Tricks of the trade* (pp. 561-580). Springer, Berlin, Heidelberg.
- [10] Detail Data (2017) GP Practices. <http://data.nicva.org/dataset/gp-practices>
- [11] Ducat, L., Philipson, L. H., & Anderson, B. J. (2014). The mental health comorbidities of diabetes. *Jama*, 312(7), 691-692.
- [12] Egede LE, Zheng D, Simpson K. Comorbid depression is associated with increased health care use and expenditures in individuals with diabetes. *Diabetes Care*. 2002;25(3):464–470. [PubMed]
- [13] Gardarsdottir H, Heerdink R, van Dijk L & Egberts A. (2007) Indications for antidepressant drug prescribing in general practice in the Netherlands. *Journal of Affective Disorders*.
- [14] Gonzalez JS. Depression. In: Peters A, Laffel L, editors. *Type 1 Diabetes Sourcebook*. Alexandria, VA: American Diabetes Association; 2013. pp. 169–179.
- [15] Health NI (2017) 2016/17 raw disease prevalence trend data for Northern Ireland. <https://www.health-ni.gov.uk/publications/201617-raw-disease-prevalence-trend-data-northern-ireland>
- [16] Henriksson S, Boëthius G, Hakansson J, & Isacson G (2016) Indications for and outcome of antidepressant medication in a general population: a prescription database and medical record study in Jämtland county, Sweden, 1995. *Acta Psychiatr Scand*
- [17] House of Commons (2016) The Big Data Dilemma. <https://www.publications.parliament.uk/pa/cm201516/cmselect/cmsctech/468/46802.htm>
- [18] Huber CA, Szucs TD, Rapold R, Reich O (2013) Identifying patients with chronic conditions using pharmacy data in Switzerland: an updated mapping approach to the classification of medications. *BMC public health* 13:1030.
- [19] iPyLeaflet (2017) An IPython Widget for Leaflet Maps. <https://github.com/ellisonbg/ipyleaflet>
- [20] Jupyter (2017) Project Jupyter. <http://jupyter.org/>
- [21] Krumholz HM (2014) Big Data And New Knowledge In Medicine: The Thinking, Training, And Tools Needed For A Learning Health System. *Health Aff (Millwood)* 33:1163–1170. doi: 10.1377/hlthaff.2014.0053
- [22] Loewenstein G, Asch DA, Friedman JY, Melichar LA, Volpp KG (2012) Can behavioural economics make us healthier?

- [23] Macdonald S, Morrison J, Maxwell M, Munoz-Arroyo R, Power A, Smith M, Sutton M, Wilson P (2009) "A coal face option": GPs' perspectives on the rise in antidepressant prescribing. *Br J Gen Pract* 59:e299–e307.
- [24] Maguire A, Hughes C, Cardwell C, O'Reilly D (2013) Psychotropic Medications and the Transition Into Care: A National Data Linkage Study. *J Am Geriatr Soc* 61:215–221. doi: 10.1111/jgs.12101
- [25] Maguire A, McCann M, Moritarty J, O'Reilly D (2014) The Grief Study: using administrative data to understand the mental health impact of bereavement. *Eur J Public Health*. doi: 10.1093/eurpub/cku165.058
- [26] McClure, J (2014) The Script Report. <http://script-report.thedetail.tv/>
- [27] Meeker D, Knight TK, Friedberg MW, Linder JA, Goldstein NJ, Fox CR, Rothfeld A, Diaz G, Doctor JN (2014) Nudging guideline-concordant antibiotic prescribing: a randomized clinical trial. *JAMA internal medicine* 174:425–431.
- [28] Mental Health Foundation (2016) Mental Health in Northern Ireland: Fundamental Facts 2016. <https://www.mentalhealth.org.uk/sites/default/files/FF16%20Northern%20ireland.pdf>
- [29] Mitchell, AJ (2016) Adherence behaviour with psychotropic medication is a form of self-medication. *Med Hypotheses*.
- [30] Morrison J, Anderson M-J, Sutton M, Munoz-Arroyo R, McDonald S, Maxwell M, Power A, Smith M, Wilson P (2009) Factors influencing variation in prescribing of antidepressants by general practices in Scotland. *Br J Gen Pract* 59:e25–e31.
- [31] MySociety (2017) MapIt. <https://mapit.mysociety.org/>
- [32] NHS England (2016) The Five Year Forward View for Mental Health. <https://www.england.nhs.uk/wp-content/uploads/2016/02/Mental-Health-Taskforce-FYFV-final.pdf>
- [33] NHS Digital (2017) NHSBSA dm+d and NHSBSA dm+d supplementary. <https://isd.digital.nhs.uk/>
- [34] NISRA (2017) Northern Ireland Multiple Deprivation Measure 2010: SOA Results. <https://www.nisra.gov.uk/publications/northern-ireland-multiple-deprivation-measure-2010-soa-results>
- [35] Numpy (2017) <http://www.numpy.org/>
- [36] O'Neill, C, McGregor, P, & Merkur, S (2012) United Kingdom (Northern Ireland). Health System Review. *Health Systems in Transition*, 14 (10).
- [37] Open Data NI (2017) GP Prescribing Data. <https://www.opendatani.gov.uk/dataset/gp-prescribing-data>
- [38] Pandas (2017). Python Data Analysis Library. <https://pandas.pydata.org/>
- [39] Pampallona S, Bollini P, Tibaldi G, et al. (2002) Patient adherence in the treatment of depression. *Br J Psychiatry*.
- [40] Rahman F, Slepian MJ (2016) Application of big-data in healthcare analytics #8212; Prospects and challenges. In: 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI). pp 13–16
- [41] Scikit-learn (2017) Machine Learning in Python. <http://scikit-learn.org/stable/>

- [42] Seaborn (2017) seaborn: statistical data visualization
<https://seaborn.pydata.org/>
- [43] Velasco E, Agheneza T, Denecke K, Kirchner G, Eckmanns T (2014) Social Media and Internet-Based Data in Global Systems for Public Health Surveillance: A Systematic Review. *Milbank Q* 92:7–33. doi: 10.1111/1468-0009.12038
- [44] Vercambre M & Gilbert F (2012) Respondents in an epidemiologic survey had fewer psychotropic prescriptions than nonrespondents: an insight into health-related selection bias using routine health insurance data. *Journal of Clinical Epidemiology*. doi:10.1016/j.jclinepi.2012.05.002
- [45] WHO Collaborating Centre for Drug Statistics Methodology (2017) Structure and Principles. https://www.whocc.no/atc/structure_and_principles/
- [46] Yang M, Kiang M, Shang W (2015) Filtering big data from social media – Building an early warning system for adverse drug reactions. *Journal of Biomedical Informatics* 54:230–240. doi: 10.1016/j.jbi.2015.01.011