Exploring Prescribing – do Lofepramine, Moclobemide, Sertraline and Mirtazapine have geographies?

Martin Charlton[*1], Chris Brunsdon[†1], John Cromby[‡2] and Lex Comber[§3]

[1]National Centre for Geocomputation, Maynooth University, Maynooth, Co Kildare, Ireland
[2]School of Business, University of Leicester, Leicester LE1 7RH
[3]School of Geography, University of Leeds, Leeds LS2 9JT

December 3, 2017

Summary

The NHS have made available their GP Practice Presentation-level Data since April 2005. We describe the data and its organisation. Initial exploration of the data reveals some of the complexity surrounding its analysis – in particular the sheer size of the data, and the necessity of understanding the British National Formulary coding system. Taking antidepressant drugs as an example we undertake some initial visualisation of the data, including time series decomposition, and regional comparisons using boxplots. Mapping specific subsets of the data suggests possible covariates; however, we stress the importance of domain understanding in try to unravel the complex interactions within these data.

KEYWORDS: NHS, Prescribing Data, Data Handling, Data Exploration.

1. Introduction

The NHS have made available their GP Practice Presentation-level Data since April 2005[**] The data are described by Rowlingson et al (2013). Downloadable files are available in CSV form, one per month. Each month's PDPI file contains around 10 million records with the following fields:

Table 1 PDPI fields

Field	Description
`SHA`	Strategic Health Authority/Area Team
`PCT`	Primary Care Trust/Clinical Commissioning Group
`PRACTICE`	Practice Code
`BNF.CODE`	British National Formulary Code
`BNF.NAME`	Individual preparation name
`ITEMS`	Number of items dispensed for this presentation
`NIC`	Net ingredient cost (£)
`ACT.COST`	Actual cost (£)
`QUANTITY`	Quantity of drug dispensed
`PERIOD`	Month timestamp (yyyymm)

There are two additional files: (i) ADDR, which details the address and postcode of each practice, with

[*] martin.charlton@mu.ie
[†] christopher.brunsdon@mu.ie
[‡] john.cromby@leicester.ac.uk
[§] a.comber@leeds.ac.uk
[**] https://digital.nhs.uk/article/4214/Prescribing?p=1

its associated **PRACTICE** code and (ii) CHEM, which details the chemical name that is addressed by the first 9 characters of the BNF code. The files are available as CSV, and the 36 files representing each year require some 15Gb of disk space. These data present some interesting challenges, of management, exploration, analysis, visualisation, and modelling. A listing of the first 5 of the 10144514 records for June 2017 gives a flavour of the nature of the data in each PDPI file (Figure 1).

SHA	PCT	PRACTICE	BNF.CODE	BNF.NAME	ITEMS	NIC	ACT.COST	QUANTITY	PERIOD
Q44	RJN	Y05218	0501011P0AAAJAJ	Phenoxymethylpenicillin Pot_Tab 250mg	1	1.04	0.98	28	201706
Q44	RJN	Y05218	0502010B0AAABAB	Fluconazole_Cap 150mg	1	20.50	19.00	25	201706
Q44	RTV	Y04937	0401010Z0AAAAAA	Zopiclone_Tab 7.5mg	6	2.27	2.78	61	201706
Q44	RTV	Y04937	0401020K0AAAHAH	Diazepam_Tab 2mg	6	2.39	2.88	94	201706
Q44	RTV	Y04937	0402010ABAAABAB	Quetiapine_Tab 25mg	7	203.18	188.95	301	201706

Figure 1 First lines of the June 2017 PDPI file

The BNF code provides a key to the prescribing information. Taking the 4th record, a prescription for 94 2mg Diazepam tablets, we see that the BNF code starts with 040102. Chapter 4 in the BNF deals with drugs for the Central Nervous system; section 4.1 are Hypnotics and Anxiolytics; subsection 4.1.1 are Hypnotics. Diazepam, with the chemical code, 0401020K0, is a drug indicated for the short term relief of severe or disabling anxiety, but may also be used for dealing with spasms, muscle contractions, sedation, and convulsions. A glossary of which gives more detailed descriptions of the field names is also available[††].

2. Data exploration

With some 841000000 records (2010-2107) some re-organisation of the data will be required before any exploration or analysis might begin. The Open Prescribing website provides some tools for data exploration[‡‡]. Rowlingson et al (2013) describe more extensive approaches for examination spatial variation in prescribing. However, we may wish to link other data, or aggregate to different geographies. For this Excel is inadequate as a research tool, and we need something like R.

The sheer variety of drug names and variants can be bewildering. For this it is necessary to have an appreciation of the organisation of the BNF. A new version appears every 6 months. BNF 72 has 1449 pages, so is not for the faint-hearted. Summaries, such as those available at the Great Manchester Joint Formulary[§§] are useful guides.

Table 2 BNF Chapters

Chapter	Description
1	Gastro-intestinal system
2	Cardiovascular system
3	Respiratory system
4	Central nervous system
5	Infections
6	Endocrine system
7	Obstetrics, gynaecology and urinary-tract disorders
8	Malignant disease and immunosuppression
9	Nutrition and blood
10	Musculoskeletal and joint diseases
11	Eye
12	Ear, nose and oropharynx
13	Skin

Chapter 4 has 11 sections (see Table 3). Of these section 4.3 has 4 subsections which covered different classes of Antidepressant drugs, 4.3.1: Tricyclic and related antidepressants, 4.3.2 Monoamine-oxidase

[††] https://digital.nhs.uk/practice-level-prescribing-summary/glossary-of-terms
[‡‡] https://openprescribing.net
[§§] http://gmmmg.nhs.uk/html/formulary_bnf_chapters.html

inhibitors, 4.3.3 Selective serotonin re-uptake inhibitors and 4.3.4 Other antidepressant drugs.

One might begin by tabulating the prescriptions, perhaps for a single drug such as the SSRI Citalopram, or those in a subsection, such as the SSRIs. This raises the issue of what to count. Drugs appear in different forms and strengths, so counting the number of tablets prescribed would have an interaction with the strength of the particular drug. Citalopram can be prescribed in 10mg, 25mg and 40mg tablets for oral administration. Diazepam is available in a range of forms for oral or rectal administration, or by injection. The **`ACT.COST`** field definition has changed over the years, but is currently

[NIC less discount]+[payment for consumables]+[payment for containers]+[out of pocket expenses]

This does permit cross-sectional comparisons over a short time period and the creation of time series.

Table 3 BNF Chapter 4 Sections

Section	Description
4.1	Hypnotics and anxiolytics
4.2	Drugs used in psychoses and related disorders
4.3	Antidepressant drugs
4.4	Central nervous system stimulants and drugs used in ADHD
4.5	Drugs used in the treatment of obesity
4.6	Drugs used in nausea and vertigo
4.7	Analgesics
4.8	Antiepileptics
4.9	Drugs used in Parkinsonism and related disorders
4.10	Drugs used in substance dependence
4.11	Drugs for dementia

3. Data organisation

Some processing of the CSV files is required to speed the task of data reading and extraction. We made the decision to carry out the exploration and analysis using R. Reading the raw CSV files is a length process, so to speed the analysis we have stored each months PDPI data in a separate R image file, named to represent the year/month combination. The ADDR file images were stored in a R list structure. To read the entire dataset of 840 million records, and extract a subset for a section or subsection takes about 25 minutes on a relatively low powered laptop (Dell Vostro with 2.20GHz Intel Core i7 CPU). A single month can be loaded in under 20 seconds.

4. Experiment: antidepressant prescribing

We decided to explore prescribing patterns for antidepressants. Taking June 2017 as an example month the total spend on antidepressants was £17.7 million, with 5.7 million items prescribed. The most frequently prescribed drugs are shown in Table 4:

Table 4 Top-prescribed antidepressants

SubSection	Description
4.3.1	Amitriptyline [2, 1 is Trazodone, 3 is Trimipramine]
4.3.2	Moclobemide [2, 1 is Tranylcypramine]
4.3.3	Citalopram [2]
	Sertraline [1]
	Fluoxetine [3]
4.3.4	Mirtazapine [2, 1 is Venlafaxine]

The numbers in square brackets are the ranks based on cost. What this suggests is that the variations in individual drug costs, and prescribing mean that comparison at the drug level will be difficult due to various confounding effects, such as drug efficacy, which are difficult to determine from these data.

Furthermore, different CCGs have different advice on preferred drugs when there are alternatives: for example, Hull and East Riding Prescribing Committee recommend Lofepramine as first choice tricyclic, with Trazodone as second; Greater Manchester recommend both with Clomipramine as second. Manchester note that tricyclics are not recommended as first-line treatment choices – SSRIs are to be preferred. This suggests that we undertake the comparison at the subsection level.

Times series plots based on the 835010379 records from August 2010 to June 2017 are fascinating:

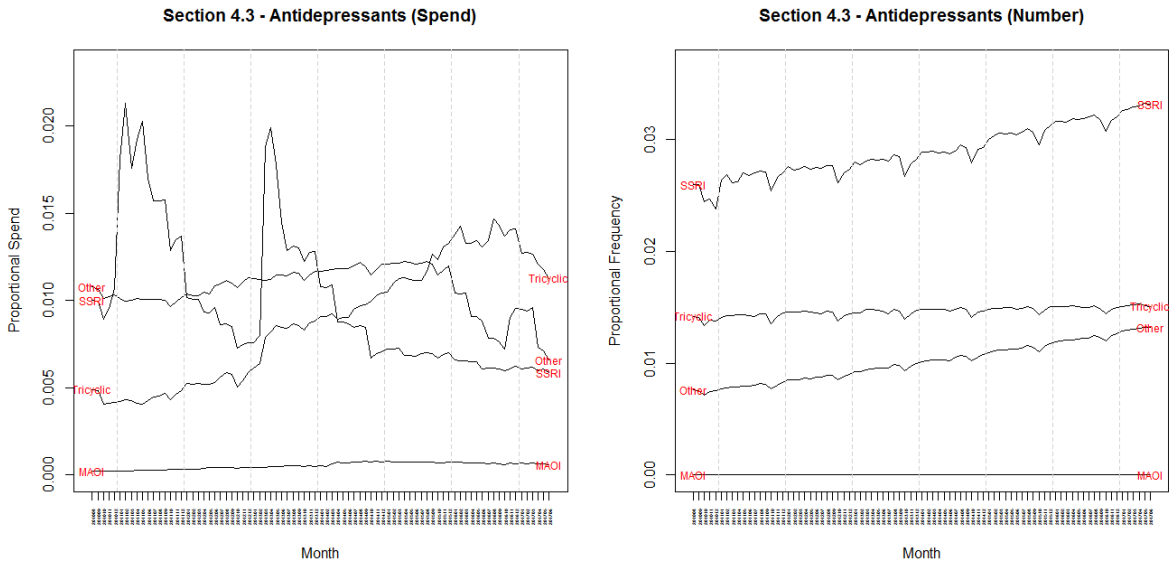

Figure 2 Time series – 2010 to 2017

The proportions based on the number of items show a steady increase, with some periodicity in the series. A decomposition of the SSRI series is shown in Figure 3:

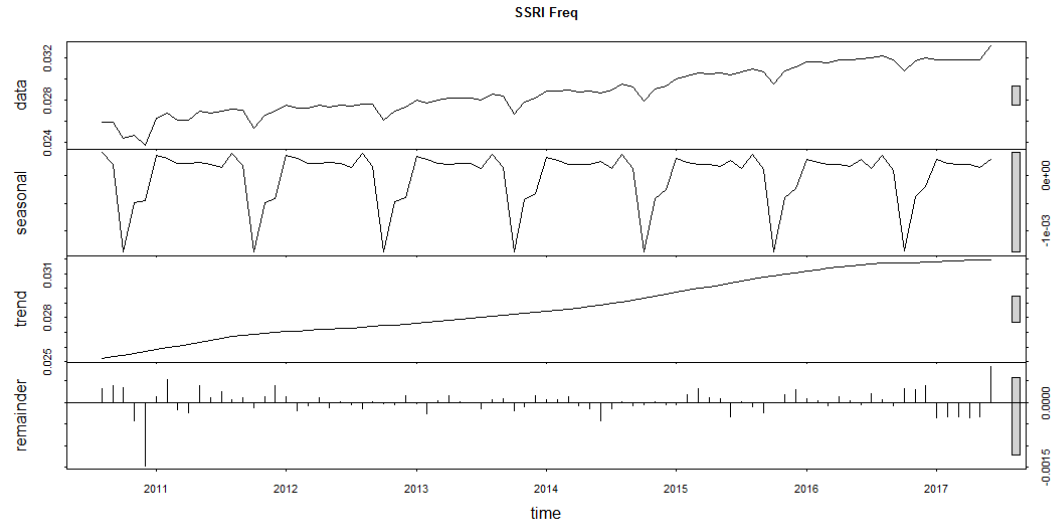

Figure 3 Time series – proportional frequency

The regular October-November-December decrease is intriguing. The variation in SSRI cost in Figure is decomposed in figure 4.

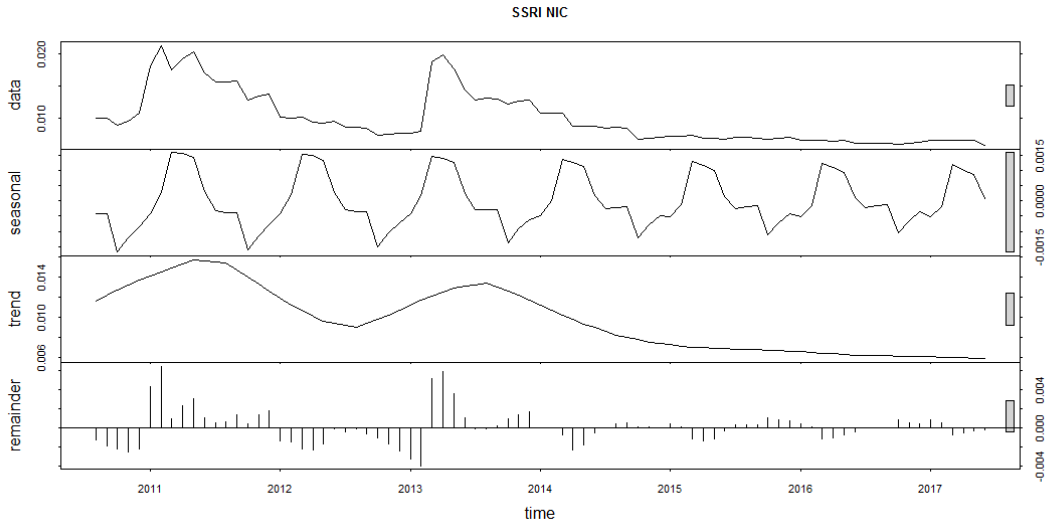

Figure 4 Time series – proportional NIC

These data would suggest some variability in costs for SSRI drugs over the period 2010-2014 followed by a stabilisation of the general prices. There's still a seasonal component which is tricky to explain.

Geography is as challenging The PCT/CCG codes can be related back to spatial data to provide maps at CCG level – one might impose county boundaries over the pattern as these provide familiar spatial clues. The postcodes can be used to add Region codes, as well as local authority and county codes, and also provide the possibilities of aggregation to other spatial units. This allows mapping at various spatial units, but also the use of other visualisation techniques such as boxplots to compare variations at practice level when cross classified by subsection and region. An example is shown in Figure 5 below:

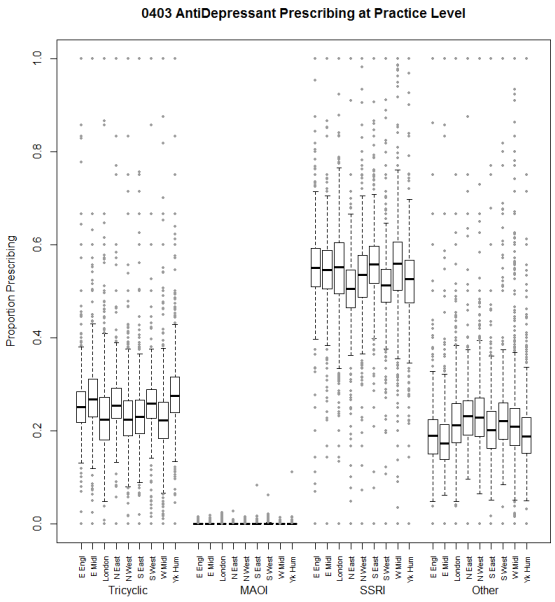

Figure 5 Proportional prescribing by Region and BNF SubSection

The preference for SSRIs over Tricyclic drugs is clear. MAOI are rarely prescribed – advice in local formularies is usually clear on this.

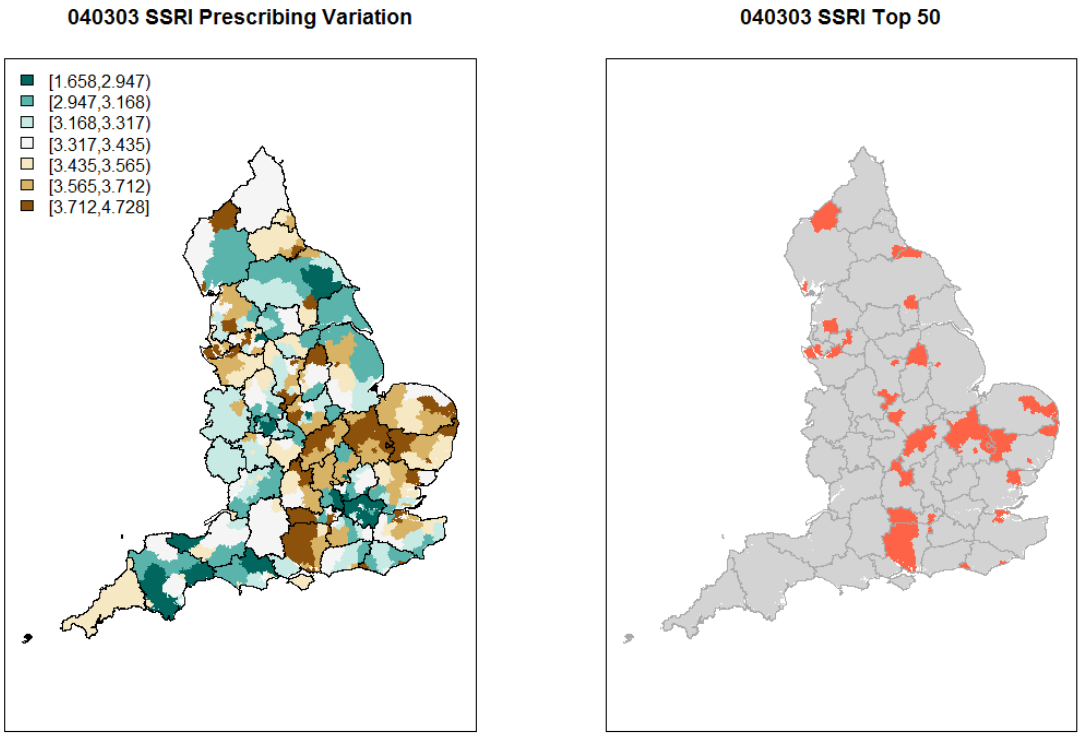

Figure 6 Spatial variation in SSRI prescribing

The spatial pattern is less obvious, although the top 10, with over 4%, are Norwich, Eastleigh, Great Yarmouth, Colchester, Brighton and Hove, Southampton, Kettering, Portsmouth, Lincoln, and Cambridge. At the other end of the prescribing scale, the lowest levels of SSRI prescribing are found in Harrow, Newham, Brent, Ealing, Redbridge, Slough, Hounslow, Enfield, Croydon, and Tower Hamlets. Potential covariates might include measures of affluence or proxies thereof. However, there are many potential confounders.

Modelling is one step further on. Practice level deprivation scores are available, and hint at some variability in prescribing at either end of the deprivation scale. The strongest correlations among the practices in least deprived areas are for female sex hormone, drugs for Parkinsonism, and drugs for arrhythmias; at the other end of the scale we find Vitamin B, antidiabetics, antacids, analgesics and topical antirheumatics. While it is tempting to conclude that affluent are dealing with the ravages of old age, and women are getting relief from the effects of the menopause, while the working class are dealing with poor diet and pain.

5. Conclusions

At the moment we are reporting work in progress. The analysis of spatial and temporal patterns anti-depressant (AD) prescribing has the potential to provide insights into local patterns of misery, as a counter to the frankly meaningless measures of and narratives around societal well-being. However, we are taking an explicitly informed approach to Data Science rather than simply engaging in a data fishing exercise. AD prescribing is very complex as these drugs are sometimes used for anxiety problems (reflecting in part the poor validity of the relevant diagnostic categories which overlap by more than 80% in some studies), but they also get prescribed for other purposes such as chronic pain and eating disorder diagnoses (SSRIs). There has been a steady long term increase AD prescribing and factors influencing AD prescribing incidence and prevalence are diverse, exhibit local variation, and seem to change over time. Critically AD prescribing is known to be influenced most vitally include underlying demography – socio-economic status is a strong predictor of distress, and thus of prescribing

– as well as a myriad of other factors including ethnic compositions, NICE guidelines, service frameworks, levels of knowledge amongst practitioners (GPs), treatment duration, and GP preferences.

For these reasons, the approach taken in this research is different to most Data Science which is fixated on machine or deep learning and data mining, under the implicit assumption that all questions can be answered by simply detecting patterns in data (Comber et al, 2016). Some researchers consider such approaches to be superior to theory-led ones (eg Mayer-Schonberger and Cukier, 2013), supplanting the need for theory (Brunsdon, 2016) but others highlight the critical importance of domain understanding when analysing large datasets (Reeves et al, 2014). In our view, such analyses can support both prediction and exploration but only with an understanding of the domain (Kitchen, 2014): otherwise, as recent research has highlighted (Harris et al, 2017), they may detect artifactual patterns or provide answers to arbitrary questions.

6. Acknowledgements

Martin Charlton gratefully acknowledges support from Science Foundation Ireland's Strategic Research Initiative (07/SRC/I1168). Chris Brunsdon and Martin Charlton also acknowledge support from Science Foundation Ireland's Investigator Award Programme (15/IA/3090).

7. Biography

Martin Charlton is Senior Lecturer in Geocomputation at Maynooth University. He is often to be found in the Roost. He has been to the Gladstone once.

Chris Brunsdon is Professor of Geocomputation at Maynooth University. He is also a regular in the Roost and visits the Gladstone when in Nottingham.

John Cromby is a Reader in Psychology. He has never been to the Roost but occasionally visits the Gladstone.

Lex Comber is Professor of Geography at the University of Leeds. He wrote the abstract of a paper for GIScience 2016 on a flight from Rhodes to London, aided by Martin and Chris. And he knows the Roost and the Gladstone.

References

Brunsdon C, 2016. Quantitative methods I: Reproducible research. *Progress in Human Geography*, 40:687

Comber, A, Brunsdon, C, Charlton, M and Harris, R, 2016, A moan, a discursion into the visualisation of very large spatial data and some rubrics for identifying big questions. In *International Conference on GIScience Short Paper Proceedings* (1)1, https://escholarship.org/uc/item/5sc537n4

Reeves A, McKee M, Basu S, and Stuckler D, 2014. The political economy of austerity and healthcare: Cross-national analysis of expenditure changes in 27 European nations 1995–2011. *Health Policy*, 115(1), 1-8

Kitchen R, 2014. Big data and human geography. *Dialogues in Human Geography* 3:262

Harris, R., O'Sullivan, D., Gahegan, M., Charlton, M., Comber, L., Longley, P., Brunsdon, C., Malleson, N., Heppenstall, A., Singleton, A. and Arribas-Bel, D., 2017. More bark than bytes? Reflections on 21+ years of geocomputation. *Environment and Planning B: Urban Analytics and City Science*, 44(4), 598-617

Mayer-Schonberger V & Cukier K, 2013 *Big data*. Houghton Mifflin Harcourt;

Rowlingson, B., Lawson, E., Taylor, B. and Diggle, P.J., 2013. Mapping English GP prescribing data: a tool for monitoring health-service inequalities. *BMJ open*, 3(1), p.e001363, see http://bmjopen.bmj.com/content/3/1/e001363.short