Big Data Research ••• (••••) •••–•••

Contents lists available at ScienceDirect

Big Data Research

www.elsevier.com/locate/bdr

Insights into Antidepressant Prescribing Using Open Health Data ☆

Brian Cleland [a], Jonathan Wallace [a], Raymond Bond [a], Michaela Black [b], Maurice Mulvenna [a], Deborah Rankin [b], Austin Tanney [c]

[a] Ulster University, Jordanstown Campus, Northern Ireland, UK
[b] Ulster University, Magee Campus, Northern Ireland, UK
[c] Analytics Engines, 1 Chlorine Gardens, Belfast, Northern Ireland, UK

ARTICLE INFO

Article history:
Received 15 October 2017
Received in revised form 6 February 2018
Accepted 25 February 2018
Available online xxxx

Keywords:
Health policy
Data analytics
Big data
Prescribing
Prevalence
Machine learning
Deprivation

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.

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 as 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,

☆ This article belongs to Big Data for Healthcare.

E-mail addresses: b.cleland@ulster.ac.uk (B. Cleland), jg.wallace@ulster.ac.uk (J. Wallace), rb.bond@ulster.ac.uk (R. Bond), mm.black@ulster.ac.uk (M. Black), md.mulvenna@ulster.ac.uk (M. Mulvenna), a.tanney@analyticsengines.com (A. Tanney).

https://doi.org/10.1016/j.bdr.2018.02.002
2214-5796/

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.

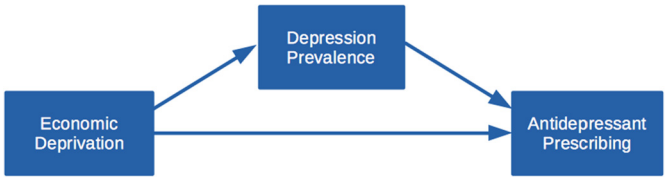

Fig. 1. Hypothesis: mental health is a mediating factor linking economic deprivation and antidepressant prescribing (arrows indicate mediating relationships).

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].

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

3.2.1. 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 Foundry (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].

Please cite this article in press as: B. Cleland et al., Insights into Antidepressant Prescribing Using Open Health Data, Big Data Res. (2018), https://doi.org/10.1016/j.bdr.2018.02.002