Investigation of Prescription Flows between General Practices and Boots Pharmacies via Spatial Interaction Modeling Techniques

By

Lenka Hašová

Dissertation

Submitted to

UNIVERSITY OF LIVERPOOL

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

16th September 2018

ABSTRACT

The purpose of this study was to apply spatial interaction models for estimating the prescription flows between General Practices (GP) and Boots pharmacies. Understanding the distribution of prescription flows is important for pharmacies to identify new markets and to evaluate the performance of current pharmacies. Analysing prescription flows is becoming increasingly important because prescription demands have increased with the associated increased life expectancy.

However, the application of spatial interaction modelling for prescription flows in the literature is limited. This is in part due to the nature of sensitive corporate information and the limited resources of the National Health Service (NHS) to analyse all prescription flows with all pharmacies. We apply several model specifications to estimate and analyse the distribution of GP-Boots pharmacy prescription flows in Merseyside, North-West England. We applied an unconstrained model, an unconstrained case enhanced by origin-destination specific demographic variables, a production-constrained specification, an attraction-constrained and doubly-constrained models using General Linear Models (GLM). We apply these models to several flow characteristics, to assess whether the nature of prescriptions has different flow generating processes.

The results of this study suggest that different flow generating processes are exhibited for different flow characteristics. The doubly constrained model had the best estimation of prescription flows greater than 100 prescriptions, accounting for a 78% model fit. However, the estimated and fitted distribution had the closest fit in the posterior predictive check for the production-constrained model with flows less than 800m, giving a model fit greater than 90%. These results support the literature, where the distribution and availability of health care services are based on human behaviour. This was related to the two stage decision process, where people firstly choice the pharmacy in the closest distance and secondly to fulfil our requirements such as other retail shopping. Furthermore, an analysis of the spatial distribution of prescription flows in Merseyside revealed a clear pattern. As expected, Boots prescription flows was highest for Liverpool, the core retail centre of Merseyside. Conversely, prescription flows was lowest for Southport. We also identified several Boots pharmacies where the performance of them could be improved by simply relocating the pharmacy or joining together two pharmacies.

By comparing various methods of prescription flow estimation, we have emphasized the importance of spatial effects in prescription flows. These include both spatial autocorrelation and heterogeneity of the movement of patients in space. By incorporating random effects of distance to the Liverpool city centre using a General Linear Mixed Model (GLMM), the interaction models gave a significantly higher model fit. The use of spatial interaction models at a smaller scale could be beneficial for the exploration of the local relationships, however, we propose a more robust theoretical research of the prescription flows nature. We have highlighted the need to incorporate the hierarchical structure of the people’s decision making process for selecting a pharmacy, to give better model estimates. Furthermore, further studies should also incorporate the two spatial effects because not accounting for these structure violates the independent and identically distributed assumption of regression. However, defining a spatial structure between origin and destinations is computationally challenging. Overall, this study has shown the value of applying spatial interaction models to prescription flows but also highlights the need for further research on this topic, to better understand the dynamics of prescription flows between GPs-Boots pharmacies.

DECLARATION

I hereby certify that this dissertation constitutes my own product, that where the language of others is set forth, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions or writings of another.

I declare that the dissertation describes original work that has not previously been presented for the award of any other degree of any institution.

ETHICS APPROVAL

This study made a formal ethics review check to the University of Liverpool Research Integrity and Ethics Committee, which ensures that the project does not involve human participants, human tissue or personal data. All the NHS data used, were aggregated for the yearly counts and does not carry any personal information. Other than that, there were used Open source data derived from Census that does not incorporate any sensitive information.

The Ethics review and approval with a reference 3990 can be accessed for request.

Signed,

Lenka Hasova

ACKNOWLEDGEMENTS

I would like to thank the CDRC, Boots and the University of Liverpool for giving me an opportunity to undertake this research. As a supervisor, Les Dolega has provided a vital guidance, support and advice that has facilitated a completion of this thesis. Daras Konstantinos also provided useful assistance concerning a variety of geographical problems and computational problems. Furthermore, the Graeme Cormac and David Rour provided an essential guidance and support for this thesis. Lastly, I would like to thank my family, all close friends and classmates for everyday life support and language advice during the whole period of my master's degree. My special thanks goes to Thomas Statham, who believed in me and helped me to stand up every time I fell over.

TABLE OF CONTENTS

1 Introduction

The demand placed on the National Health Service (NHS) and health care services is increasing as the life expectancy for both males and females living in England and Wales has increased. One study found that life expectancy in England and Wales will increase 79.5 and 83.3 years for both men and women to 85.7 and 87.6 years respectively, between 2012-2030 (ONS 2013; Bowers 2018). Combined with an increasing population size, the demand placed on General Practice (GP)s and pharmacies is increasing.

The life expectancy also varies between different regions, related to life style choices and level of deprivation. Moreover, within these regions, the availability of GP and pharmacies changes, to meet these different demographic demands. Nevertheless, GP's and pharmacies are available to more than 84% of the population, within walking distance of 1.6km (Todd, et al., 2015). This means that the level demand and supporting health care services are strongly influenced by the local population. For example, areas with a mainly an elderly population will experience a better supporting network of GP's and pharmacies within short distance, to support the demand of this demographic group. This suggests that a Geography of GP-pharmacy spatial interactions exists.

These contributing factors mean that the NHS and supporting pharmacies must better understand the dynamics of how demand changes over both time and space. More specifically, understanding the demand of prescription flows between GPs and pharmacies. Understanding these dynamics better are important from an operational point of view for both costs and logistical reasons. More importantly, its an obligation of the NHS to provide health care to every citizen to enable them to live a healthy life. This supports the need to apply spatial analysis for strategic planning of both GP and pharmacies, such as developing a new facility.

In England and Wales, the health services are commissioned by NHS, resulting in a unique health care structure within Europe. Whilst the NHS supports the interaction between GPs and pharmacies economically and ensuring safe and effective services to patients (NPA, 2009), with budgets an analysis of all GP-pharmacy interactions has been limited. Therefore, pharmacy chains must apply their own analysis to better understand the impacts of the structure and generation process of the prescription flows between GPs and pharmacies.

The literature on GPs and pharmacy have mainly focused on the availability of them and their effects on the population or the effects of populations on health care (Ikram, et al., 2015; Todd, et al., 2015). However, there is a limited number of studies who have assessed the interaction between GPs and pharmacies. Although widely acknowledged as a sound methodology, the application of Spatial Interaction Modeling methods to the context of prescription flow interactions between GPs and pharmacies is limited. This has been previously hampered by both computational power and the required software to apply the modelling approach with large datasets, such as prescription flows. Based on Gravity models, spatial interaction modelling can be applied to explore the spatial relationship of prescription flows between GPs and pharmacies. This can also be applied to predict the consequences of changes in their generation or attraction conditions (Wilson, 1967; Rodrigue, 2017). Overall, there is a need to better understand the prescription flow interactions between GPs and pharmacies.

We apply several model specifications to investigate the prescription flows between GPs and one of the leading pharmacies in England and Wales, Boots UK. We apply such an approach to the Merseyside County, to better understand the dynamics of prescription flows within this unique Geography. We apply these models to different flow characteristics to gain better insights into which

Boots UK pharmacies have the highest performance and to also identify areas where a new pharmacy could be located. It's hoped that by applying several different model specifications and different flow characteristics, that such an approach could be used by the wider literature.

The aim of this study was to apply spatial interaction modelling methods to prescription flows between General Practices (GPs) and Boots pharmacies in Merseyside, to evaluate the performance of Boots pharmacies. The objectives of this study are;

1. Investigate which Boots pharmacies have low and high performance, defined through number of prescription flows.
2. Evaluate which spatial interaction models in the literature best describes GP-Boots pharmacy prescription flows.
3. To assess which socio-economic factors are most related to GP-Boots pharmacy prescription flows.
4. To explore the distribution of high and low spatial interactions of GP-Boots prescription flows within Merseyside.

2 Literature review

2.1 Boots within the health care services

Boots United Kingdom (UK) creates an important share of the pharmacy market. However, Boots also provides health care services internationally with an expanding global presence. Boots UK has over 2500 stores, making up the largest share of the pharmaceutical market by provides both prescription and over the counter health services. Its market share is influenced by its attractiveness to its consumers, as a place for purchasing other health and beauty products alongside pharmaceutical ones. They also provide other healthcare services, including Boots opticians. Therefore, Boots faces competition from many sectors of the retail industry and their position within different markets requires a specific approach for both operational efficiency and strategic planning, including (UKEssays, 2013).

2.2 Healthcare services availability

With expected UK population increase and associated demand for prescription services, assessing the current and future demand for individual Boots UK stores is necessary. Moreover, the socio-economic background of an area influences the demand on health care services (Hiscock, et al., 2008) and their accessibility also varies by Geographic space. For example, one study applied simple linear regression to explore the relationship between pharmacies and socio-demographic factors Qato, et al. (2017). They found that pharmaceutical varied significantly by different socio-demographic characteristics, associated with different areas.

The NHS is a unique public health service structure within Europe that provides comprehensive healthcare, free for the use of people resident within the UK. It has a strong network of more than 40,000 GPs, which make in total over 1.1 million prescription items each year (Moody & Mindel, 2017). With the associated increased demand of prescription services is increasing over time, there is a need to analyse the interaction of prescription flows to pharmacies, where they are dispensed. Moreover, GP-pharmacy interactions must coverage for all areas of the UK and meet the current and future demands, including the aging demographic structure of the UK population.

Several studies have focused on assessing health care availability to the demographics of a population. Law, et al. (2011) used network analysis to measure Geographic accessibility to pharmacies in Ontario, Canada and examined the impact of pharmacy closures by Monte Carlo simulation. They study found that 63.6% of the population resides within 800m distance of pharmacies and that the closure of 30% of pharmacies, decreases this to just 56% of the population residing within 800m distance. Another study examined the accessibility of both pharmacies and GP's in England by Geographic Information System (GIS) and general linear models (GLM) (Todd, et al. 2015). The results show that 84.8% of the population lives within 20 min walk (1.6 km) of the GP premises and 89,2% population lives within 20 min walk of pharmacies. Furthermore Ikram, et al. (2015) used Kernel Density Estimation to measure the level of access to pharmacies that sell over-the-counter health items.
Other studies have assessed the accessibility of health care in relation to socio-economic characteristics and associated local Geography. One study found that the proportion of the population living within a 20-minute walk from a pharmacy increases from 90% for those residing in urban areas, 68% for those living on the urban- fringe and just 20% for those living in rural areas. Moreover, health care accessibility is also influenced by the level of deprivation. For example, the

proportion of people having access to GPs ranges from 66% in the least deprived areas to 76% in the most deprived areas; a result of positive care law (Todd, et al., 2015). Another study by Dijkstra (2013) that focuses on the geographic distribution of population prevalence to diabetes 2 medication, found a strong influence of social welfare, Social welfare varied by high- and low-income groups, in addition to the average house price and the proportion of elderly people. Overall, it can be said that the accessibility to health care services is considered dependent on the local Geography and its demographic characteristics, therefore a Geography of prescription flows exists.

2.3 Prescription flows as a spatial flow

The NHS defines patients flow as the number of people supported by the NHS. The NHS capacity is defined as a function of the capacity of the service to patient flows and when patient flows exceed the NHS capacity, this leads to overloaded waiting rooms. Ensuring a high quality of service and reducing the waiting time is made by incorporating more staff and improving the quality of each step from the moment the patient enters the facility (The Health Foundation, 2013). Conversely, within the Geography field, a spatial flow is defined as the interaction between georeferenced places, including aggregated migration flows, information exchanges or the flow of goods (Tao & Thill, 2016).

2.4 Spatial interaction models in the wider literature

The literature on flow dynamics and their spatial interactions has largely been expanded by the retail and service sector, which experienced extensive restructuring in the 20th century (Wilson, 1967). A spatial interaction model explains flows as a function of the origin and destination demand and the cost of the interaction between origin and destination. Different families of the spatial interaction models are recognized by the wider literature (Wilson, 1971; Fotheringham, 1983; Eyre, 1999), and can be generalized into four cases;

1. Unconstrained models,
2. Single constrained models – production or origin constrained case
3. Single constrained models – attraction or destination constrained case
4. Doubly constrained models, also called production-attraction or origin-destination constrained case

The production constrained model has been widely applied in the retail and shopping analysis to model the consumer's demand (Huff, 1963; 1964) or competing destinations (Fotheringham, 1983) and the attraction constrained model has been applied in modelling residential locations (Senior, 1973). One of the common techniques is an entropy maximisation (Batty & Mackie, 1972). This mathematical technique is a measure of entropy based on probability distribution of the stated Wilson (1970). Fotheringham (1983) claims that traditional spatial interaction models are often misspecified because of an occurrence of the spatial structure effect, caused by the distance decay effect, which assumes that distance is the only factor affecting the decay parameter, therefore his competing destinations model incorporate an entropy maximisation to achieve more correctly specified model. Lo (1991) claims that it is difficult to produce costumer preference patterns as each situation has a unique spatial structure. Despite the many extensions of this approach, Wilsons core ideas remains at the heart of such spatial interaction approaches.

This has led to the development of sophisticated methods and an important part of location strategy, within a competitive market (Eyre, 1999). Spatial interaction models emerged three major schools of analytical thought: one based on statistical equilibrium approach, second based on choice-theoretic approach and third based upon the neural network approach. However, all of those have a same view that the movement of entities creates the spatial interaction (LeSage & Fisher, 2008). Moreover, the availability of spatiotemporal data derived from sensor and GPS technologies has led to a resurgence in the application of spatial interaction methods. With open-source software platforms, including QGIS and Rstudio, this high volume of data can be enriched with other information sources, including social media data and analysed using spatial analysis.

The application of spatial interaction methods in the literature of patients flows between health services is limited, in part because of the challenge of combining the interaction between origin or GP and destination or pharmacies. There are two properties of spatial data that affect the choice of analysis. Firstly, spatial dependence, which refers to the first law of Geography (Tobler, 1970) where things closer in geographic space are more related to each other. Secondly, spatial heterogeneity accounts for the tendency of observations with each location to have a certain degree of uniqueness (Fischer, 2006). Moreover, if such methods have been applied in this field, they may not be publicised due to the nature of sensitive corporate information.

Typically, spatial interaction models are applied using a Poisson distribution. Yano (1993) claims that if interaction corresponds to Poisson distribution, then the interaction process is a result of a discrete probability process and the people's movement is constant and independent. The Poisson log-linear regression for the family of spatial interaction models was proposed by Flowerdew & Aitkin (1982), because of several limitations the flows have. Firstly, flows are often an integer counts of people so should be modelled as discrete entities, secondly, flows are not normally distributed and lastly, zero flows are problematic since the logarithm of zero are undefined (Oshan, 2016). The specification of the Poisson model generally assumes that the flow is drawn by Poisson distribution with mean, which is presumed to be logarithmically linked to the linear combination of variables. Since the mean of flows distribution is equal to the variance of the distribution, any factor or variable that affects one flow will also affect the other.

This also means that the assumption behind the homoscedasticity (error terms of independent variables is the same across) are not applicable to the Poisson model (Flowerdew & Aitkin, 1982). This issue of calibration of highly skewed flow data was also examined by Guy (1991), who compared different models explaining the people's movement into a retail centre and demonstrated that Poisson transformation of linear models is beneficial in the more flexible model specification, giving more consistent regression estimates and simpler model calibration and validation.

Congdon (2000) introduced a Bayesian approach to estimate and predict hospitalization flows. The study focused on emergency admissions to hospitals, to create a method for the better planning of patient care. The problem with such an approach is that the common Markov Chain Monte Carlo (MCMC) algorithms used within Bayesian inference are not computationally efficient and this approach requires the modeller to learn non-trivial programming languages. Another study by Jiang, et al. (2018) compared conventional methods including General Linear Models (GLM) against machine learning approaches for forecasting patientsarrival volumes to Accident and Emergency Departments. Although the machine learning approach had a higher forecast accuracy, incorporating space within such an approach is non-trivial. Overall, there is a need to apply different methods to explore which is more appropriate for current and future prescription flows between GPs and Boots pharmacies.

2.5 Local Spatial Interactions

We are also interested in how spatial interactions vary between GPs and Boots pharmacies. Linear regression analysis is the most commonly used statistical modelling technique and predictive tool. However, such an approach is not appropriate where spatial dependency exists because it violates the independent and identically distributed (iid) assumption (Getis, 1992; 2008). There are two main approaches in the wider literature for modelling local spatial interactions include spatial regimes and geographically weighted regression (GWR). The first method can be characterised as stratification, where each regime corresponds to a well-defined subset of locations (Anselin, 1992). Whereas the latter is a locally varying regression model. The first method was introduced by Brunsdon, et al. (1996), motivated by the idea of a nonparametric regression method used for exploration of non-stationarity of a regression relationship for spatial data.

As a relatively new method, GWR is sporadically represented in literature, and even fewer examples of an application on spatial flows can be found. One of the relevant examples can be found in the work of Nakaya (2001), who applied GWR on origin-constrained/doubly-specific spatial interaction models, to examine regional variations. Spatial dependency is modelled through a spatial weight's matrix. The research refers to competing for destination model (Fotheringham, et al., 2001) and verifies its assumption of spatial variation. Furthermore, Nakaya claims that it's reasonable to expect that spatial non-stationarity appears in every data that contains spatial information or spatial interaction. Concern about the presence of the spatial effects on the retail flows was already given by Fotheringham (1983), who stated that the stores may generate a greater volume of sales when they are clustered than when they are dispersed, and so the effects of agglomeration or competition are present.

Attention must also be given to a family of spatial lag/autoregressive and spatial error models. LeSage & Fisher (2008) suggested a sparse spatial weight matrix to capture either origin-based spatial dependencies or destination-based spatial dependencies for the application of migration flows. This paper emphasized the importance of spatial regression models for applied econometrics and linked maximum likelihood algorithms and Bayesian estimation as to origin-destination flows. A relevant example of Autoregressive models can be an application of Generalized Linear mixed models incorporating spatial random effect on the mapping of Dengue Fever in Thailand (Lekdee & Ingsrisawang, 2013). A similar approach that incorporates a spatial random effect can be seen in Multilevel Models (MLM). Dong, et al. (2015) used this approach to simultaneously measure horizontal and vertical spatial effects of residential land parcels among the districts in Beijing, China.

2.6 Decision making process and human behaviours of spatial interactions

The interaction between an origin and destination is a function of the decision-making process of individuals and defined by many factors, from which the main ones are distance and attractiveness. I individuals tend to interact less with destinations they view as less attractive and located further away than other destinations. However, many authors that aim to theoretically base and describe the spatial interaction of competing destinations by measuring human interaction, claims that this is considered as a two-staged decision-making process (Fotheringham, 1983; 1986; Eyre, 1999; Fotheringham, et al., 2001). In the first stage, the individual decides which group from the destinations he wants to travel to and in then decides about the specific destination.

The theoretical principle of the spatial interaction models is based on random utility maximization process, which assume that individuals are making a 'flat' information decision strategy. For example, from a set of destinations, it is assumed that the individual chooses one destination from all possibilities. However, this is a misleading concept and spatial Interaction models aim to describe human behaviour, rather than just one outcome. So, there is a need to examine the very principles of destination decision making (Fotheringham, et al., 2000). The result is that individuals do not consider all destinations equally once a destination is considered desirable, they may not consider all the other possible destination options, as the gravity model assumes (Eyre, 1999).

Fotheringham (1986) extended this simplistic assumption, by considered a hierarchical decision-making process. Here the spatial clusters of destinations are firstly selected and then only destinations within these selected clusters are evaluated, which provides a more plausible assumption than Gravity models and reduces any spatial bias (Fotheringham, et al., 2000). Liaw & Bartels (1982) developed a two-level logistic model of migration to respond to this problem. They stated that each decision process is determined by different variables, so they introduced a two-stage model consisting of a Departure model and destination model. Other examples could be found in the work of Birkin & Clarke (1995). Another way of more accurate representation of human behaviour is based on random utility theory that presume people maximise their utility by making choices between alternatives. Such models can be found in Anas (1982) or more recently in Mohammadian & Kanaroglou (2003).

3 Methodology

3.1 Study area

This study focused on the Merseyside metropolitan county in North West England. The area covers a population of approximately 1.38 million people and consists of 5 boroughs; Liverpool, Knowsley, St. Helens, Sefton and Wirral. This study area is unique because the Geography and demographic characteristics strongly shape the spatial interactions within Merseyside. Firstly, the Wirral peninsula is separated from the other boroughs by the River Mersey and Sefton is also shaped by the coast. Secondly, Merseyside is characterized by clear urban and rural communities. For example, there is a higher proportion of rural areas for Wirral, whereas Liverpool is characterized by a higher proportion of urban areas containing younger populations. Moreover, the city of Liverpool is the main retail catchment for this area.

Figure 1: Overview on study area

3.2 Data

All data used in this study are secondary data sources, which combines industry and open-source data.

3.2.1 Dependent variable

We used the number of Items on a flow between the GPs and Boots pharmacies as the dependent variable. Boots UK provided prescription flow data from General Practice (GP)s to pharmacies. This contained the number of items on each flow per month from March 2012 to February 2018. This is provided by the National Health Service as an Open Source (NHS, 2018).

3.2.2 Covariates

Boots also provided spatial information containing prescribers and dispenser's information, Boots pharmacies information and the latest population estimates. The origin mass is represented by a number of prescriptions prescribed by each GP in a year and the destination mass is defined as a number of prescriptions dispensed by Boots pharmacy in a year.

3.3 Data processing

In order to meet the aims and objectives of this study, data pre-processing was conducted on all flow information. Firstly all the flows from a GP's in Merseyside to pharmacies outside of Merseyside was removed. Whilst this may affect the performance of some GPs, particularly for those GPs on the edge of this study area, the aim of this study was to analyse interactions within Merseyside only. Secondly, all observations with zero item flows or missing spatial information were excluded because they are not meaningful. Lastly, the aim of this study was to analyse interaction flows for 2017 only so all other flows were excluded. Overall, the original dataset contained 1.75 million flows but after data cleaning, the analysis was conducted on just 6044 flows or 0.35% of the original data.

3.4 Exploratory Analysis

We used the programming language R for most of the data analysis and ArcGIS for data visualization.

3.4.1 Data description

Table 1 compares the UK origin-destination flows of yearly prescriptions for Boots and their competitors. Whilst the proportion of prescriptions for Boots stores represents approximately only 20% of total flows, when compared to its competitors, their flows make up twice as much as their nearest competitor (Table 2).

Table 1: Yearly proportions of flows

Year	*Total flow*	*% of the all flows*	*Boots flow*	*% of yearly flow*	*Competition flows*	*% of yearly flow*
2012	235556	13.5	53291	22.6	182265	77.4
2013	286280	16.4	61923	21.6	224357	78.4
2014	294092	16.8	62028	21.1	232064	78.9
2015	293965	16.8	60514	20.6	233451	79.4
2016	295148	16.9	59584	20.2	235564	79.8
2017	295528	16.9	58476	19.8	237052	80.2
2018	48797	2.8	9414	19.3	39383	80.7
Total flow	1749366	100	365230	20.9	1384136	79.1

Table 2: Top ten prescription dispensers

Pharmacy brand name	*Count*	*%*
BOOTS	365230	20.9
LLOYDS PHARMACY	173559	9.9
ROWLANDS PHARMACY	160660	9.2
APPLICANCE SUPPLIER	115132	6.6
ASDA PHARMACY	92085	5.3
TESCO INSTORE PHARMACY	63386	3.6
COHENS CHEMIST	49822	2.8
WELL	43018	2.5
SUPERDRUG PHARMACY	32908	1.9
SEDEM PHARMACY	28721	1.6
Total flow	1749366	100

Extracting the flows only for the Merseyside, the monthly flow dynamics can be explored. From 2013 to 2017, there is a clear pattern of higher and lower prescription flows for all pharmacies (Figure 2). This tends to follow a seasonal trend of disease outbreaks. Moreover, it’s clear that the number of prescriptions dispensed has increased over time. Conversely, the number of prescriptions dispensed from Boots stores has remained more stable over time (Figure 3).

Figure 2: Amount of dispensed Items by all pharmacies in Merseyside between 2013 and 2017

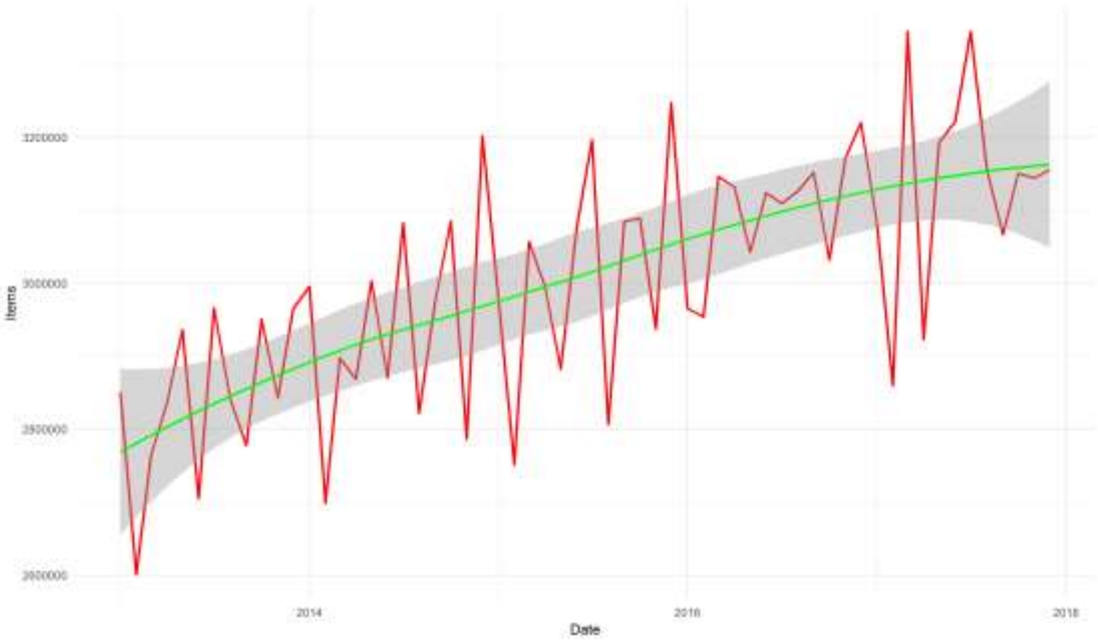
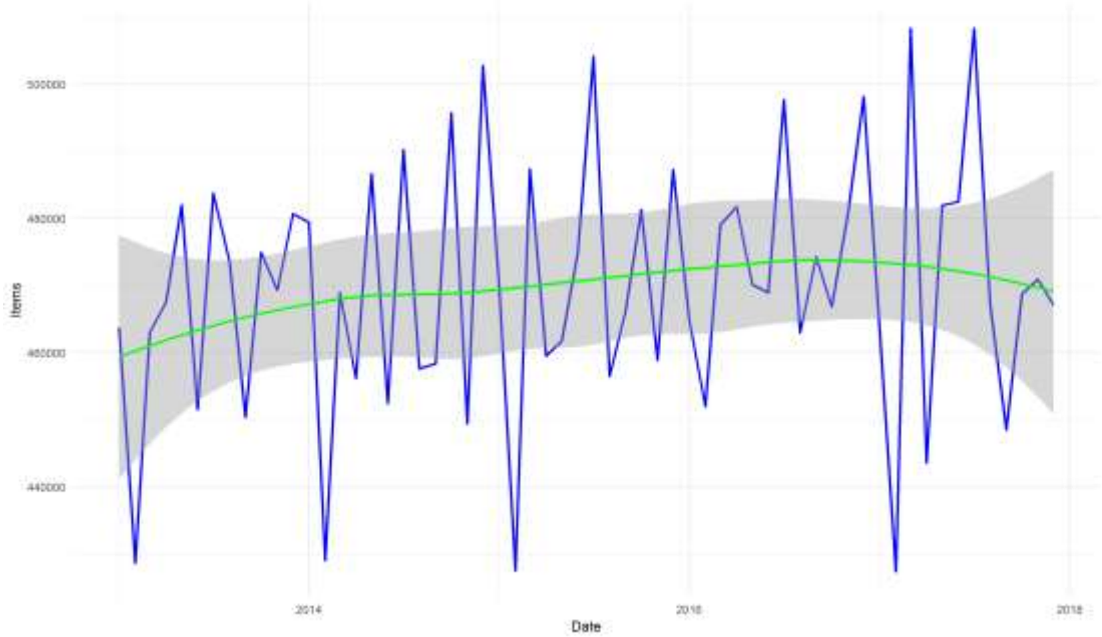


Figure 3: Amount of dispensed Items by Boots pharmacies in Merseyside between 2013 and 2017



An analysis of the monthly prescription flows for 2017 only further reveals this seasonal pattern of dispensed prescriptions (Figure 4). The highest number of dispensed items was recorded in March and July, whereas the least dispensed items was recorded in February and April. Furthermore, an analysis of just prescription items dispensed by Boots pharmacies shows the same pattern (Figure 5).

Figure 4: Amount of dispensed Items by all pharmacies in Merseyside during 2017

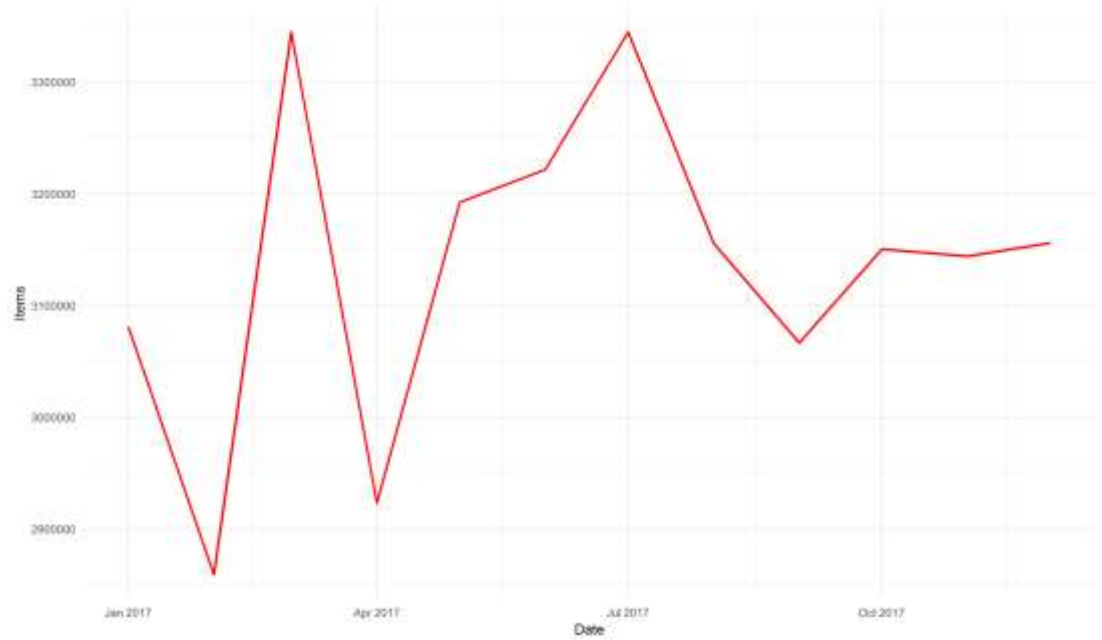
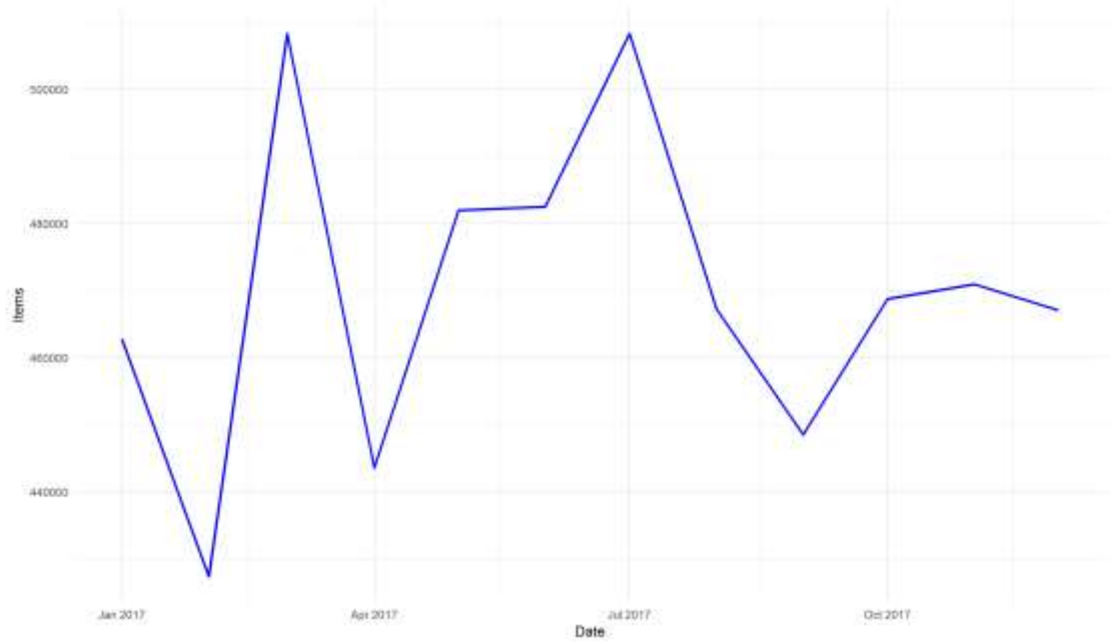


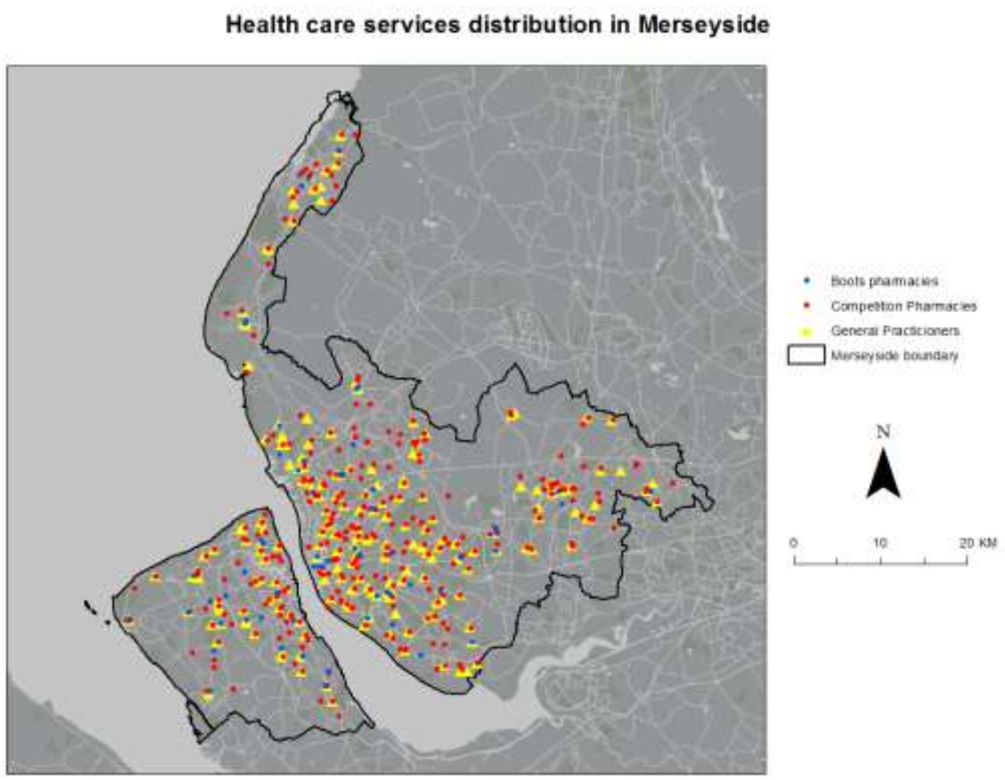
Figure 5: Amount of dispensed Items by Boots pharmacies in Merseyside during 2017



3.4.2 Spatial distribution

Although the data-processing removed some pharmacies in Merseyside, this study area covers 259 GP centres, 61 Boots pharmacies and 336 competing pharmacies. The variety of health services with different areas creates a unique structure dependent on the socio-demographic characteristics of the local community.

Figure 6: Distribution of health care services in Merseyside



To explore the characteristics of prescription flows, we created density maps to inspect densities of origin-destination points. The density maps show that there is significant clustering of GP centres (Figure 7) and Boots pharmacies (Figure 8). Both maps show that the highest density of GPs and Boots pharmacies in the study area distributed in Liverpool city centre. The lowest densities are distributed on the Wirral and East Merseyside.

Figure 7: Density of GP centres

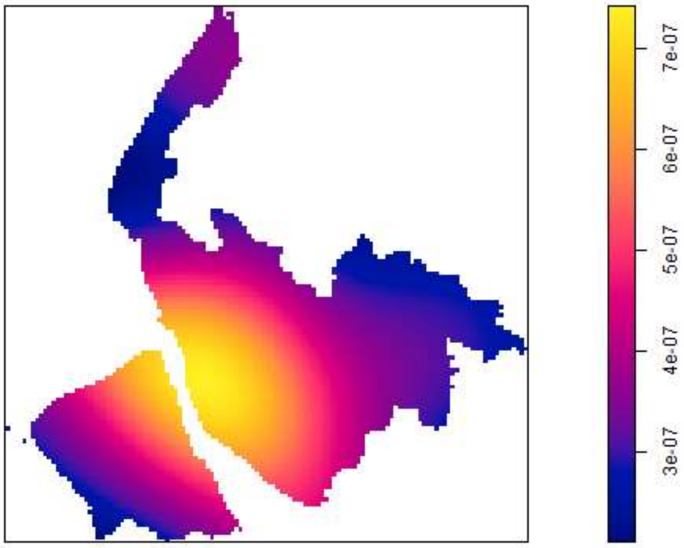


Figure 8: Density of Boots pharmacies

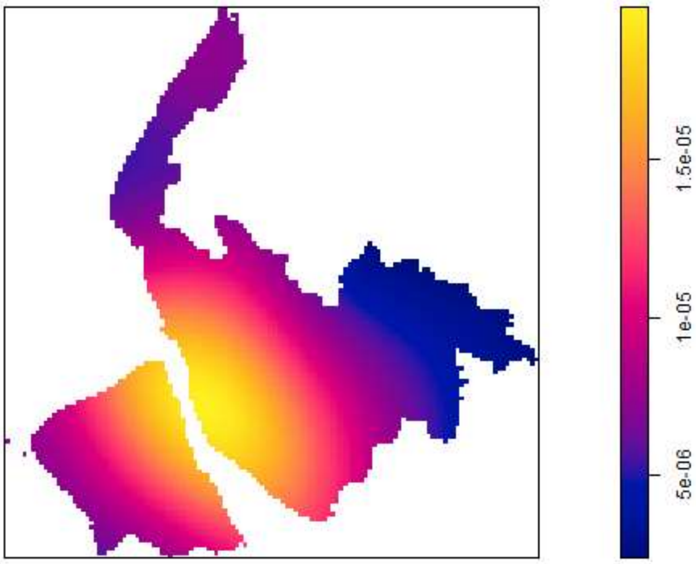


Figure 9: KDE map of the GP centres

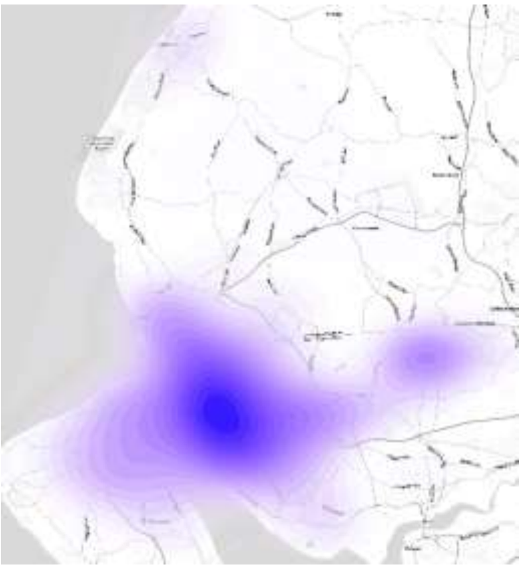


Figure 10: KDE map of Boots pharmacies

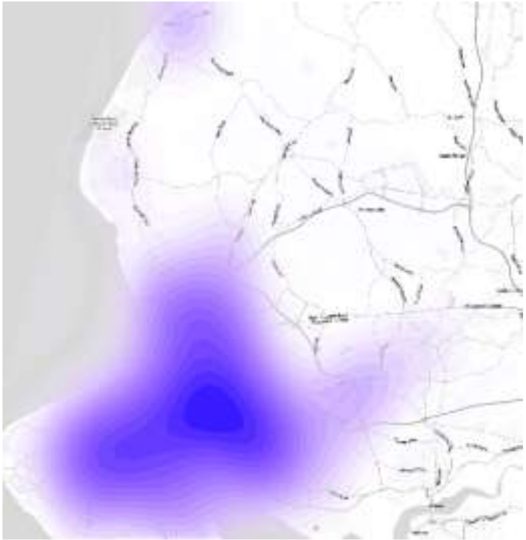
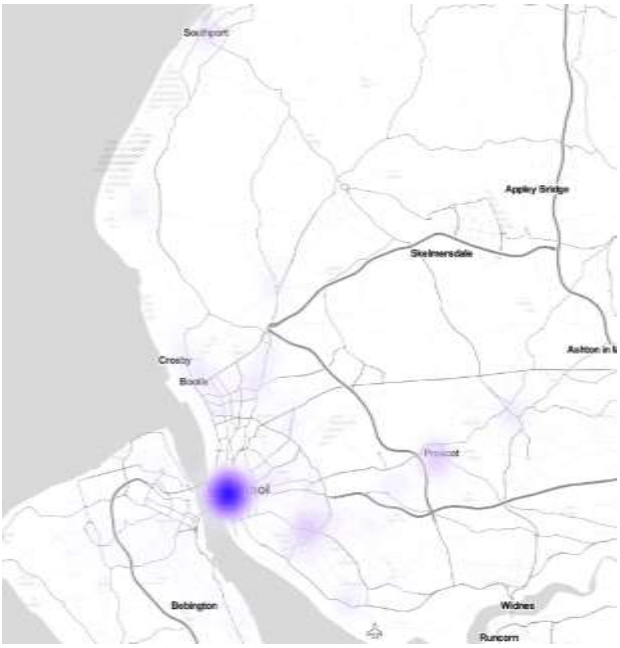


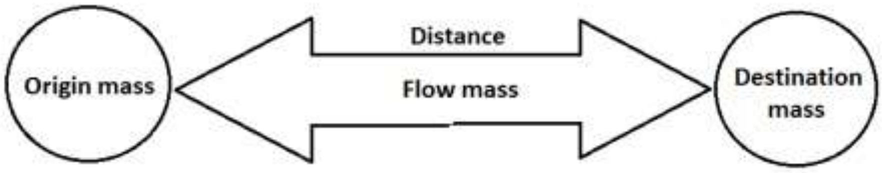
Figure 11: KDE map of flow destinations



3.5 Statistical modelling

To explore the interaction between GP's and pharmacies, we adopt four different model specifications of spatial interaction models from the literature (Figure 12). We used a Generalized Linear Model (GLM) framework to explore which is the best for describing the prescription flows from GPs to Boots pharmacies.

Figure 12: Composition of spatial interaction model



3.5.1 Model specification

There was constructed five spatial interaction models to simulate the possible variation of prescription flow nature.

Firstly, the unconstrained model assumes that the number of items on a flow can be simply explained by the mass of items prescribed at the origin GP and the mass of items dispensed at the destination pharmacy, such as the basic gravity model. However, that can often result in a misconception of interaction, which often has a much more complex economic structure (Fotheringham, 1983; Fotheringham, et al., 2001). Second, production constrained model assumes that the flows are constrained by the prescription production of each GP, i.e. the performance of each GP's prescription production is unique. The third model, attraction constrained, have a similar assumption about the pharmacies, so that the flows are constrained by the prescription dispensation power of each Boots pharmacy. The fourth model then is a combination of both constrained version and it assumes that the flow of prescriptions is dependent on the unique character of the GP as well as the Boots pharmacies. Because of the methodology behind the constrained models doesn't allow to incorporate other variables that could possibly explain the flow variation (Patuelli & Arbia, 2016; Dennett, 2017), the last model complemented by origin/destination specific variables is built in order to compare the performance of constrained and unconstrained models.

Table 3: describes the four different model specifications applied in this section of the analysis.

Model	*Description*
Unconstrained model	gravity model that predicts the prescription flow by GP and pharmacy prescription mass and the distance between the GP and pharmacy
Single, production/origin constrained model	gravity model that predicts the prescription flow by GP's fixed effect, pharmacies prescription mass and the distance between the GP and pharmacy
Single attraction/destination constrained model	gravity model that predicts the prescription flow by pharmacies fixed effect, GP's prescription mass and the distance between the GP and pharmacy
Doubly constrained model	gravity model that predicts the prescription flow by both GP's and pharmacies fixed effects and the distance between the GP and pharmacy

The unconstrained model is a general form of geographical gravity model and written as;

$$T_{ij} = \beta O_i D_j f(c_{ij}). \quad (1)$$

Where T_{ij} is flow mass between the origin and destination, O_i and D_j represents mass elements of origin and destination, constant β is to be estimated and $f(c_{ij})$ is a function of distance between the origin and destination. This is an analogy of the classical hypothesis that the interaction varies proportionally with the product of the two masses and inversely with distance (Eyre, 1999).

For the single constrained models, the origin – destination vector is constrained by the origin or destination function;

$$\sum_j T_{ij} = O_i . \quad (2)$$

or;

$$\sum_i T_{ij} = D_j, \quad (3)$$

where O_i are known parameters but D_j are unknown ones.

For the doubly constrained model, it incorporates a constraint for both origin and destination effects and therefore predicts the size of individual flows between each origin and destination;

$$T_{ij} = A_i O_i B_i D_i d - ij\beta \quad (4)$$

Where

$$O_i = \sum T_{ij} \quad (5)$$

$$D_j = \sum T_{ij} \quad (6)$$

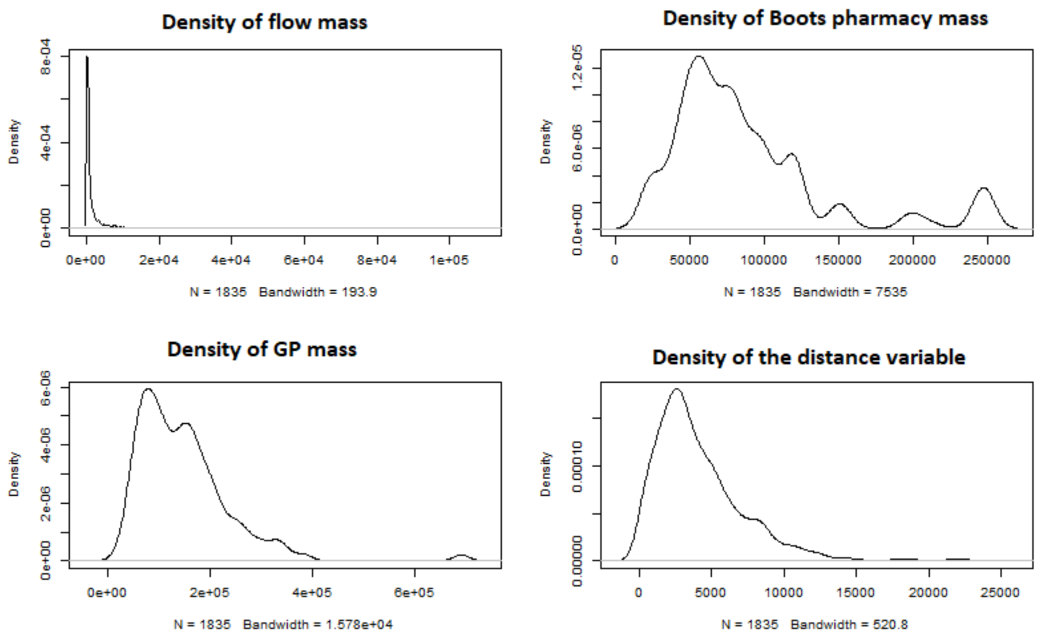
And

$$A_i = \frac{1}{\sum_j B_j D_j d_{ij}^{-\beta}} \tag{7}$$

$$B_j = \frac{1}{\sum_i A_i O_i d_{ij}^{-\beta}} \tag{8}$$

For all model specifications we define the density of flows a linear combinations of density of Boots pharmacy mass, density of GP mass and density of distance;

Figure 13: Histograms of each model component



Unconstrained case:

$$I_{PG} = \beta G_m + P_{ID} + f(D_{PG}) \tag{9}$$

Production-constrained case:

$$I_{PG} = G_{ID} + \beta P_m + f(D_{PG}) \tag{10}$$

Attraction constrained case:

$$I_{PG} = \beta G_m + P_{ID} + f(D_{PG}) \tag{11}$$

Doubly constrained case:

$$I_{PG} = G_{ID} + P_{ID} + f(D_{PG}) \tag{12}$$

Where I_{PG} is the number of prescriptions on the flow, G_m is vector of a 'mass' term of the origins/general practitioners, so number of items prescribed by that GP in a year, P_m is the 'mass' term of the destinations/pharmacies; number of items dispensed in a year, P_{ID} is a constraint defined by vector of the Pharmacy unique codes; ODS codes , G_{ID} is vector of GP's unique ODS codes, β is the estimate to be generated, and $f(D_{PG})$ is distance decay parameter, which measures the distance between the GPs and Boots pharmacies. This assumes that the interaction between GPs and pharmacies decreases with distance. This is expressed as a negative exponential power function;

$$\exp(-\beta \ln D_{PG}). \tag{13}$$

Here we write the function defined as $-\beta \ln D_{PG}$ is simply substituted by $-\beta D_{PG}$ because it better describes the rapid drop off in the observed distance of the interaction Oshan (2016) and Dennett (2017).

This is further described by Oshan (2016) and Dennett (2017), who explains that the function based on inverse power law much better represents the rapid drop off in the observed distance of the interaction. In result, the function defined as $-\beta \ln D_{PG}$ is simply substituted by $-\beta D_{PG}$ in the model. Therefore, the distance vector variable in the model remains unaltered.

3.5.2 General Linear Models

Since the distribution of the dependent variable is highly skewed, the use of a linear model with a Gaussian distribution is not appropriate. Instead we adopt a Poisson log-normal distribution using the GLM framework. Where for the unconstrained model;

$$. I_{PG} = \exp(\beta G_m + \beta P_m + f(D_{PG})) \tag{14}$$

We introduce fixed effects within the Poisson regression to avoid the need to iteratively compute each of the fixed effects in separate models (Oshan, 2016). This is especially true for the computationally demanding doubly-constrained model. These fixed effects behaves like a categorical variable in the regression model so that the first category of the variable is taken as a dummy category and all the following are generated in compare with the first one, which results in n-1 coefficients generated. Therefore, the fixed effects in the constrained models can be interpreted such that the mean predicted flows I_{PG} , is $\exp(\mathrm{B}_G coefficient)$ times larger if they originate from location B_G (Oshan, 2016). Because of this, the order of the data must be taken into account. As the importance of this study is given to high flows between the GP's and pharmacies, the flow data were ordered decreasingly, so that the highest flow in is taken as the dummy variable and all other flows are compared against this.

Table 4: Highest flow used as the comparative flow in regression

	Set 1, 2, 3, 4	*Set 5, 6*
Boots pharmacy_ODS code	FKM99	FN923
Boots_Format	Pharmacy Format 2 (lower Affluence)	Pharmacy Format 2 (lower Affluence)
NHS_contract type	Standard 40hr	Standard 40hr
Boots Name	BOOTS 5974 - LIVERPOOL WHISTON	BOOTS 1307 - WIRRAL ARROWE PARK COMMONFIELD HEALTHCENTRE
Boots_Items _PY	211714	100116
GP_ODS	N83028	N85009
GP_Name	ASTON HEALTHCARE LIMITED	COMMONFIELD RD SURGERY
GP_Flag	Branch	Main
GP_Partners	3	5
GP_Items_PY	690347	139782
Items	109227	90318
distance	2268.621	3.952059
distance_city_centre	3758.77	7165.346

3.6 Catchment analysis

A separate analysis was conducted using the unconstrained model and several socio-demographic factors. This was to explore the relationship their between prescription flows and also enhance the predictive power of the model. This model was selected over the other four model specifications because it represents a traditional way of modelling prescription flows in the literature, assuming that GPs and pharmacies interactions are related to the local demographic characteristics. We simply extend the model through;

$$I_{PG} = \exp(\, \beta G_m + \beta P_m + (-\,\beta D_{PG}) + \beta Var_a \ldots \ldots \ldots \beta Var_x \,) \qquad (15)$$

3.6.1 Covariates

Based on the literature, we selected potential sociodemographic covariates and their impacts on prescription flows. The full list of variables definition and generation can be found in Appendix, however, the overview of their extraction is provided. We included walking distances of 800m between the origin and destination points in order to explore the demographic influence of the area. This is because several studies have shown that most of the UK population resides within 1.6 km distance or a 20 minute walk between pharmacies and GP's and more than half of the population reside within 800 m distance of pharmacies (Law, et al., 2011; Todd, et al., 2015).

We included other demographic factors, including total daytime population, total night time population, the proportion of the elderly population defined as 65 plus, the proportion of people with ethnicity other than white or white-mixed, the proportion of people without a car and Indices of Income Deprivation (IMD) score. The proportion of elderly population was included because this group represents lower levels of health compared to other age groups and so it can be expected that these groups have higher prescription rates than other groups. IMD score was calculated as the mean IMD score falling within each of the 800m catchments for both the GPs and pharmacies.

We incorporated another covariate that calculated the Euclidian distance from each Boots pharmacy to the Liverpool city centre, in explore the effects of urbanity level and people's density level on prescription flows. To compare competition to Boots pharmacies, we included the number of items dispensed per year by their competitors in each catchment as well as the amount of items dispensed by Boots pharmacies per year in each catchment. We also defined another covariate that described the attractiveness of the Boots pharmacies in relation to shopping opportunities or falling within retail catchments. Based on definition of retail attractiveness used in Huff's shopping probability model (Huff, 1963), the definition of pharmacy retail attractiveness is as follows;

$$Attractivnes\ of\ the\ pharmacy = \frac{Size\ of\ the\ closest\ retail\ catchment}{Distance\ (Euclidian)\ to\ closest\ retail\ catchment} \tag{1}$$

We also included Boots format classification of their Boots pharmacies. This divides their pharmacies based their customer characteristics such as their behaviour in addition to the size of the pharmacy. Another Boots covariate we included was the type of contract with the NHS or retail park hour's contract. Finally, we include a covariate which describes the type of GP, as well as another which gives the number of doctors in each GP.

3.6.2 Variables validation

We applied the Variance Inflation Factor (VIF) to ensure the covariates are not strongly correlated or have multicollinearity. Covariates that are strongly correlated explain the same thing and therefore do not add any value in applying them to our models. The VIF showed that 9 variables had high multicollinearity, so were removed from the models. These included; Pharmacy shopping attraction, pharmacy proportion of elders, ethnically diverse population and population without the car, pharmacy night time population density and mean deprivation score, GP's proportion of elders and population without the car and lastly GP's mean deprivation score.

3.7 Empirical analysis

In total, 5 models were applied to the dataset to explore which best explains prescription flows;

1. Unconstrained spatial interaction model
2. Unconstrained spatial interaction model complemented by possible influential variables
3. Origin constrained spatial interaction model
4. Destination constrained spatial interaction model
5. Doubly constrained spatial interaction model

Firstly, we applied these models to all 6055 flows. According to (Tobler, 1987) selective information aggregation and elimination is an effective approach for detecting flow spatial patterns. To explore the effects of different flows, we defined 5 more flows (Table 4). We firstly selected the top 500 flows with the highest amount of prescriptions, to explore which Boots pharmacies are the most productive. The third analysis was applied to flows with more than 100 prescriptions, as frequently applied by Boots UK. The fourth analysis looked at flows greater than 771 items or the mean number of prescription flows for 2017. The last two analysis are based on prescription flows that are within 1.6 kilometres and 800 metres of GPs and Boots pharmacies.

Table 5: Global statistics of dependent variable (Items/Flow) for each tested data set

Data Set	*Min*	*Mean*	*Median*	*Max*	*Flow count*
1 - All flows	1	771	16	109227	6044
2 - 500 highest flows	1299	8140	3515	109227	500
3 - Flows with more than 100 Items	101	2503	426	109227	1835
4 - Flows with more than Mean Items	773	6212	2220	109227	685
5 - Flows within 1600m	2	8683	3268	90318	354
6 - Flows within 800m	2	16688	9655	90318	141

3.8 Local variation

We applied an extension of the GLM to the unconstrained model using a Generalized Linear Mixed Model (GLMM);

$$I_{PG} = \exp(\beta G_m + \beta P_{ID} + (-\beta D_{PG}) + \varepsilon) \tag{16}$$

This allowed us to explore the local heterogeneity of prescription flows as a function of the distance to the city centre, through the incorporation of both fixed and random effects. The GLMM fits the random-effect parameter by the maximum likelihood, which is very convenient in simple cases, where covariates are normally distributed and the random effects are nested effects. However, the GLMM can be challenging to fit in more complex models (Bolker, et al., 2009). Therefore this model was applied only to the selected flows of GPs within 800m of pharmacies. The variable 'distance to the city centre' was chosen over the others for several reasons. Firstly, it can be taken as a local spatial effect simulating the urbanity level and peoples density. Secondly, it is possibly the least biased variable towards the model.

3.9 Model selection

To measure the model fit or how well a statistical model describes how well it fits a set of observations, we use three indicators. We apply the three most commonly applied the R-squared, Akaike's Information Criterion (AIC) and Root Mean Square Error (RMSE). (Anselin, 1992; Nakaya, 2001; Patuelli & Arbia, 2016). The R-squared represents the proportion of variance in the outcome that is explained by the model, where a higher value indicates a better model fit. Conversely, as the number of variables increases, the R-squared value naturally increases. AIC penalize the inclusion of additional variables, where lower values indicate a better model fit and therefore offers a better indicator of model fit. The biggest limitation of AIC is that it does not really provide a test of the model performance, but relates its quality to other models (Kassambara, 2018). RMSE measures the average squared difference between the observed actual outcome values and the values predicted by the model. Similar to AIC lower values of RMSE indicate a better model fit (Kassambara, 2018).

We also applied a predictive check to visualize the predictive power of each of the models. Here we generated 300 samples from each of the estimated models and cross-validate against the predicted and actual flow distribution. The predictive check is based on the assumptions of a presence of a predictive and inferential uncertainty. The first is an uncertainty arising with the fact that the model formula does not capture all the elements explaining the true data or their right form in which the elements generate the interaction. The later represents a fact that we never get to know the true parameters, just their estimates from the data sample available. Therefore, generated graphs simulate the feasible scenarios under the model and uses those to assess uncertainty (Arribas-Bel, 2017). Nevertheless, we also take into consideration the nature of flows and the purpose of this study when considering which model best describes Boots UK prescription flows.

4 Results

4.1 Modelling Interactions

We first present the distribution spatial interactions defined by the five spatial interaction models (table 3.5.1) applied to the different flow interactions within Merseyside, defined in section 3 (table 4). Figures 14;15 shows the flow interaction between GPs and pharmacies, defined as the volume of prescriptions flows. Figure 14 shows that people tend to use Boots pharmacies that are closest to their GP. This is because the distribution of prescriptions are very skewed, where is a very high density of prescription flows for the smallest number of items. In other words, as the number of items increases, the density of prescription flows decreases. For all flow interactions, the mean and the median remain very close to each other and very much at a peak of the distribution (Figure 14). Interestingly, as the volume and distance of flows decreases, the number of items dispensed decreases. Moreover, as the volume and distance of flows decreases, the mean distribution of flows shifts to the right or a more normal distribution.

As expected, as both the number of flows and flows by distance decreases, the spatial interaction of flows significantly decreases (Figure 15). Interestingly, set 2; 6 shows a clear divide between prescription flows in Merseyside across the river Mersey (Figure 15). This suggests that this geographical barrier plays a strong influence on spatial interactions of prescription flows in Merseyside. Moreover, although the interactions represent straight-line Euclidean distances, the interactions suggest they follow road networks. This again supports the literature on the decision process of patients selecting a pharmacy, which tends to be the closest one, along a road network.

Figure 14: Histograms of the dependent variable Items (number of prescriptions on a flow) in each dataset

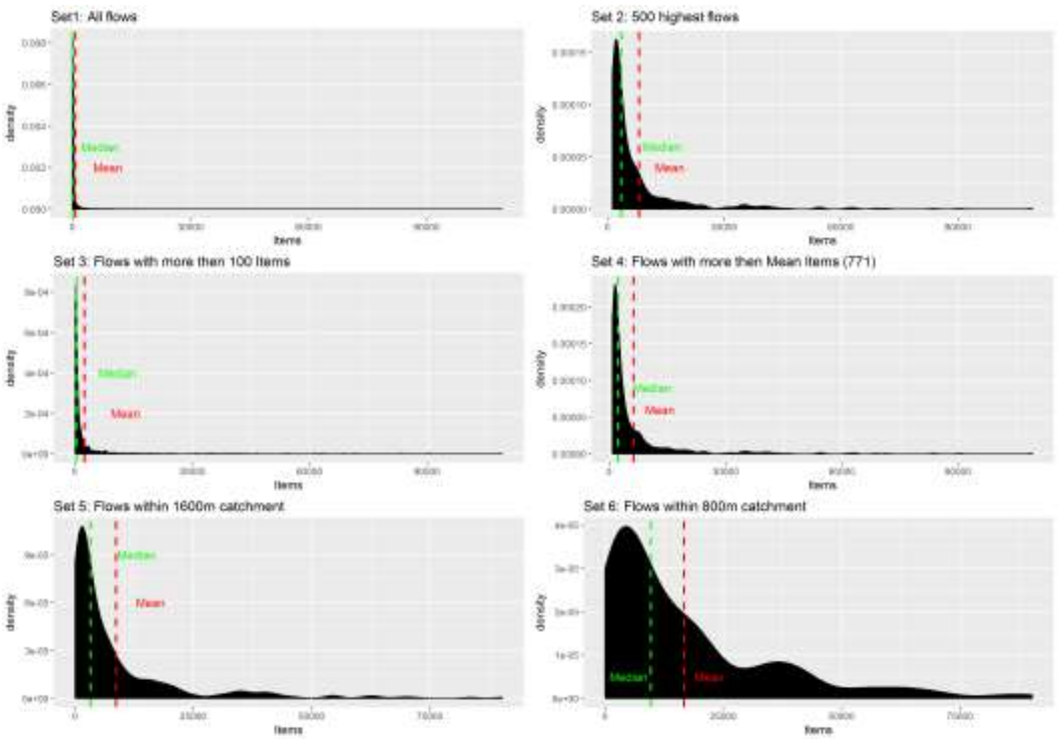
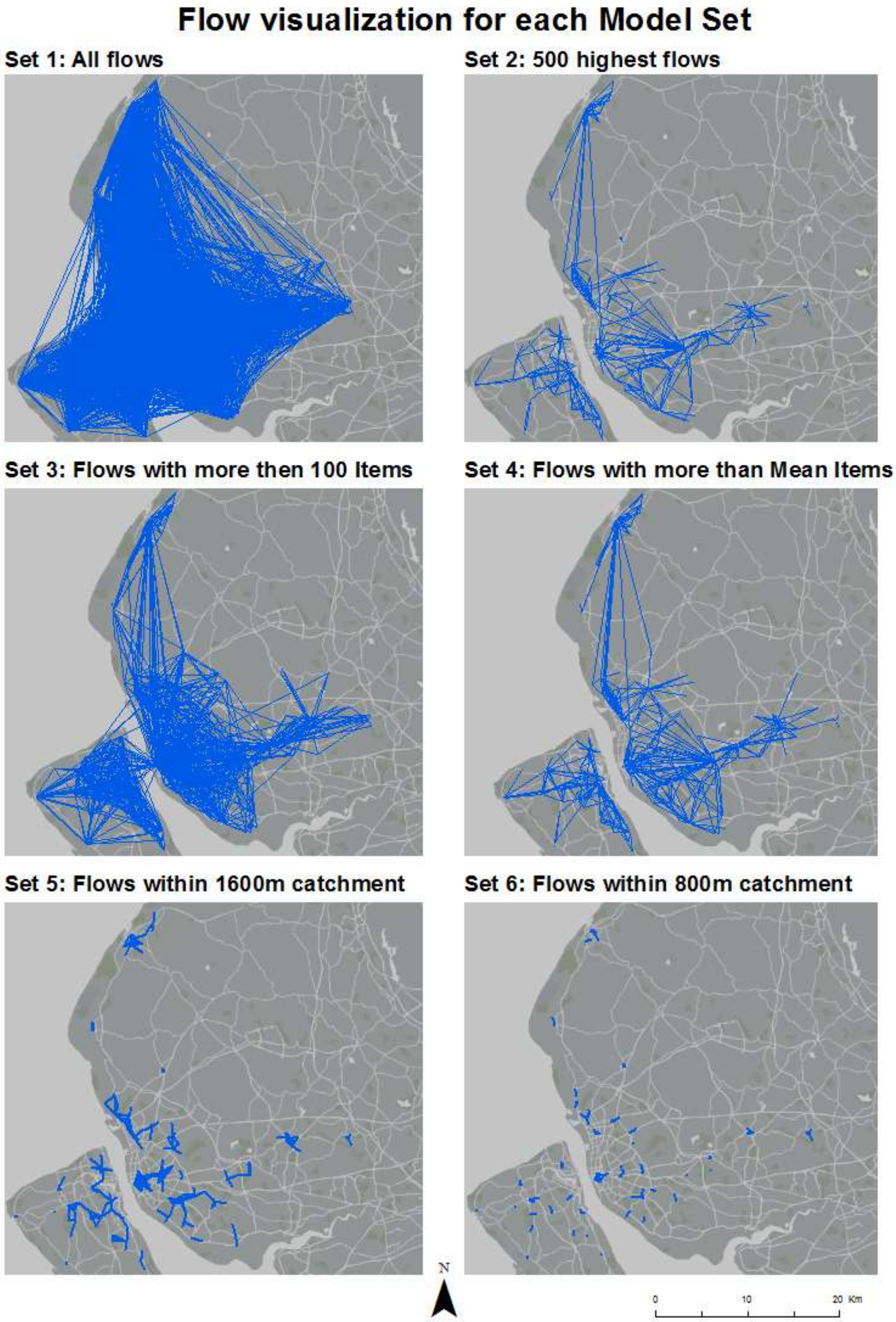


Figure 15: Visualization of the prescription flows



4.1.1 Unconstrained model

The results of the simple unconstrained models are shown in Table 6, alongside the application to the different flow interactions. All the covariates in this model indicate that there is a significant relationship with prescription flows. Generally, the effect of pharmacy mass variable has a stronger association than the GP mass variable. For the distance covariate, the influence of distance is strongest for flows within 800m distance and weakest for the flows within 1.6km distance.

Table 6: Coefficient estimates of unconstrained models

	Set 1		***Set 2***		***Set 3***	
(Intercept/Items)	-10.54	***	-4.334	***	-9.169	***
log(GP_Items_PY)	0.7733	***	0.5342	***	0.7038	***
log(Boots_Items_PY)	0.9759	***	0.685	***	0.9224	***
distance	-0.0008131	***	-0.0004923	***	-0.0006867	***
	Set 4		**Set 5**		**Set 6**	
(Intercept/Items)	-6.011	***	-22.73	***	-0.9089	***
log(GP_Items_PY)	0.579	***	1.329	***	0.4731	***
log(Boots_Items_PY)	0.7802	***	1.334	***	0.5298	***
distance	-0.0005473	***	-0.0004516	***	-0.002891	***

*Signif. codes: 0 '***' 0.001 '**'*

4.1.2 Production-constrained model

By introducing the constraints in the form of the GP's unique ODS codes, we examined the fixed effects of each practice on the interaction model. The effect of the distance on the flow strengthened in all data selections ranging from 0.03% to 0.5% prescription difference for 1m distance, as well as the prescription mass dispensed by pharmacy strengthened in most of the model sets except the set of flows within 1.6km. The GP's effect then varies from -6.5 to 2.8 throughout the model sets. The least versatile is the fourth model set in compare with the sixth model set that has observed the highest differences. As an example, it can be said that the Lance Lane Medical Centre is 0.5 times less likely to generate flows then the Aston Healthcare Limited considering the flows above its mean. On the other hand, considering flows within 800m catchment, the Lance Lane Medical Centre is 1.2 times less likely to generate flows.

Table 7: Shortened regression table of production-constrained model of all Sets showing the lowest and highest coefficient estimates

	Set 1		***Set 2***		***Set 3***		***Set 4***		***Set 5***		***Set 6***	
(Intercept)	-1.942	***	0.1708	***	-0.1935	***	-0.3538	***	-5.63	***	0.2598	***
log(BO_I_PY)	0.9321	***	0.8163	***	0.8993	***	0.8797	***	1.273	***	1.167	***
distance	-0.00072	***	-0.00038	***	-0.00059	***	-0.00043	***	-0.00044	***	-0.00527	***
GP_A	2.843	***	1.805	***	1.046	***	1.939	***	2.127	***	1.344	***
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
GP_Z	-6.489	***	-2.608	***	-4.451	***	-2.324	***	-6.503	***	-10.21	***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

4.1.3 Attraction-constrained model

Similarly to the production constrained model, attraction-constrained considers a fixed effect of each pharmacy on the interaction. The negative effect of the distance on the flows is very similar to production constrained model, ranging from 0.07 % to 0.33% decrease of prescriptions on flow with 1m distance. Pharmacy effects are less versatile here than in production-constrained model and vary from -2.4 to 2.6 throughout the data selections. The highest variability is observed in fifth data selection. As an example of the effect of particular pharmacy on prescription, here it can be said that the Boots pharmacy in Liverpool Clayton Square Shopping Centre is 0.3 times more likely to generate flows than the Liverpool Whiston Boots pharmacy when considering all the flows, although, examining the flows within 1.6km distance, the Liverpool Clayton Square Shopping Centre is 0.7 times more likely to generate flows in compare with the Liverpool Whiston Boots pharmacy.

Table 8: Shortened regression table of attraction-constrained model of all Sets showing the lowest and highest coefficient estimates

	Set 1		***Set 2***		***Set 3***		***Set 4***		***Set 5***		***Set 6***	
(Intercept)	0.1209	***	3.019	***	1.042	***	2.769	***	-7.664	***	4.375	***
log(GP_Items_PY)	0.7715	***	0.5329	***	0.6911	***	0.5393	***	1.302	***	0.5533	***
distance	-0.0008	***	-0.0004	***	-0.0006	***	-0.0004	***	-0.0005	***	-0.003	***
Boots pharmacy_A	1.621	***	1.939	***	1.733	***	1.958	***	2.623	***	1.227	***
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
Boots pharmacy_Z	-1.194	***	-1.811	***	-1.086	***	-1.526	***	-2.409	***	-1.796	***

*Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1*

4.1.4 Doubly constrained model

Here the interaction is examined with both fixed effects of GP's and fixed effects of pharmacies, so the coefficients are generated for each particular origin and destination in the data. The GP's effect varies from -8.2 to 5.5 throughout the datasets and pharmacies effect varies from -2.1 to 10.7 throughout. An example of difference in pharmacy effect can't be given with the Liverpool Clayton Square Shopping Centre Boots pharmacy, which is 1.3 times more likely to generate flows than the Liverpool Whiston Boots pharmacy when accounting for the all flows, but when accounting for flows of more than 100 items, same pharmacy is 1.5 times more likely to generate flows then the Liverpool Whiston Boots pharmacy. Along with the pharmacies, the example of the difference between GP's effect can be given with the Lance Lane Medical Centre, which is 2.2 times more likely to originate flows than the Aston Healthcare Limited when considering all possible flows, however, considering flows above 100 prescriptions, same GP centre is 0.1 times less likely to originate flows in compare with the Aston Healthcare Limited.

Table 9: Shortened regression table of doubly-constrained model of all Sets showing the lowest and highest coefficient estimates

	Set 1		*Set 2*		*Set 3*		*Set 4*		*Set 5*		*Set 6*	
(Intercept)	6.207	***	8.656	***	7.974	***	7.552	***	7.652	***	14.29	***
distance	-0.0009	***	-0.0005	***	-0.0007	***	-0.0005	***	-0.0005	***	-0.007	***
Boots pharmacy_A	10.7	***	7.632	***	9.82	***	8.092	***	8.116	***	2.796	***
.		.		.		.		.		.		.
Boots pharmacy_Z	-0.3362	***	-1.64	***	-0.12	***	-1.056	***	-2.13	***	-1.108	***
GP_A	5.474	***	3.854	***	3.509	***	4.651	***	2.64	***	1.064	***
.		.		.		.		.		.		.
GP_Z	-7.03	***	-7.765	***	-8.167	***	-6.781	***	-6.439	***	-11.73	***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

4.2 Catchment analysis

Comparing the coefficients of the catchment regression analysis (Table 10), there can be observed a strong relationship of the GP Flag division and Boots pharmacy Format specification on prescription in most of the data selections. In comparison, the least influential parameters are those defining the competition in surrounding of pharmacies, distance to city centre and the population estimate variables. A note must be given to the proportion of ethnically diverse population in GP catchments, which has a higher negative relationship to flows then the GP's total population. Therefore, it can be said that an increase in the proportion of ethnically diverse groups results in a decrease in the number of prescriptions flows. Significantly, this negative effect reverses for flows within 800m, where there is a positive relationship. Similarly, distance to the city centre and pharmacy Format has a positive effect with 800m catchment but a negative effect in all other selection. Furthermore, variables such as Higher and Lower affluence pharmacy Format, Local pharmacy format or 40 hours NHS pharmacy contract show a negative effect in all Sets except 1.6 km catchments.

Table 10: Regression coefficients of unconstrained models with an origin/destination specific variables

	Set 1		*Set 2*		*Set 3*		*Set 4*		*Set 5*		*Set 6*	
(Intercept/Items)	-5.443	***	-4.95	***	-5.111	***	-5.278	***	-9.821	***	-2.003	***
log(GP_Items_PY)	0.5685	***	0.3538	***	0.4887	***	0.3802	***	1.023	***	0.4783	***
log(Boots_Items_PY)	0.897	***	0.9424	***	0.9188	***	0.9654	***	0.8038	***	0.5771	***
distance	-0.0008052	***	-0.0004228	***	-0.0006658	***	-0.0004879	***	-0.0004524	***	-0.002427	***
distance_to_city_centre	-0.000001507	***	-0.00001521	***	-0.000006968	***	-0.0000128	***	-0.00002631	***	0.000004489	***
GP_FlagMain	-0.812	***	-0.5082	***	-0.7405	***	-0.5769	***	-1.384	***	-0.07222	***
GP_FlagMain with Branch	-0.7331	***	-0.4295	***	-0.6103	***	-0.5402	***	-1.319	***	-0.5101	***
GP_Partners	0.05023	***	0.06298	***	0.05132	***	0.06266	***	-0.004211	***	0.0606	***
GP_nightime_population	-0.0006186	***	-0.0004396	***	-0.0006204	***	-0.0005094	***	-0.0001319	***	0.0002503	***
GP_ethnicity_proportion	-0.01744	***	-0.002354	***	-0.02049	***	-0.0128	***	-0.0163	***	0.01427	***
Boots_daytime_population	0.0002386	***	0.001044	***	0.000315	***	0.0009401	***	-0.000932	***	0.0007137	***
FormatDestination H&B	0.5231	***	0.3766	***	0.4933	***	0.3651	***	0.4255	***	-0.05926	***
FormatLocal Boots H&B	0.3388	***	0.6625	***	0.397	***	0.549	***	-0.6037	***	0.5476	***
FormatPharmacy Format 1 (Higher Affluence)	0.5493	***	1.155	***	0.6498	***	0.9901	***	-0.9941	***	0.6823	***
FormatPharmacy Format 2 (lower Affluence)	0.372	***	1.071	***	0.5402	***	0.9024	***	-0.924	***	0.7967	***
Boots_competitors_Items_PY	-7.083E-07	***	3.403E-07	***	-0.000000195	***	0.000000473	***	-0.000002797	***	-6.236E-07	***
NHS_DescRetail park exempt	-0.6625	***	-0.5208	***	-0.6389	***	-0.5203	***	-0.5439	***	-	
NHS_DescStandard 40hr	-0.349	***	-0.3192	***	-0.2019	***	-0.2093	***	0.02512	***	-0.928	***
Competitors_Items_PY	-0.000001037	***	-7.599E-07	***	-0.000001092	***	-9.079E-07	***	-0.000001125	***	-0.000001143	***

*Signif. codes: 0 '***' 0.001 '**'*

4.3 Model validation

Firstly, each of the models was diagnosed for normality, homoscedasticity, multicollinearity and outliers. All the data selections and applied models revealed outlying residuals, non-normality and possible bias, see example Figures 16-18

Figure 16: Diagnostics of simple unconstrained model of Set 1; all flows

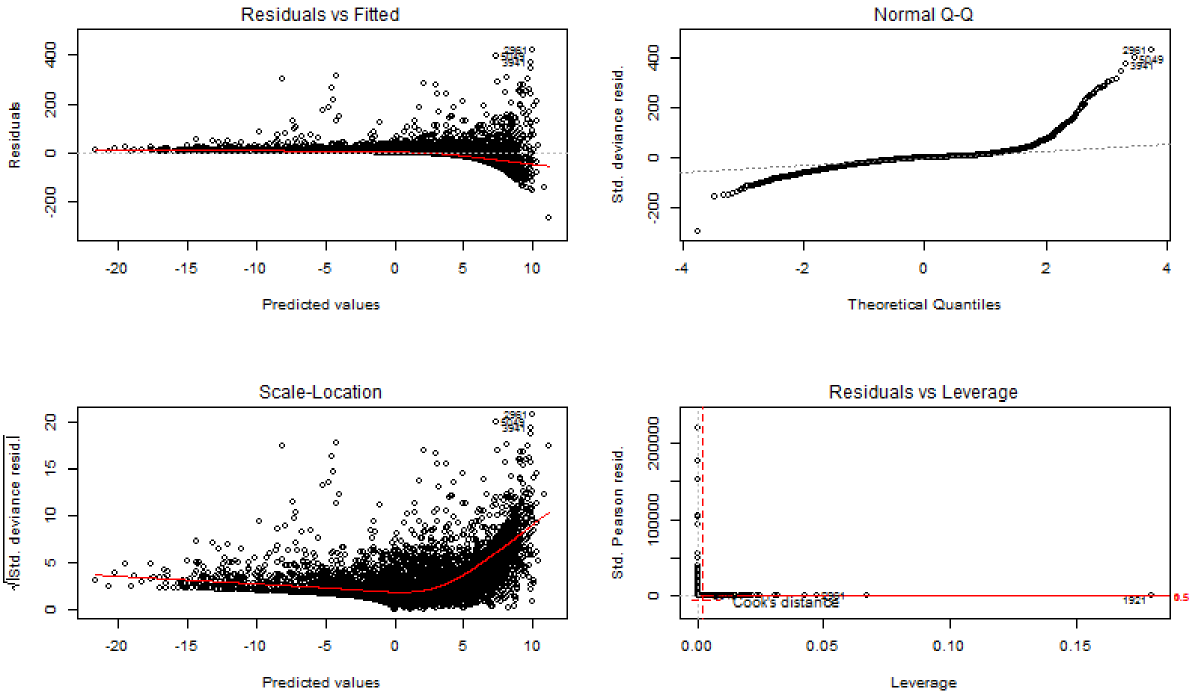


Figure 17: Diagnostics of Production-constrained model of Set 3; Flows above 100 items

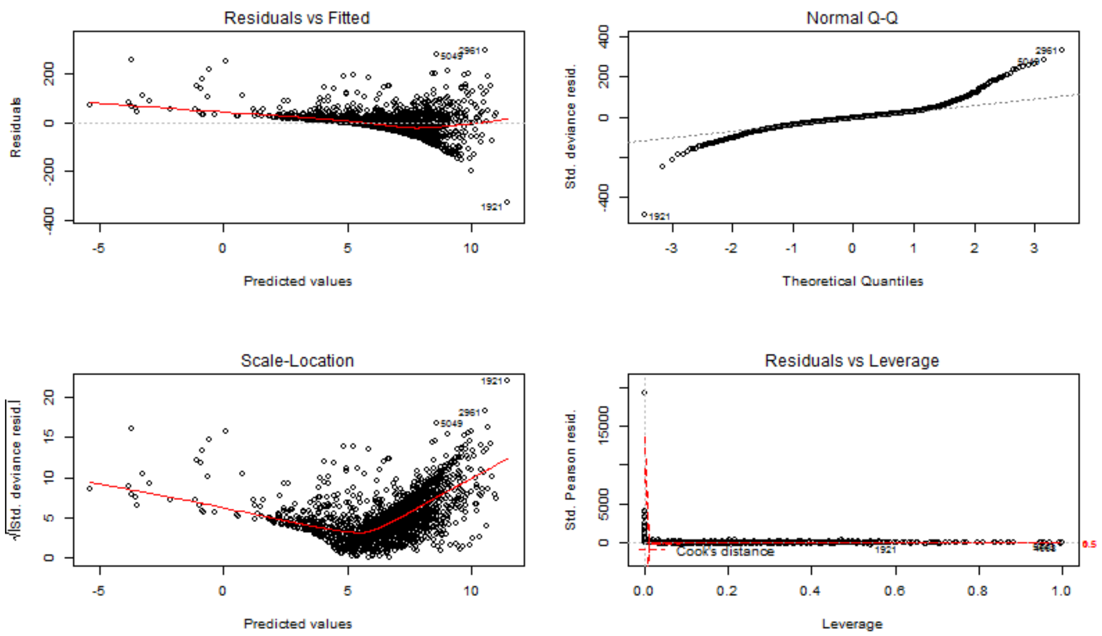
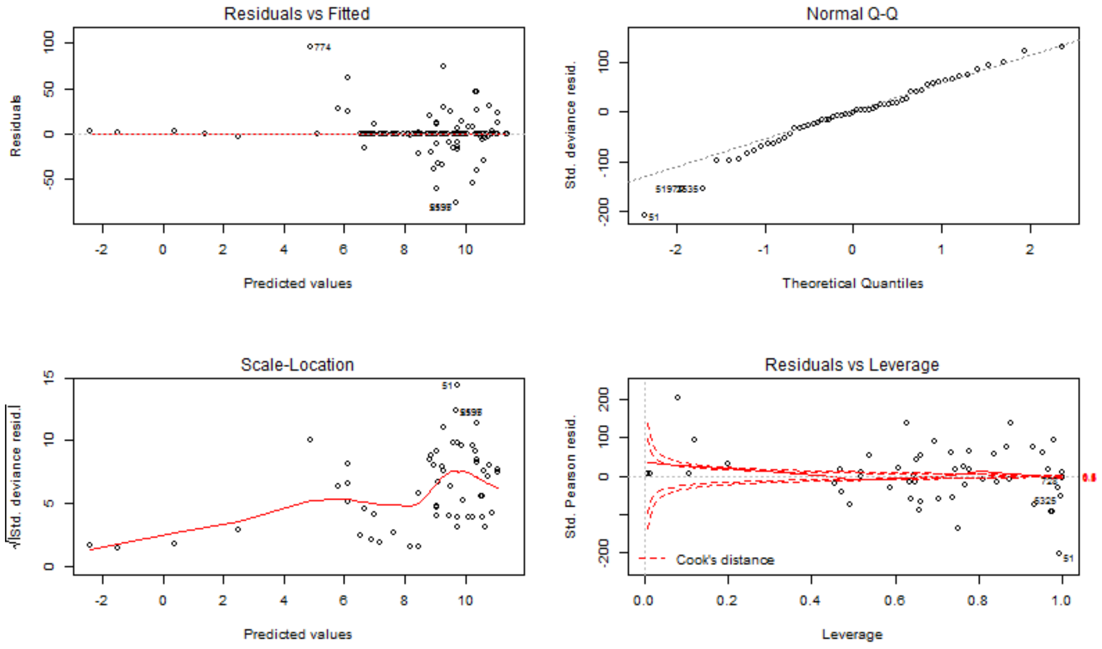
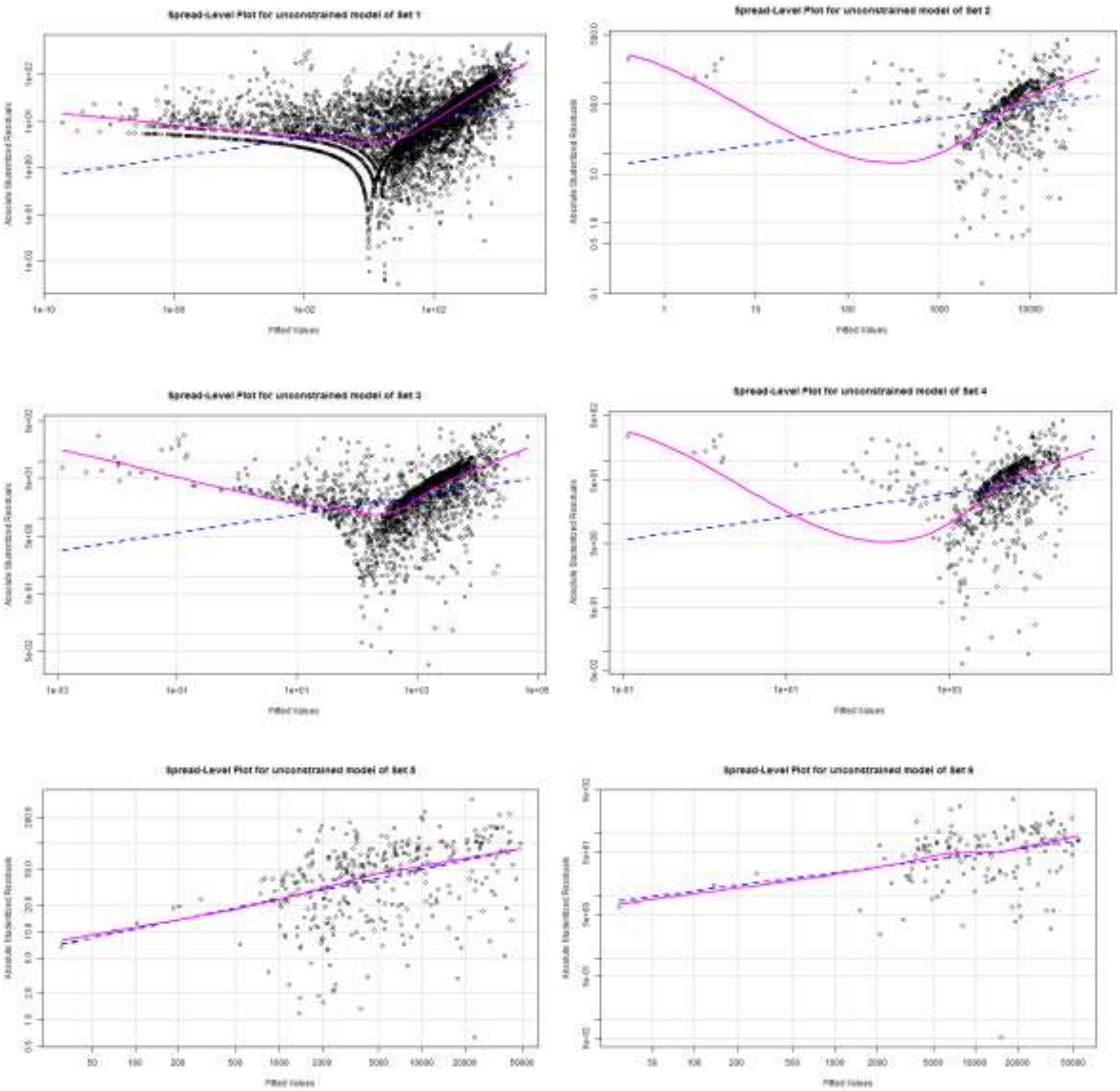


Figure 18: Diagnostics of Doubly-constrained model of Set 6: flows within 800m distance



Although anomalous residuals are present in most models, this is most significant for the very high flows. We decided to include all outliers because this is factual information and the exclusion of one results in new ones being replaced. The fulfilment of the linearity of the model is achieved in Dataset 5 and 6 (Figure 24), where the distribution of residuals and fitted values are close. There can be observed a clear, similar trend of the error variance distribution of Set 1, 3 (Figures 19, 21, 23). It shows a stripe-like structure that is divided into two wings, which strongly suggests a two-stage hierarchical structure of the model. Set 2 and 4 shows similar structure as both represents 'high flows', although here, the residuals vs fitted values are focused on a right side of the plot (Figures 20, 22). Overall this suggest there the way how the flows are generated is a two-stage process, so flows with a smaller number of prescriptions are generated differently than the flows with a larger number of prescriptions, which support the assumption of diverse decisions people make when choosing their pharmacy destination.

Figure 19-24: Plots examining constancy of error variance of simple unconstrained model for all 6 data selections



For All flows, there was un-constant error variance in all interaction models. This is a possible consequence of the nature of the data. Firstly, the dependent variable is extremely skewed and have a rapid decay of the prescriptions on the flow, secondly, independent variable distance follows the inverse power function and so decay even more rapidly. From the estimated regressions, the doubly-constrained model shows the best fit and can explain up to 83% of the variations. Comparing the single constrained models, production-constrained indicates better fit then attraction-constrained, its AIC is 24% higher as well as R squared increased by 7%. Unconstrained models generally have the highest AIC, RMSE and explain the least variations. Complementing model with geodemographic explanatory variables can explain up to 72% of the variations, which is just 2% more than the simple unconstrained model. Therefore, there is a noticeable similar model fit between the unconstrained model with covariates and the attraction-constrained model. Overall, it can be said, that modelling prescription flow can be achieved best by the incorporation of GP's fixed effects or by incorporation of both GP and pharmacy fixed effects.

Table 11: Validation of tested models in their datasets

	Set 1 All flows				**Set 2** 500 highest flows			
Model	**df**	**R^2 - logit**	**AIC**	**RMSE**	**df**	**R^2 - logit**	**AIC**	**RMSE**
Unconstrained	4	0.70	7207411	3211	4	0.38	3395918	10553
Unconstrained + covariates	19	0.72	6726401	3008	19	0.51	2694044	8770
Production constrained	261	0.80	4896482	2479	220	0.74	1440299	7152
Attraction constrained	63	0.73	6527061	2945	63	0.56	2403091	8456
Doubly constrained	320	0.83	4068601	2156	276	0.84	874750	4683
	Set 3 Flows with more than 100 Items				**Set 4** Flows with more than Mean Items (771)			
Model	**df**	**R^2 - logit**	**AIC**	**RMSE**	**df**	**R^2 - logit**	**AIC**	**RMSE**
Unconstrained	4	0.56	5868403	5830	4	0.42	4098561	9247
Unconstrained + covariates	19	0.60	5338938	5313	19	0.52	3416730	7879
Production constrained	256	0.73	3701149	4482	239	0.72	2011549	6654
Attraction constrained	63	0.62	5178741	5254	63	0.55	3164241	7764
Doubly constrained	315	0.78	3004134	3725	297	0.80	1420083	4892
	Set 5 Flows within 1600m catchment				**Set 6** Flows within 800m catchment			
Model	**df**	**R^2 - logit**	**AIC**	**RMSE**	**df**	**R^2 - logit**	**AIC**	**RMSE**
Unconstrained	4	0.66	1716476	8793	4	0.60	986644	12423
Unconstrained + covariates	19	0.73	1386378	7582	18	0.72	693924	10087
Production constrained	193	0.90	493985	3919	110	0.96	98024	3307
Attraction constrained	62	0.76	1190592	6885	57	0.81	479343	7841
Doubly constrained	233	0.94	284785	2976	124	0.98	58413	2319

The selection of the 500 highest flows and flows above the mean flows, shows that some of the high flows also have a long distance and its error distribution has significantly different from the distribution of All flows. Although the doubly-constrained model for those two selections can explain a lot of variations, considering all of the applied models, it can be said that using the flows with a large number of prescriptions results in worst model fit.

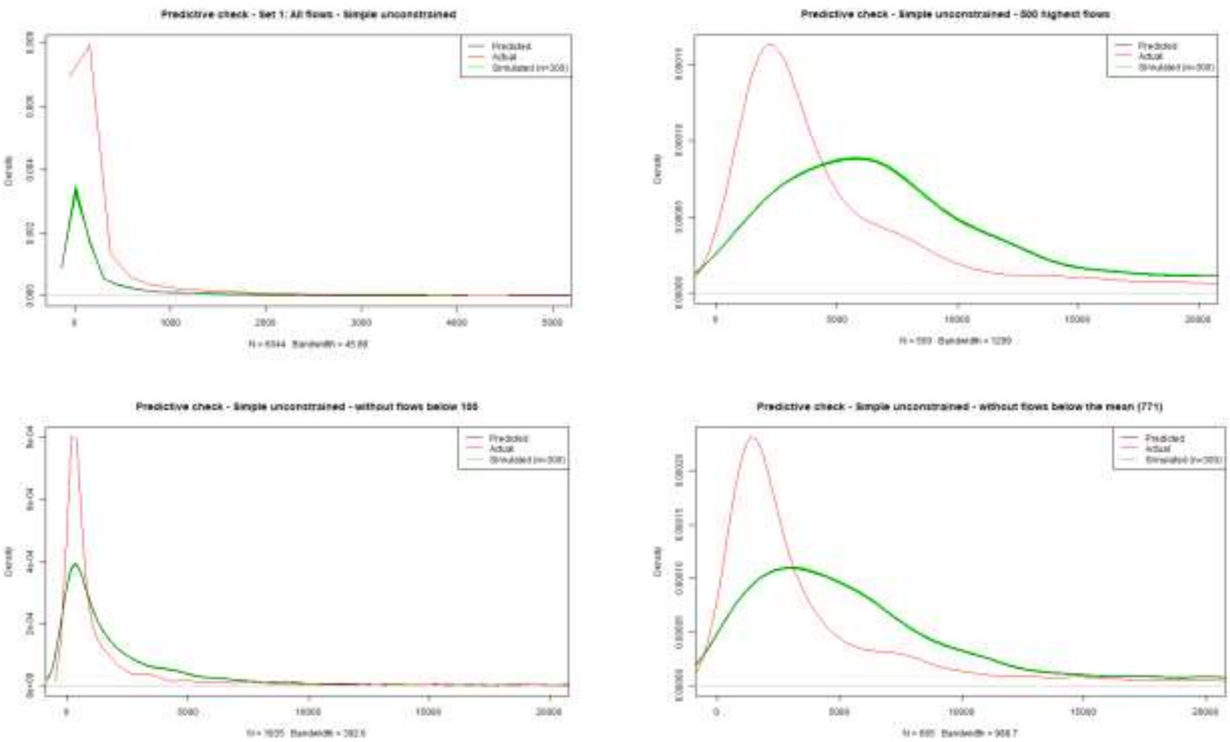
The residuals distribution of models using the flows with a number of items larger than 100, is visibly more representative to all flows than the other selection. Although the AIC number lowered in all cases, the RMSE increased in all cases and the model explains approximately 10% less of the variations than models applied on full flow data. The models of 1.6 m distance flow selection show very familiar residual distribution to Set 1. Although its unconstrained model explains as much as 66% of variations, its production-constrained and doubly-constrained model can explain more than 90% of the variations. Although the AIC dropped significantly in a doubly-constrained model, the RMSE does not lower below the RMSE of a doubly-constrained model of all flows. Furthermore, there can be seen a larger gap between the model fit of the unconstrained model and doubly-constrained model.

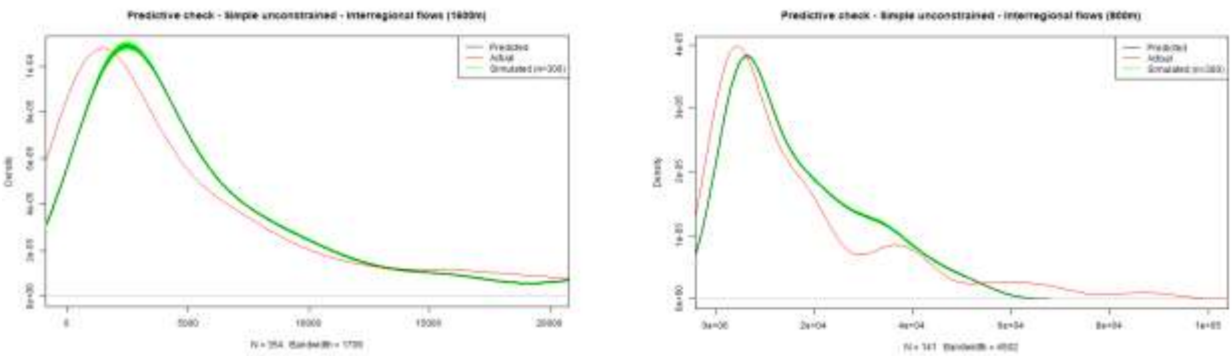
The same pattern as in Set 5 can be observed in Set 6; flows within the 800 m distance. Although the doubly-constrained model results in the best model fit from all tested models, there is a larger gap between the simple interaction model and the constrained models. The reason for such a model fit and variation in Set 5 and 6 can be possibly explained firstly, by the low number of observations, origins and destination incorporated in a model, in comparison with the full flow data and secondly by the nature of selection, which assume that the important interaction occurs on limited distance between the GP and pharmacy.

4.4 Cross-validation

The generated prediction check plots are the last piece of the validation to consider for evaluation of the best model. The graph basically reveals how the model behaves on various ranges of distribution when generating predictions. Comparing the selections of flows, the simulations for an unconstrained model are fitted the best to actual data in the selections of 1.6 km and 800 m distanced flows, where the simulated distribution very much follow the shape of actual distribution (Figure 25, 26), and the worst in the selections of 500 highest flows and flows above the mean, where the peak is shifted towards a normal distribution and the prediction mostly overfit the actual data (Figure 27, 28). Prediction of unconstrained model for Set 1 and 3 appears to be more under-fitting but respects the shape of the actual distribution (Figure 29, 30).

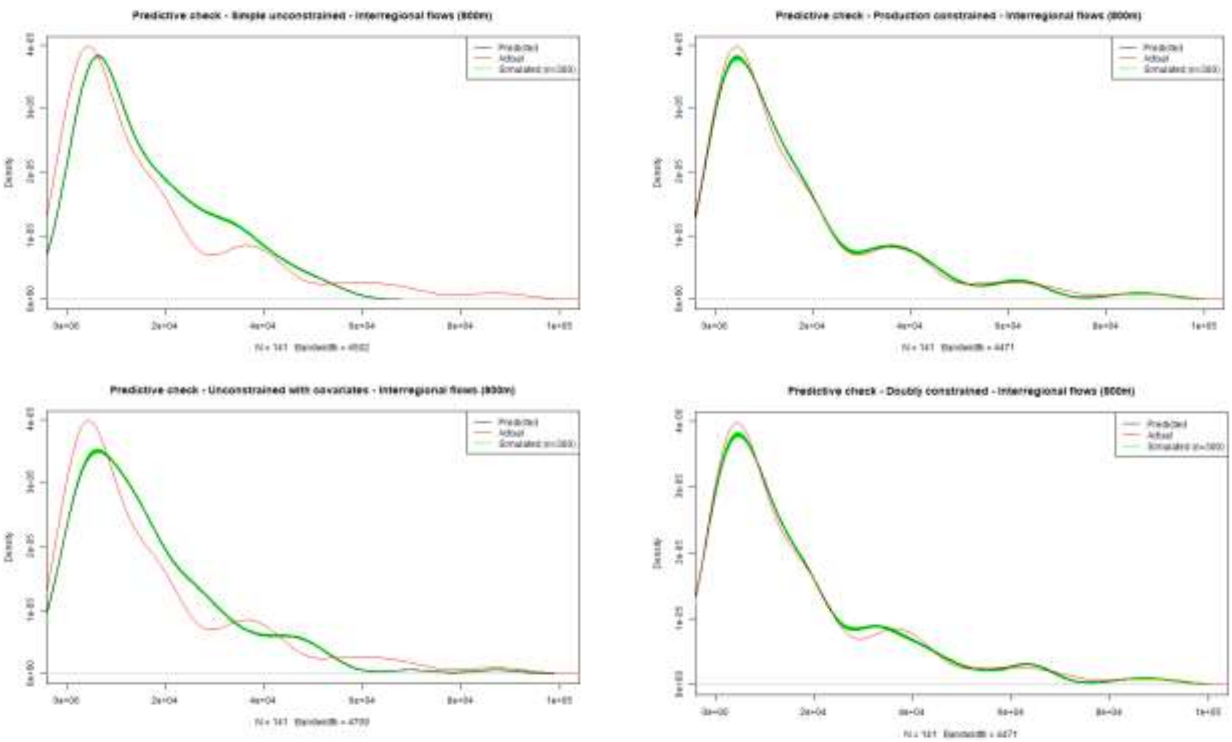
Figure 25-30: Predictive check for unconstrained models of all sets





There are observed significant differences between the simulations of the interaction models. Simulation on unconstrained interactions results in the least fitted model, where the prediction generally holds the global shape of actual data bud does not follow the local shape of the actual distribution (Figure 31-32). In compare, the production-constrained and doubly-constrained very much follows both global and local shape of the actual distribution (Figure 33-34).

Figure 31-34: Predictive check, demonstration of interaction models fit for Set 6



4.5 Random effects

Inclusion the random effect of the distance to the city centre was computational difficult within the constrained models, although, there is a significant improvement in model fit within the unconstrained models. Most importantly, the random effect accounts for the spatial distribution of flows related to urbanity, which is almost necessary if the scale of the study is bigger than local.

Table 12: Validation of mixed-effect models accounting for random effect of distance to city centre

Unconstrained model	*df*	*AIC*	*marginal R2*	*conditional R2*
Set 1: All flows	5	6527574	0.99303	0.999954
Set 2: 500 highest flows	5	2403627	0.740715	0.999923
Set 3: Flows above 100 items	5	5179251	0.964279	0.999937
Set 4: Flows above mean items	5	3164770	0.828176	0.999925
Set 5: Flows within 1.6km distance	5	1191096	0.857352	0.999956
Set 6: Flows within 800m distance	5	479800	0.843593	0.999981

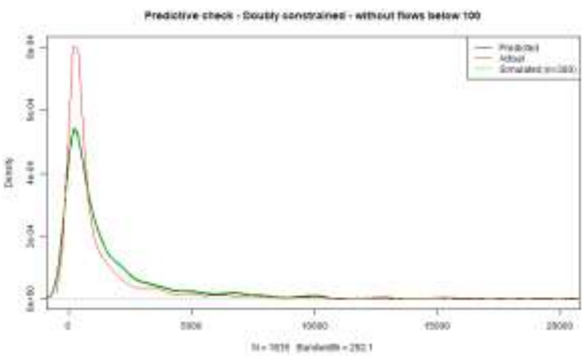
4.6 Model Selection

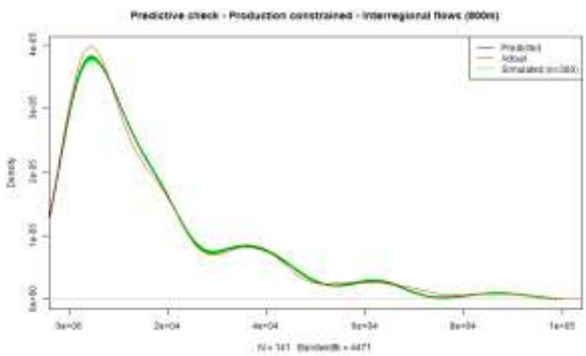
The model fit and validation revealed very diverse behaviour of interaction models in various situations. It can be said that the less observation captured, the less valuable become the general gravity model and the more valuable become the constrained models. Therefore, the model selection for a spatial interaction between GP's and pharmacies is greatly dependent on a scale and a purpose of analysis.

Important assumption made is, the interaction is overall much more dependent on the effect of GP surgeries than on effect of pharmacies, as the prediction distribution to actual distribution is more closely fitted. Consequently, the production-constrained and doubly-constrained models are the most relevant on most of the scales.

To reach the aims of this analysis, model a spatial interaction between GP's and pharmacies on a scale of Merseyside, it is essential to consider all the possibly non-random flows to capture the most of the possible variations. Therefore, a doubly-constrained model considering all flows above 100 prescriptions per year, that can explain up to 78 % of the variability, can be considered as the best fit (Tables 2-4 in Appendix). Although the predictive check of this model is not efficient in predicting flows with an extremely high number of prescriptions (Figure 35). Those 'high flows' are mostly a result of a close distance between GP and pharmacy, and so if the aim of the study would be to assess the interaction of the closes surgeries and its effect on pharmacies, the production-constrained model considering flows within the 800m distance would be a more suitable choice. In this case, there could be explained up to 96% of the data variations with the lowest uncertainty (Figure 36).

Figure 35-36: Prediction check for the selected models





The result of the doubly-constrained model for flows over 100 prescriptions can be easily interpreted with two maps that demonstrate the coefficient estimates for each origin and destination (Figure 37, 38).

The regression coefficients for Boots pharmacy show clearly the highest and the lowest estimates at Merseyside area. The Boots pharmacies that are most likely to attract the patients are majorly located on a north of Sefton borough and on the outskirts of Liverpool, Knowsley and St.Helens, while the pharmacies which attract the patients least likely are located on a north of Liverpool and south Sefton with a traces in a city centre. Interestingly, the estimates of the GP shows almost reverse distribution. The GP's that are least likely to generate flows are located on a north of the Sefton borough and outskirts of other boroughs, while the GP's that are more likely to attract patients are located on a north of Liverpool and south of the Sefton. Although the estimates are visibly clustered around certain areas, the pattern strongly reminds the pattern of urbanity of the area (Figure 39).

Figure 37: Map of pharmacy coefficient estimates of doubly-constrained model for flows over 100 items

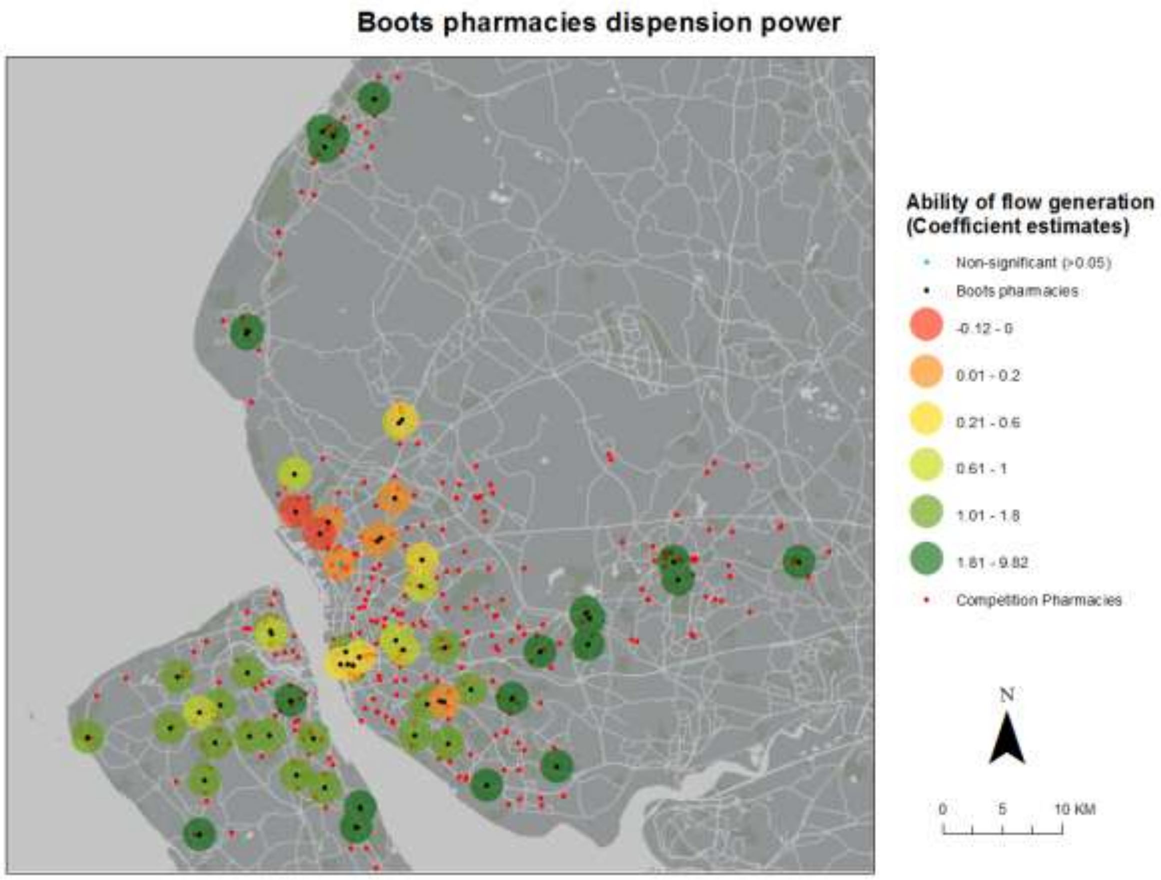


Figure 38: Map of GP's coefficient estimates of doubly-constrained model for flows over 100 items

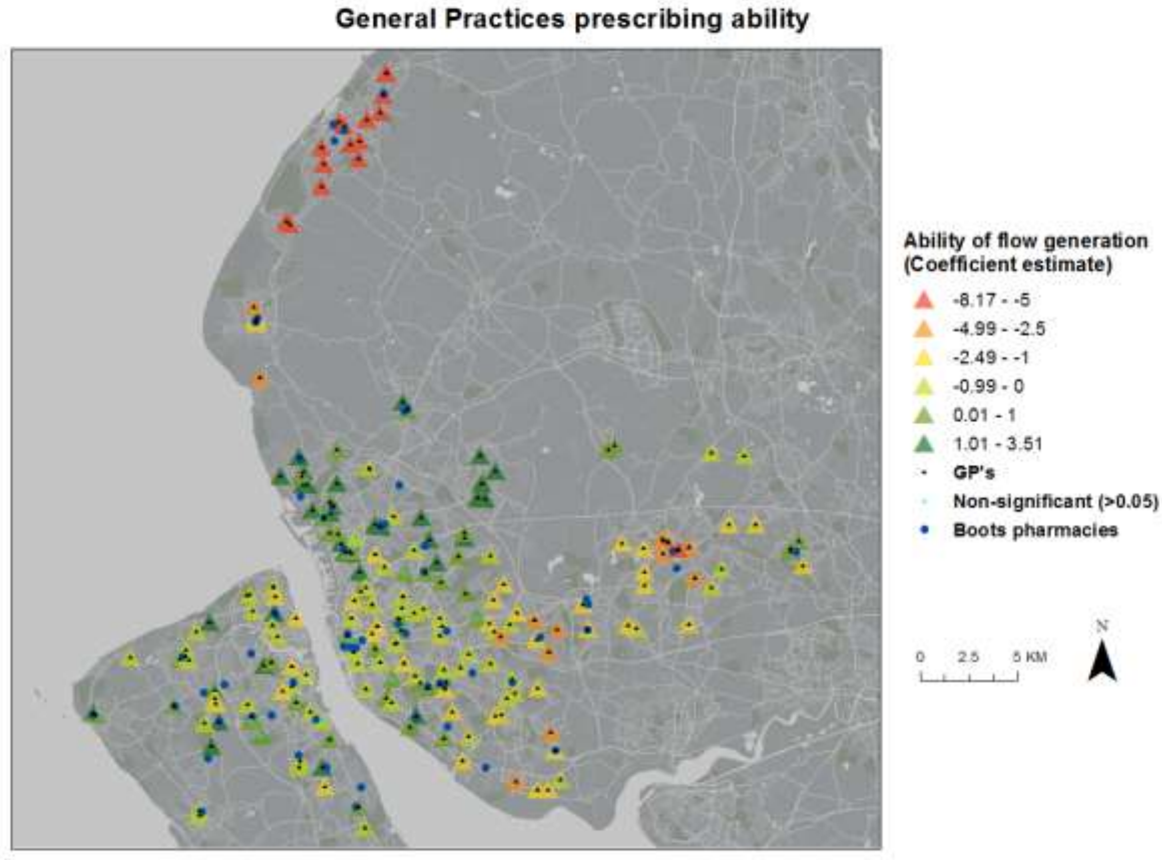
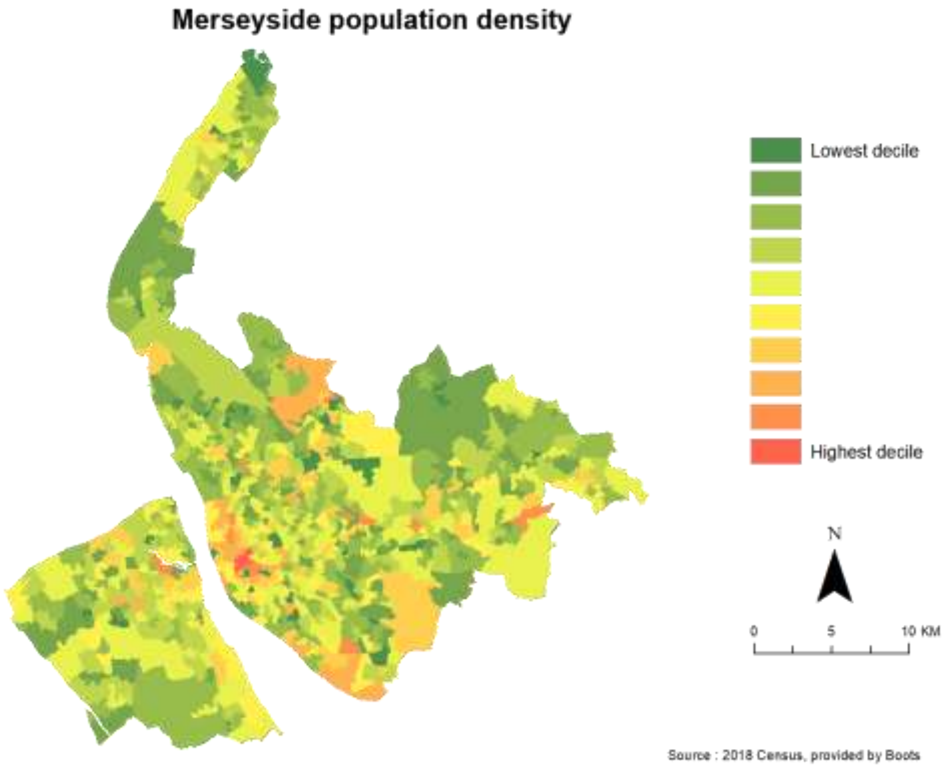


Figure 39: Population density derived from provided population estimates (2018)



5 Discussion

5.1 Application

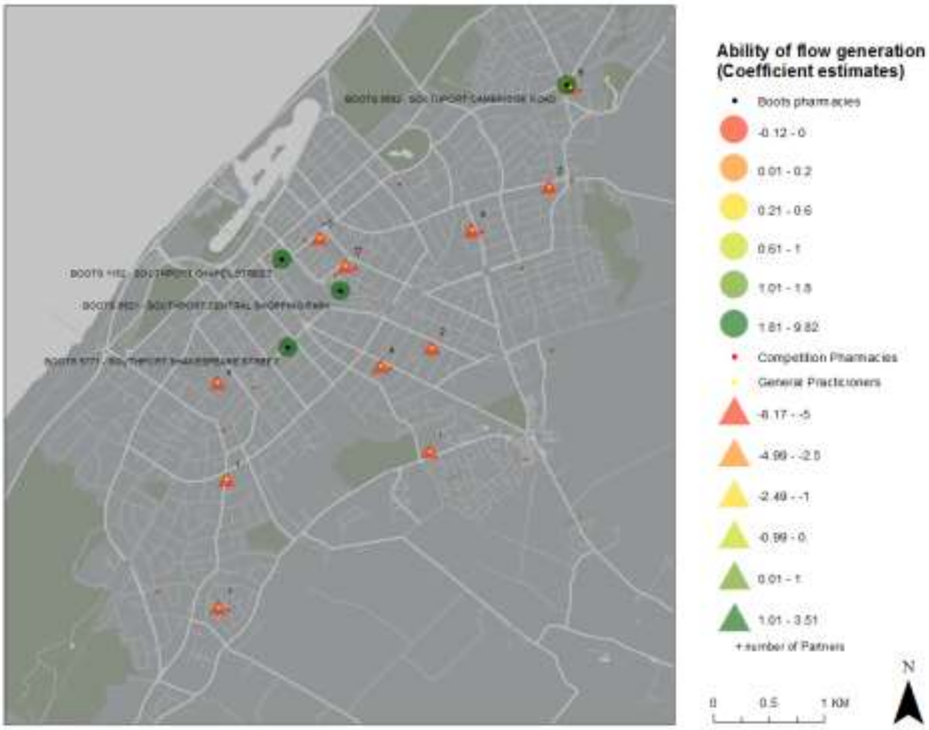
By applying different model specification using a GML framework to different prescription flow interactions between GPs and Boots pharmacies, we assessed the nature of Boots prescription flows in Merseyside. The performance of the constrained models proved to be the most efficient and easiest to use for understanding spatial interactions of prescription flows, compared to the traditional geodemographic catchment analysis using the unconstrained model. Moreover, the simulation-based cross validation showed the constrained model has the highest predictive performance compared to the unconstrained ones. Therefore, in other application of forecasting future prescription flow interactions, this model would also be preferred.

Although, the constrained model of flows was the best model, it explained just 78% of the variations. Moreover, we demonstrated the limitations of this approach for analyzing the distribution of above 100 prescriptions. This is because these close interactions behave significantly different than other flows and are result of a direct influx of patients from the GP. Instead, the production-constrained model performed better for close-distanced interactions. This suggest that the close-distanced interactions have a different interaction flow generating process, compared to the other flow types. This suggests that the hierarchical structure dependence of people's decision making plays a bigger role, supporting the literature.

The limitation of the constrained models is that fixed effects are included and have a comparative characteristic. This was because the highest prescription flow was the dummy variable where all other flows were compared against. Since this high flow is at a distance more than 2000m, the application of the different model specifications for lower flows do not perform as well as the high ones. Whilst we could have specified different criteria for the dummy variable, which gives us better coverage of most of the flows apart from the very highest and lowest, we had no prior information of what defines "best flows".

According to the maps demonstrating the coefficient estimates, a unique spatial distribution of prescription flows can be observed in Southport, Sefton borough. The GP's flow generation effect in this area is the lowest in Merseyside, which can firstly relate to the geographical position of the area. Secondly GP's are evenly distributed throughout this area. Thirdly, the interactions are also directly related to competing pharmacies, where GP-Boots pharmacy relationships are higher than their competitors. Therefore, the low density of Boots pharmacies in this area, are related to their high prescription flow generating process.

Figure 40: Efficiency of the flow generation in Southport area

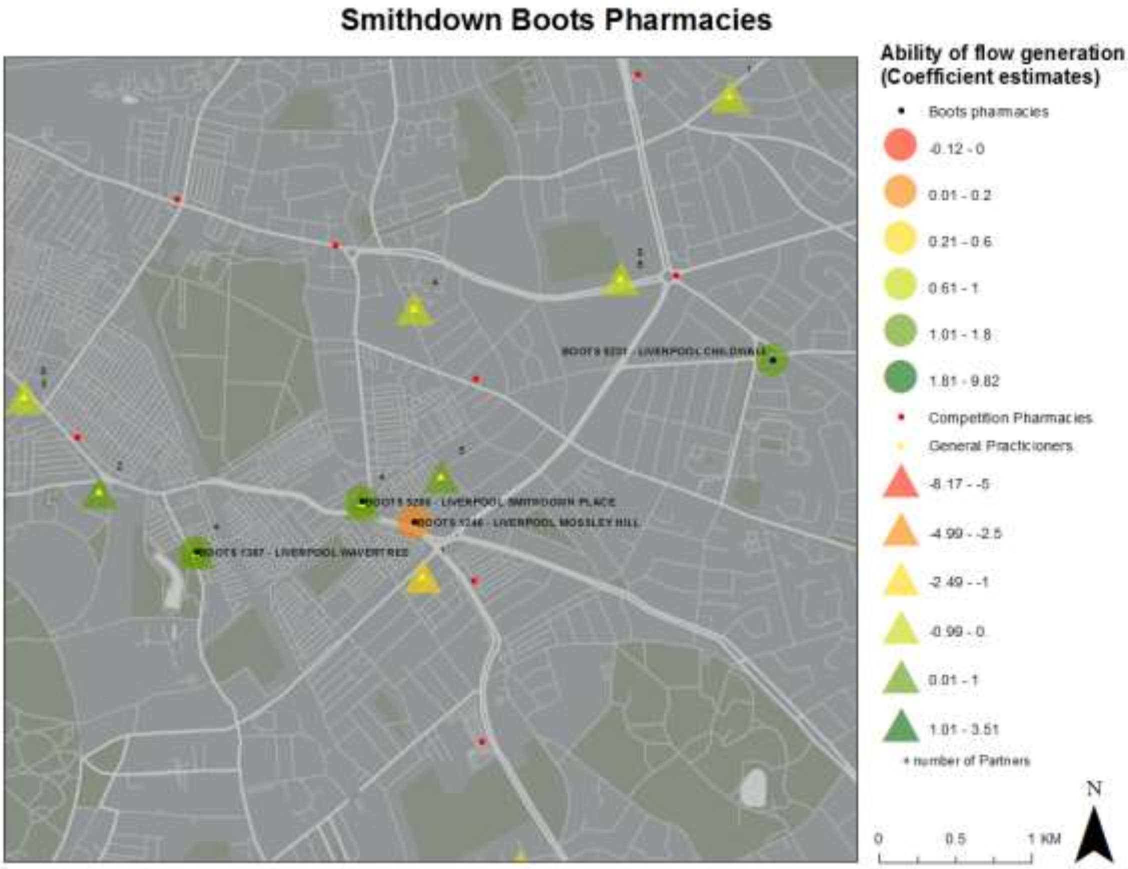


Similar situation of performance cluster is observed in south Sefton with low efficient Boots pharmacy flow production. The lack of close-distanced relationships with GP's and competing pharmacies allocated in those close distances cause a loss of prescription flows and its unsustainability for Boots pharmacies in that area.

Figure 41: Worst performing Boots pharmacies in south Sefton area

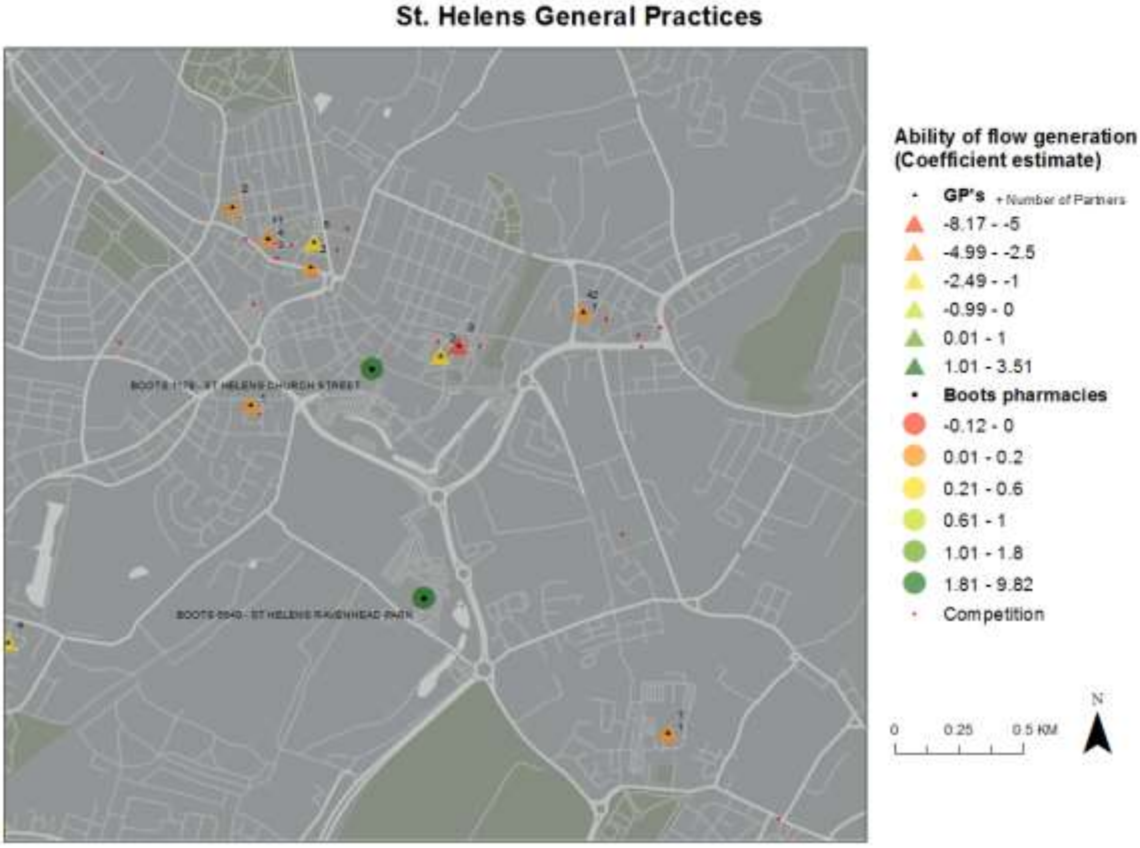
Although the pattern of the models coefficient estimates refers to spatial dependence, some of the estimates remains low throughout multiple models and those could potentially represent pharmacies or GP's that requires change. For example, the Mossley Hill Boots pharmacy is located on a high street, surrounded by three GP's. However, most of the flows in that area are attracted by Smithdown Place Boots pharmacy that is located just 210 meters away and is incorporated with a GP surgery. This suggest that the Mosley Hill pharmacy could be joint to Smithdown Place pharmacy without loss of prescriptions to their competitors.

Figure 42: Closer look at Smithdown pharmacies and their unusual situation



Investigating the GP's estimates, the central area of Saint Helens shows cluster of GP's with low flow generation ability. In that area most of the flows are attracted by Boots competitors that are more closely located to the GP's. By relocating the Boots stores on a high street (Saint Helens Church Street Boots) and a retail park (Saint Helens Ravenhead Park), their flow attraction ability is extremely high. This suggest that the Boots might not be using the full potential of the area and its GP's and allocating the pharmacy in closer distance to GP cluster on high street could potentially generate higher Boots prescription flows in this area.

Figure 43: Closer look at St. Helens centre that shows unusual results



All these results are highly influenced by the dummy variable 'best flow' and using a different dummy variable would result in different estimates. For example, using the lowest flow as a 'best flow'. Overall, the prescription flow estimates between origins and destinations should be taken cautiously because the predictive uncertainty increases with decreasing distance of interaction.

In conclusion, there has been made two assumptions about the prescription flows that require future theoretical and methodological investigation. Firstly, this analysis suggests a presence of spatial hierarchical structure of the prescription flows based on urbanity that should be further explored. Secondly, it suggests a presence of clustering of the flow generation process based on human behaviour and their decision making. For example, when the patient leaves GP surgery, his first choice will be the closest pharmacy, which is most like stocked with the needed item. The second choice could be a pharmacy in the patient's residence, where he possibly know the pharmacists and can easily access needed medicaments. Thirdly, a choice might be made to go to a pharmacy in a city centre, high street or shopping park, where the patient can make additional shopping purchases within the same trip.

5.2 Limitation and challenges

Geographical data and flows especially are mostly very high in volume due to the nature of their collection and its type. Availability of such a data increases in time, however, several limitations remains in the methodology of its analysis. Firstly, most of the existing methods are focused on the certain perspective of the flow data, secondly, many traditional methods can't process large data volume, and thirdly, the spatial aspect of the flows can be often difficult to store and address.

Therefore, it can be said that the main practical issues are linked to data aggregation procedure. This study uses as little as 0.4% of the original data due to aggregation and omitting missing values, which can give a misrepresentation of the true generating flow process.

Moreover, the nature of dependency within the prescription flow data violates the independent and identically distributed assumption of regression models. Figures 15-20 illustrate that the residuals of the models have a periodical, non-random structure, which can only be adjusted by manipulation with the data. This suggests that the rules by which are flow generated vary across the data, therefore, the spatial interaction model are possibly most effective when applying on specific selections of the data.

One of the limitations of the favoured constrained models is a difficulty linked to the inclusion of other variables that could possibly explain the flow values. Assessing the interaction on a county level, it has been confirmed the presence of spatial effects on the prescription data. The importance of a spatial dependence and heterogeneity is suggested firstly, by the significance of variables such as 'distance to the city centre' and 'Boots pharmacy format', which describes the allocation of origins and destinations within the space and population structure. Secondly by accounting for random effects of 'distance to the city centre', that enhance the generation flow and lastly by the generated density plots. The Liverpool city centre attracts a significant proportion of prescription flows in Merseyside, so that it should be treated differently from the rest of the study area.

5.3 Further research

Another step towards a better estimation of flows is incorporating spatial effects through spatial modelling techniques, such as Bayesian approaches. Introducing a Bayesian Hierarchical structure allows each of the parameters including the spatial one to be modelled as stochastic processes at each level. LeSage (2007) introduced a Bayesian hierarchical regression model with estimated individual effect parameters and latent spatial effect structured to follow a spatial autoregressive process, to investigate origin-destination commodity flows. Another example could be a recent study also by LeSage (2016), which examine commuting-to-work flows using Poisson pseudo-maximum likelihood method extended to allow for spatial dependence between flows from nearby regions.

Furthermore, using recently introduced technologies of machine-learning such as deep-learning models which can help to detect spatial and temporal flow patterns of sparse are non-normal data and build an accurate prediction. There can be found several recently published articles that use deep-learning to predict flow patterns in traffic flow (Polson & Sokolov, 2017; Yi, et al., 2017; Ezzatabadipour, et al., 2017).

In spite of the relevance of suggested methods, the major and primary step in further research is to build a better understanding of the prescription flows. It is necessary to detect and describe the underlying structure of flows with convenient statistical or econometric tools. Building a matrixes for origin-destination flows with single set of entities is already challenging task, but building matrixes of origin-destination flows for two sets of entities such as GP's and pharmacies is far more difficult and is not sufficiently described by current literature. Lastly, expanding the focus of the analysis on all pharmacy chains, could reveal hidden patterns, increase the prediction accuracy or even tackle the gaps in prescription flow generation. This would require better performance hardware, more specific parameters calibration or conducting the analysis on a much smaller scale such as boroughs, urban districts or towns.

6 Conclusion

There has been demonstrated various literature discussing the distribution and availability of health care services (Law, et al., 2011; Todd, et al., 2015), as well as problematics of spatial flow analysis and related issues with analysing human behaviour (Fotheringham, et al., 2001; Rodrigue, 2017). Analysing prescription flows is important to identify potential customers and identify underperforming pharmacies. However, there lacks an application of spatial interaction modelling of prescription flow between GPs and pharmacies (Griffith, 2009). Here we presented five model specifications which was applied to different flows characteristics. This was to examine which one best estimates prescription flows between GPs and pharmacies in North West England.

The results indicate that the doubly constrained model best estimated prescription flows, with more than 100 prescriptions, accounting for a 78% model fit. However, for flows with less than 100 prescriptions, it didn’t have the best estimation because of the different prescription flow generating process Instead the production-constrained model applied on flows as close as 800 metres, that results in more than 90% accurate prediction. We also analysed the association between different covariates and the spatial distribution of high and low prescription flows. As expected, Liverpool had the highest volume of prescription flows, given its presence as the retail core of the region. In contrast, Southport had the lowest volume of prescription flows. Moreover, we identified several Boots pharmacies where the movement of these stores could give higher prescription flows and

Throughout the investigation, there has been found a presence of an underlying hierarchical structure of the prescription flow linked to both spatial dependence and heterogeneity of patient’s movement in space. By presenting the random effect of distance to Liverpool city centre, the interaction models illustrated much higher precision and model fit. Use of spatial interaction models at a smaller scale could be beneficial for exploration of the local relationships, however, we propose a more robust theoretical research of the prescription flows nature. It is essential to identify the hierarchical structure of the people's decision behind the choice of pharmacy to give better future estimates. Overall, this study highlights the value of applying spatial interaction models for estimating prescription flows between GPs and Boots pharmacies. This methodology could be extended to predicting future prescription flows.

REFERENCES

The Health Foundation, 2013. *Improving patient flow,* s.l.: The Health Fundation.

Anas, A., 1982. *Residential location markets and urban transportation,* New York: Academic Press.

Anselin, L., 1992. *Spatial Data Analysis with GIS: An Introduction to Application in the Social Sciences,* s.l.: National Center for Geographic Information and Analysis.

Arribas-Bel, D., 2017. Spatial Econometrics. *An Introduction to Spatial Regression.*

Batty, M. & Mackie, S., 1972. The calibration of gravity, entropy, and related models of spatial interaction. *Environment and Planning,* Volume 4, pp. 205-233.

Birkin, M. & Clarke, G., 1995. Using microsimulation methods to synthesise census data. *The census user's handbook,* pp. 363-387.

Bolker, B. M. et al., 2009. Generalized linear mixed models: apractical guide for ecology and evolution. *Trends in Ecology & Evolution,* 24(3), pp. 127-135.

Bowers, S., 2018. 2,500 GPs needed to meet future demand. *The Sunday Times*, 10 June.

Brunsdon, C., Fotheringham, A. S. & Charlton, M., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis,* Volume 28, pp. 281-298.

Congdon, P., 2000. A bayesian Approach to Prediction Using the Gravity Model, with an Aplication to Patient Flow Modeling. *Geographical Analysis,* 32(3), pp. 205-224.

Dennett, A., 2017. *Dr Ds Idiots Guide to Spatial Interaction Modelling for Dummies.* [Online]
Available at: https://rpubs.com/adam_dennett/257231
[Accessed 2018].

Dijkstra, A. et al., 2013. Using Spatial Analysis to Predict Health Care Use at the Local Level: A Case Study of Type 2 Diabetes Medication Use and Its Association with Demographic Change and Socioeconomic Status. *PLOS ONE,* 8(8).

Dong, G., Harris, R., Jones, K. & Yu, J., 2015. Multilevel Modelling with Spatial Interaction Effects with Application to an Emerging Land Market in Beijing, China. *PLoS ONE,* 10(6).

Eyre, H. A., 1999. *Measuring the performance of spatial interaction models in practice,* s.l.: The University of Leeds.

Ezzatabadipour, M. et al., 2017. Deep Learning as a Tool to Predict Flow Patterns inTwo-Phase Flow. *CoRR,* Volume abs/1705.07117.

Fischer, M. M., 2006. *Spatial Analysis and GeoComputation; Selected Essays.* Heidelberg: Springer.

Flowerdew, R. & Aitkin, M., 1982. A Method of Fitting the Gravity Model Based on the Poisson Distribution. *Journal of Regional Science,* 22(2), pp. 191-202.

Fotheringham, A., 1986. Modelling hierarchical destination choice. *'Environment and Planning A,* Volume 18, pp. 401-418.

Fotheringham, A. S., 1983. A new set of spatial-interaction models: the theory of competing destinations. *Environment and Planning A,* Volume 15, pp. 15-36.

Fotheringham, A. S., 1983. Some theoretical aspects of destination choice and their relevance to production-constrained gravity models. *Environment and Planning A,* Volume 15, pp. 1121-1132.

Fotheringham, A. S., Brunsdon, C. & Charlton, M., 2000. *Quantitative Geography.* London: SAGE Publications Ltd.

Fotheringham, A. S. et al., 2001. Hierarchical destination choice and spatial interaction modelling: a simulation experiment. *Environment and Planning A,* Volume 33, pp. 901 - 920.

Getis, A., 2008. A History of the Concept of SpatialAutocorrelation: A Geographer's Perspective. *Geographical Analysis,* Volume 40, pp. 297-309.

Getis, A. & Ord, K. J., 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis,* 24(3), pp. 189-206.

Griffith, D. A., 2009. Modeling spatial autocorrelation in spatial interaction data: empirical evidence from 2002 Germany journey-to-work flows. *Journal of Geographical systems,* 11(2), pp. 117-140.

Guy, C. M., 1991. Spatial interaction modelling in retail planning practice; the need for robust statistical methods. *Environment ond Planning B: Planning one! Donign,* Volume 18, pp. 191-203.

Hiscock, R., Perace, J., Blakely, T. & Witten, K., 2008. Is Neighborhood Access to Health Care Provision Associated with Individual-Level Utilization and Satisfaction?. *Health Service Research,* 43(6), pp. 2183-2200.

Huff, D., 1963. A probabilistic analysis of shopping centre trade areas. *Land Economics,* Volume 39, pp. 81-90.

Huff, D. L., 1964. Defining and estimating a trading area. *Journal of Marketing,* Volume 28, pp. 37-38.

Ikram, S. Z., Hu, Y. & Wang, F., 2015. Disparities in Spatial Accessibility of Pharmacies in Baton Rouge, Louisiana. *Geogr Rev,* Volume 105, pp. 492-510.

Jiang, S., Chin, K. & Tsui, K. L., 2018. A universal deep learning approach for modeling the flow of patients under different severities. *Computer Methods and Programs in Biomedicine,* Volume 154, pp. 191-203.

Kassambara, A., 2018. Machine Learning Essentials: Regression Model Validation. *Practical Guide to R,* Volume 1.

Law, M. R., Dijkstra, A. & Douillard, J. A., 2011. Geographic Accessibility of Community Pharmacies in Ontario. *Healthcare Policy,* 6(3), pp. 36-46.

Lekdee, K. & Ingsrisawang, L., 2013. Generalized Linear Mixed Models with Spatial Random Effects For Spatio-Temporal Data: An Application to Dengue Fever Mapping. *Journal of Mathematics and Statistics,* 9(2), pp. 137-143.

LeSage J.P., S. E., 2016. A Bayesian spatial interaction model variant of the Poisson pseudo-maximum likelihood estimator. In: *Spatial Econometric Interaction Modelling. Advances in Spatial Science (The Regional Science Series).* s.l.:Springer.

LeSage, J. P., 2005. Spatial Econometric Modeling of Origin-Destination flows. In: *Lecture 7.* s.l.:Department of Economics, University of Toledo.

LeSage, J. P., 2013. A Spatial Interaction Model With Spatially Structured Origin and Destination Effects. *Journal of Geographic Systems,* 15(3), pp. 265-289.

LeSage, J. P. & Fisher, M. M., 2008. Spatial econometric methods for modeling origin destination flows. *WU Vienna University of Economics and Business.*

Liaw, K. L. & Bartels, C. A., 1982. Estimation and interpretation of a nonlinear migration model. *Geographical Analysis,* Volume 14, pp. 229-245.

Lo, L., 1991. Spatial structure and spatial interaction: a simulation approach. *Environment and Planning A,* Volume 23, pp. 1279 - 1300.

Martin, D., Gale, C., Cockings, S. & Hartfoot, A., 2018. Origin-destination geodemographics for analysis of travel to work flows. *Computers, Environment and Urban Systems,* Volume 67, pp. 68-79.

Mennis, J. & Guo, D., 2009. Spatial data mining and geographic knowledge discovery—An introduction. *Computers, Environment and Urban Systems,* Volume 33, pp. 403 - 408.

Mohammadian, A. & Kanaroglou, P. S., 2003. *Applications of Spatial Multinomial Logit Model to Transportation Planning,* Lucerne: 10th International Conference on Travel Behaviour Research .

Moody, A. & Mindel, J., 2017. *Health Survey for England 2016,* s.l.: NHS Digital.

Nakaya, T., 2001. Local Spatial Interaction Modelling Based on the Geographically Weighted Regression Approach. *Geojournal,* Volume 53, pp. 347 - 358.

NHS, 2018. *NHS Business Services Authority (NHSBSA).* [Online]
Available at: https://www.nhsbsa.nhs.uk/prescription-data
[Accessed 2018].

NPA, 2009. *Improving Comunication Between community pharmacy and general practice,* s.l.: National Pharmacy Association.

O'Kelly, M. E., 2009. Applied retail location models using spatial interaction tools. *The SAGE Handbook of Spatial Analysis,* pp. 419-442.

ONS, 2013. *Historic and projected mortality data from the period and cohort life tables, 2012-based, UK, 1981–2062,,* Newport: Office for National Statistics.

Oshan, T., 2016. A primer for working with the Spatial Interaction modeling (SpInt) module in the python spatial analysis library (PySAL). *Journal of ERSA: REGION,* 3(2), p. R11–R23.

Patuelli, R. & Arbia, G., 2016. *Spatial Econometric Interaction Modelling.* s.l.:Springer.

Polson, N. G. & Sokolov, V. O., 2017. Deep learning for shorterm traffic flow prediction. *Transportation Research Part C: Emerging Technologies,* Volume 79, pp. 1-17.

Qato, D. M. et al., 2017. The availability of pharmacies in the United States: 2007–2015. *PLOS ONE,* 12(8).

Rodrigue, J. P., 2017. Spatial Interactions and the Gravity Model. In: *The Geography of Transport Systems.* Fourth ed. New York: s.n.

Senior, M. L., 1973. Approaches to Residential Location Modelling 1: Urban Ecological and Spatial Interaction Models (A Review). *Environment and Planning,,* Volume 5, pp. 165-197.

Spark, C. S., 2015. Spatial Regimes and Geographically Weighted Regression in R. *Reggression in R.*

Stopka, T. J. et al., 2013. Factors Associated with Presence of Pharmacies and Pharmacies that Sell Syringes Over-the-Counter in Los Angeles County. *Journal of Urban Health: Bulletin of the New York Academy of Medicine,* 90(6), pp. 1079-1090.

Sun, D., Tsukawa, R. K., Kim, H. & He, Z., 2000. Spatio-temporal interaction with disease mapping. *Statistics in Medicine,* 19(15), pp. 2015-2035.

Tao, R. & Thill, J. C., 2016. Spatial Cluster Detection in Spatial Flow Data. *Geographical Analysis,* 48(4), pp. 1 - 18.

Timofeeva, A. & Tesselkina, K., 2016. Local Spatial Interaction Modelling of Graduate Flows. *AIP Conference Proceedings ,* p. 1772.

Tobler, W., 1987. Experiments in Migration Mapping by Computer. *The American Cartographer,* Volume 14, pp. 155-63.

Tobler, W. R., 1970. A computer movie simulating urban growth in the Detroit region.. *Economic Geography,* Volume 46, pp. 234-240.

Todd, A. et al., 2015. An area-level analysis of accessibility to general practice and community pharmacy services in England by urbanity and social deprivation. *BMJ Open,* Volume 5.

UKEssays, 2013. *Boots ltd Current strategic position.* [Online]
Available at: https://www.ukessays.com/essays/business-strategy/boots-ltd-current-strategic-position.php?vref=1
[Accessed 5 September 2018].

Wilson, A., 1967. A statistical theory of spatial distribution models. *Transportation Research,* 1(3), pp. 256-269.

Wilson, A., 1971. A family of spatial interactin models, and associated developements. *Environmental and Planning ,* Volume 3, pp. 1 - 32.

Wilson, A. G., 1970. Advances and problems in distribution modelling. *Transportation Research,* Volume 4, pp. 1-8.

Yano, K., 1993. Integration of spatial interactin models: towards genral theory of spatial interaction. *Geographical Reports of Tokyo Metropolitan University,* Volume 64A, pp. 367-387.

Yi, H., Jung, H. & Bae, S., 2017. Deep Neural Networks for traffic flow prediction. *IEEE International Conference on Big Data and Smart Computing (BigComp),* pp. 328-331.

APENDICES

Table 1: Description of the catchment analysis variables

Variable	*Description*	
GP_Items_PY	Number of Items prescribed by each GP in year 2017.	
Boots_Items_PY	Number of Items dispensed by each Boots pharmacy in year 2017.	
distance	Euclidian distance in metres between particular GP and Boots pharmacy.	
distance_to_city_centre	Distance to the city centre in metres, particularly to Liverpool Central train station.	
GP_Flag	Main	(GP type based on NHS classification describing a size and purpose.
	Main with Branch	
	Branch	
GP_Partners	Number of GP Practiocioners in a surgery	
GP_nightime_population	Night-time population in each GP 800 catchment. Mean value of each LSOA population estimates.	
GP_ethnicity_proportion	A proportion of the ethnically different population then the white population. Generated as a sum of population estimates for LSOA's in 800m GP catchment.	
Boots_daytime_population	Day-time population in each Boots pharmacy 800 catchment. Mean value of each LSOA population estimates.	
Format	Destination H&B	Large Health and Beauty Store in High Street or Retail Park Location. Customers make infrequent visits but average transaction value and footfall are high.
(Boots main classification developed by using customer shopping habits such as shopping baskets, frequency of visit etc.)	Local Boots H&B	Smaller high street or retail park store. Frequent visits with Smaller average transaction value.
	Pharmacy Format 1 (Higher Affluence)	Pharmacy dominant small store in High Affluence area
	Pharmacy Format 2 (lower Affluence)	Pharmacy dominant small store in Low affluence area
	Convenience Retail	Convenience based store
NHS_Description	Retail park exempt	(Type of NHS contract, based on hours a pharmacist can trade)
	Standard 40hr	
	Extended 100hr	
Competitors_Items_PY	Sum of Items dispensed by the competing pharmacies in 800m catchment of Boots pharmacies in year 2017.	
Boots_competitors_Items_PY	Sum of Items dispensed by the other Boots pharmacies in 800m catchment of Boots pharmacies in year 2017.	
BOOTS_attraction	Retail Attractiveness of the Boots pharmacy. (Equation 10)	
GP_elder_population_proportion	A proportion of the elderly population. Generated as a sum of population estimates for LSOA's in 800m GP catchment.	
BOOTS_ethnicity_proportion	A proportion of the ethnically different population then the white population. Generated as a sum of population estimates for LSOA's in 800m Boots pharmacy catchment.	
BOOTS_nightime_population	Night-time population in each Boots pharmacy 800 catchment. Mean value of each LSOA population estimates.	
BOOTS_elder_population proportion	A proportion of the elderly population. Generated as a sum of population estimates for LSOA's in 800m Boots pharmacy catchment.	
GP_MEAN_imd_score	IMD score of each GP's 800m catchment. Generated as MEAN value of LSOA's inside the catchment.	
GP_no_car_proportion	A proportion of the population without car. Generated as a sum of population estimates for LSOA's in 800m GP catchment.	
BOOTS_MEAN_imd_score	IMD score of each Boots pharmacy 800m catchment. Generated as MEAN value of LSOA's inside the catchment.	
BOOTS_no_car_proportion	A proportion of the population without car. Generated as a sum of population estimates for LSOA's in 800m Boots pharmacy catchment.	

Table 2: Regression coefficients of production constrained model for Set 3; flows above 100 items

Coefficients	*Estimate*	
(Intercept)	-0.1935	***
log(BO_I_PY)	0.8993	***
distance	-0.0005855	***
GP_ODSN83608	1.046	***
GP_ODSN83028	1.038	***
GP_ODSN85002	0.8925	***
GP_ODSN82110	0.8374	***
GP_ODSN84018	0.8266	***
GP_ODSN85006	0.7777	***
GP_ODSN83045	0.7599	***
GP_ODSN84017	0.688	***
GP_ODSN85008	0.6108	***
GP_ODSN85009	0.5414	***
GP_ODSN85023	0.5086	***
GP_ODSN83013	0.4434	***
GP_ODSN83001	0.4361	***
GP_ODSN84025	0.4128	***
GP_ODSN83005	0.3578	***
GP_ODSN83041	0.3542	***
GP_ODSN83009	0.342	***
GP_ODSN85032	0.3387	***
GP_ODSN82073	0.3013	***
GP_ODSN82037	0.2967	***
GP_ODSN84003	0.2963	***
GP_ODSN85044	0.2697	***
GP_ODSN83049	0.2672	***
GP_ODSN84012	0.2523	***
GP_ODSN84618	0.2415	***
GP_ODSN83008	0.2266	***
GP_ODSN85047	0.2225	***
GP_ODSN85007	0.2034	***
GP_ODSN85640	0.1654	***
GP_ODSN83019	0.1564	***
GP_ODSN85046	0.1074	***
GP_ODSN85058	0.07431	***
GP_ODSN83002	0.02086	
GP_ODSN84026	0.009085	
GP_ODSN84020	0.00772	
GP_ODSN83010	0.005442	
GP_ODSN84027	-0.02276	
GP_ODSN85005	-0.04	*
GP_ODSN82026	-0.05896	***
GP_ODSN82009	-0.06396	***
GP_ODSN84013	-0.08285	***
GP_ODSN82053	-0.09444	***
GP_ODSN82018	-0.09867	***
GP_ODSN83032	-0.1081	***
GP_ODSN82116	-0.1405	***
GP_ODSN83024	-0.1412	***
GP_ODSN85034	-0.1583	***
GP_ODSN84006	-0.1857	***
GP_ODSN85022	-0.1886	***
GP_ODSN83031	-0.1917	***
GP_ODSN82094	-0.2171	***
GP_ODSN82113	-0.2307	***
GP_ODSN82079	-0.2398	***
GP_ODSN84014	-0.2408	***
GP_ODSN82617	-0.2458	***
GP_ODSN85001	-0.2503	***
GP_ODSN84024	-0.2543	***
GP_ODSN82062	-0.2609	***
GP_ODSN84021	-0.2612	***
GP_ODSN83609	-0.2711	***
GP_ODSN84035	-0.2795	***
GP_ODSN84615	-0.3001	***
GP_ODSN84005	-0.3142	***
GP_ODSN84028	-0.3193	***
GP_ODSN85020	-0.3468	***
GP_ODSN85003	-0.363	***
GP_ODSY02510	-0.3667	***
GP_ODSN83035	-0.4106	***
GP_ODSN82084	-0.4129	***
GP_ODSN85051	-0.4146	***
GP_ODSN85013	-0.4252	***
GP_ODSN82067	-0.4575	***
GP_ODSN85625	-0.4591	***
GP_ODSN82092	-0.485	***
GP_ODSN82074	-0.495	***
GP_ODSN84001	-0.5188	***
GP_ODSN83055	-0.5236	***
GP_ODSN82100	-0.5513	***
GP_ODSN83603	-0.5517	***
GP_ODSN85057	-0.59	***
GP_ODSN83014	-0.599	***
GP_ODSN82655	-0.602	***
GP_ODSN85012	-0.6073	***
GP_ODSN83021	-0.6361	***
GP_ODSN82671	-0.6466	***
GP_ODSN83012	-0.6498	***

GP_ODSN85643	-0.6638	***	***GP_ODSN85059***	-1.122	***
GP_ODSN85027	-0.6679	***	***GP_ODSN82093***	-1.129	***
GP_ODSN85016	-0.6711	***	***GP_ODSN82641***	-1.136	***
GP_ODSN84613	-0.6763	***	***GP_ODSN82076***	-1.138	***
GP_ODSN82041	-0.6813	***	***GP_ODSN84605***	-1.144	***
GP_ODSN82633	-0.6851	***	***GP_ODSN82054***	-1.158	***
GP_ODSN82117	-0.6874	***	***GP_ODSN84004***	-1.17	***
GP_ODSN82059	-0.6907	***	***GP_ODSN85037***	-1.194	***
GP_ODSN85048	-0.7065	***	***GP_ODSN82650***	-1.207	***
GP_ODSN85017	-0.7247	***	***GP_ODSN83018***	-1.208	***
GP_ODSN84037	-0.729	***	***GP_ODSN83628***	-1.223	***
GP_ODSN82035	-0.735	***	***GP_ODSN82663***	-1.227	***
GP_ODSN82108	-0.7413	***	***GP_ODSN83635***	-1.232	***
GP_ODSN82034	-0.7552	***	***GP_ODSN82082***	-1.247	***
GP_ODSN82070	-0.7612	***	***GP_ODSN83611***	-1.248	***
GP_ODSN85018	-0.7687	***	***GP_ODSN85028***	-1.252	***
GP_ODSN82050	-0.777	***	***GP_ODSN85031***	-1.266	***
GP_ODSN82066	-0.7849	***	***GP_ODSN83633***	-1.272	***
GP_ODSN82024	-0.7853	***	***GP_ODSN85024***	-1.273	***
GP_ODSN84011	-0.7925	***	***GP_ODSN85616***	-1.289	***
GP_ODSN85620	-0.8061	***	***GP_ODSN82676***	-1.298	***
GP_ODSN83601	-0.8325	***	***GP_ODSN82052***	-1.304	***
GP_ODSN84015	-0.8445	***	***GP_ODSN84008***	-1.308	***
GP_ODSN85038	-0.8471	***	***GP_ODSN82033***	-1.313	***
GP_ODSN83033	-0.8503	***	***GP_ODSN83020***	-1.315	***
GP_ODSN84010	-0.8588	***	***GP_ODSN82058***	-1.319	***
GP_ODSN85054	-0.8982	***	***GP_ODSN82022***	-1.321	***
GP_ODSY00446	-0.8985	***	***GP_ODSN83017***	-1.323	***
GP_ODSN83030	-0.9086	***	***GP_ODSN84016***	-1.342	***
GP_ODSN84007	-0.9258	***	***GP_ODSN84625***	-1.347	***
GP_ODSN83022	-0.9268	***	***GP_ODSN85633***	-1.349	***
GP_ODSN85014	-0.9351	***	***GP_ODSN82669***	-1.353	***
GP_ODSN83053	-0.9516	***	***GP_ODSN82109***	-1.36	***
GP_ODSN82049	-0.9778	***	***GP_ODSN82046***	-1.368	***
GP_ODSN82019	-0.9808	***	***GP_ODSN82115***	-1.382	***
GP_ODSN82014	-0.996	***	***GP_ODSN84034***	-1.382	***
GP_ODSN85040	-1.001	***	***GP_ODSN82664***	-1.401	***
GP_ODSN84041	-1.006	***	***GP_ODSN83614***	-1.405	***
GP_ODSN84611	-1.018	***	***GP_ODSN82621***	-1.406	***
GP_ODSN83054	-1.025	***	***GP_ODSN84621***	-1.415	***
GP_ODSN82087	-1.06	***	***GP_ODSN84617***	-1.43	***
GP_ODSN84627	-1.065	***	***GP_ODSN82662***	-1.436	***
GP_ODSN84023	-1.085	***	***GP_ODSN82083***	-1.437	***
GP_ODSN84043	-1.085	***	***GP_ODSN82097***	-1.476	***
GP_ODSN84029	-1.092	***	***GP_ODSN83624***	-1.476	***
GP_ODSN84038	-1.094	***	***GP_ODSN82077***	-1.482	***
GP_ODSN85019	-1.113	***	***GP_ODSN82670***	-1.492	***
GP_ODSY02610	-1.116	***	***GP_ODSN82004***	-1.498	***

GP_ODSN82645	-1.51	***
GP_ODSN84614	-1.512	***
GP_ODSN85617	-1.561	***
GP_ODSN82048	-1.571	***
GP_ODSN82078	-1.585	***
GP_ODSN82104	-1.586	***
GP_ODSN85053	-1.594	***
GP_ODSN83006	-1.596	***
GP_ODSN82090	-1.597	***
GP_ODSN83023	-1.61	***
GP_ODSN83050	-1.632	***
GP_ODSN82646	-1.634	***
GP_ODSN82106	-1.645	***
GP_ODSN83027	-1.657	***
GP_ODSN82003	-1.661	***
GP_ODSN83003	-1.681	***
GP_ODSN82086	-1.696	***
GP_ODSN84019	-1.706	***
GP_ODSN82011	-1.731	***
GP_ODSN85021	-1.746	***
GP_ODSN85648	-1.772	***
GP_ODSN85052	-1.773	***
GP_ODSN83605	-1.787	***
GP_ODSN85025	-1.796	***
GP_ODSY02162	-1.796	***
GP_ODSN82103	-1.817	***
GP_ODSN82060	-1.825	***
GP_ODSN82665	-1.849	***
GP_ODSN82101	-1.863	***
GP_ODSN85015	-1.887	***
GP_ODSN83621	-1.9	***
GP_ODSN83007	-1.907	***
GP_ODSN82091	-1.951	***
GP_ODSN85634	-1.952	***
GP_ODSN83025	-1.96	***
GP_ODSN82099	-1.965	***
GP_ODSN83043	-1.968	***
GP_ODSN82039	-1.999	***
GP_ODSN83637	-2.026	***
GP_ODSN85629	-2.028	***
GP_ODSN84002	-2.111	***
GP_ODSN82065	-2.115	***
GP_ODSN83015	-2.139	***
GP_ODSN82036	-2.147	***
GP_ODSN82648	-2.155	***
GP_ODSN82623	-2.156	***
GP_ODSN83047	-2.194	***
GP_ODSN82647	-2.227	***
GP_ODSN84626	-2.253	***
GP_ODSN84630	-2.269	***
GP_ODSN82002	-2.273	***
GP_ODSN83060	-2.306	***
GP_ODSN82651	-2.319	***
GP_ODSN82081	-2.324	***
GP_ODSN82678	-2.329	***
GP_ODSN82089	-2.356	***
GP_ODSN83026	-2.394	***
GP_ODSN82107	-2.432	***
GP_ODSN83610	-2.438	***
GP_ODSN84036	-2.469	***
GP_ODSN82095	-2.536	***
GP_ODSN83620	-2.622	***
GP_ODSN82668	-2.696	***
GP_ODSN83604	-2.708	***
GP_ODSY00110	-3.021	***
GP_ODSN85619	-3.026	***
GP_ODSN83622	-3.135	***
GP_ODSN82051	-3.252	***
GP_ODSY02511	-3.488	***
GP_ODSN83619	-3.734	***
GP_ODSN83638	-4.451	***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'*

Table 3: Regression coefficients of attraction-constrained model for Set 3: Flows above 100 items

Coefficients:	***Estimate***	
(Intercept)	1.042	***
log(GP_Items_PY)	0.6911	***
distance	-0.00065	***
Ph_ODSFV731	1.733	***
Ph_ODSFKM99	1.573	***
Ph_ODSFJV69	1.262	***
Ph_ODSFD635	1.143	***
Ph_ODSFGN97	1.035	***
Ph_ODSFV122	1.001	***
Ph_ODSFC972	0.9765	***
Ph_ODSFPX42	0.9601	***
Ph_ODSFWG19	0.9593	***
Ph_ODSFNL49	0.9338	***
Ph_ODSFN264	0.915	***
Ph_ODSFHK81	0.85	***
Ph_ODSFGX53	0.8243	***
Ph_ODSFK207	0.8015	***
Ph_ODSFH796	0.7985	***
Ph_ODSFN923	0.7965	***
Ph_ODSFM378	0.5996	***
Ph_ODSFQT68	0.5912	***
Ph_ODSFLM92	0.5738	***
Ph_ODSFYL72	0.5646	***
Ph_ODSFY215	0.5454	***
Ph_ODSFAN25	0.5405	***
Ph_ODSFJV77	0.4956	***
Ph_ODSFLP52	0.4246	***
Ph_ODSFCT19	0.3986	***
Ph_ODSFC368	0.3783	***
Ph_ODSFXX14	0.3251	***
Ph_ODSFF987	0.2899	***
Ph_ODSFDV38	0.2747	***
Ph_ODSFAQ22	0.1743	***
Ph_ODSFL617	0.1368	***
Ph_ODSFJJ30	0.1105	***
Ph_ODSFEV79	0.1078	***
Ph_ODSFEN99	0.1048	***
Ph_ODSFF015	0.1041	***
Ph_ODSFG049	0.09955	***
Ph_ODSFQM51	0.08386	***
Ph_ODSFVR87	0.06122	***
Ph_ODSFK182	0.03563	***
Ph_ODSFP502	0.01052	.
Ph_ODSFVF47	0.008785	
Ph_ODSFHM91	-0.02282	***
Ph_ODSFH645	-0.04181	***
Ph_ODSFVN86	-0.05715	***
Ph_ODSFMR00	-0.1062	***
Ph_ODSFM249	-0.1348	***
Ph_ODSFQX26	-0.1717	***
Ph_ODSFMD34	-0.2049	***
Ph_ODSFR460	-0.213	***
Ph_ODSFLV18	-0.2222	***
Ph_ODSFYV68	-0.2254	***
Ph_ODSFRK88	-0.2961	***
Ph_ODSFK272	-0.359	***
Ph_ODSFL048	-0.4024	***
Ph_ODSFF360	-0.5783	***
Ph_ODSFF769	-0.6417	***
Ph_ODSFXV39	-0.8695	***
Ph_ODSFT423	-1.039	***
Ph_ODSFFM98	-1.081	***
Ph_ODSFMC89	-1.086	***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'*

Table 4: Regression table of doubly-constrained model of flows over 100 items

Coefficients	Estimate	Significance
(Intercept)	7.974	***
distance	-0.0007381	***
GP_ODSN83608	3.509	***
GP_ODSN82110	2.692	***
GP_ODSN82037	2.255	***
GP_ODSN84025	2.1	***
GP_ODSN83009	2.071	***
GP_ODSN84003	1.949	***
GP_ODSN84020	1.936	***
GP_ODSN84027	1.898	***
GP_ODSN82053	1.883	***
GP_ODSN83055	1.812	***
GP_ODSN84035	1.799	***
GP_ODSN84028	1.791	***
GP_ODSN84026	1.788	***
GP_ODSN84615	1.788	***
GP_ODSN83032	1.71	***
GP_ODSN82018	1.556	***
GP_ODSN84001	1.537	***
GP_ODSN83033	1.512	***
GP_ODSN85640	1.512	***
GP_ODSN85002	1.293	***
GP_ODSN84015	1.261	***
GP_ODSN82655	1.236	***
GP_ODSN84011	1.221	***
GP_ODSN85023	1.173	***
GP_ODSN82073	1.136	***
GP_ODSN85009	1.134	***
GP_ODSN83601	1.118	***
GP_ODSN85006	1.103	***
GP_ODSN84029	1.102	***
GP_ODSN85044	1.098	***
GP_ODSN84007	1.062	***
GP_ODSN84038	1.045	***
GP_ODSN85008	1.039	***
GP_ODSN84041	1.028	***
GP_ODSN82049	1.019	***
GP_ODSN84004	1.011	***
GP_ODSN84023	1.009	***
GP_ODSN84043	0.9843	***
GP_ODSN84605	0.9559	***
GP_ODSN83045	0.9188	***
GP_ODSN84010	0.906	***
GP_ODSY00446	0.8597	***
GP_ODSN85032	0.8571	***
GP_ODSN84627	0.846	***
GP_ODSN82113	0.841	***
GP_ODSN82026	0.8308	***
GP_ODSN84016	0.8088	***
GP_ODSN84034	0.7466	***
GP_ODSN82079	0.6729	***
GP_ODSN85046	0.6698	***
GP_ODSN82067	0.6431	***
GP_ODSN83028	0.6113	***
GP_ODSN82062	0.6005	***
GP_ODSN82097	0.5753	***
GP_ODSN84621	0.5753	***
GP_ODSN82019	0.5235	***
GP_ODSN82087	0.5142	***
GP_ODSN85022	0.5112	***
GP_ODSN83005	0.474	***
GP_ODSN82009	0.47	***
GP_ODSN82048	0.4604	***
GP_ODSN84019	0.445	***
GP_ODSN82074	0.4211	***
GP_ODSN85020	0.3903	***
GP_ODSN85625	0.3862	***
GP_ODSN85034	0.3846	***
GP_ODSN82052	0.3835	***
GP_ODSN82070	0.382	***
GP_ODSN82671	0.3674	***
GP_ODSN82617	0.3192	***
GP_ODSN82059	0.3017	***
GP_ODSN85012	0.2992	***
GP_ODSN82083	0.2893	***
GP_ODSN82117	0.2382	***
GP_ODSN85013	0.2095	***
GP_ODSN83041	0.1996	***
GP_ODSN82108	0.1465	***
GP_ODSN82663	0.1437	***
GP_ODSN85051	0.1417	***
GP_ODSN83002	0.1312	***
GP_ODSN85620	0.1177	***
GP_ODSN85001	0.113	***
GP_ODSN82024	0.1116	***
GP_ODSN83049	0.1098	**
GP_ODSN85027	0.107	***
GP_ODSN85017	0.1016	***

GP_ODSN82077	0.09742	***	***GP_ODSN82623***	-0.4533	***
GP_ODSN82041	0.07692	***	***GP_ODSN82646***	-0.481	***
GP_ODSN85047	0.0702	**	***GP_ODSN82645***	-0.4908	***
GP_ODSN82078	0.06972	**	***GP_ODSN82647***	-0.5114	***
GP_ODSN85643	0.04528	.	***GP_ODSN85037***	-0.5223	***
GP_ODSN85018	0.0374		***GP_ODSN82676***	-0.5371	***
GP_ODSN82076	0.02764		***GP_ODSN82011***	-0.5397	***
GP_ODSN82086	0.02314		***GP_ODSN82678***	-0.5413	***
GP_ODSN82633	0.001606		***GP_ODSN85616***	-0.5505	***
GP_ODSN84002	-0.01078		***GP_ODSN83008***	-0.5676	***
GP_ODSN85016	-0.03734		***GP_ODSN85007***	-0.5785	***
GP_ODSN85038	-0.05648	*	***GP_ODSN82115***	-0.5788	***
GP_ODSN82641	-0.07838	***	***GP_ODSN85031***	-0.5939	***
GP_ODSN85005	-0.07884	**	***GP_ODSN82099***	-0.6517	***
GP_ODSN85048	-0.08424	***	***GP_ODSN85028***	-0.6646	***
GP_ODSN85057	-0.08851	***	***GP_ODSN83030***	-0.6847	***
GP_ODSN82100	-0.08889	***	***GP_ODSN82060***	-0.7349	***
GP_ODSN82014	-0.1051	***	***GP_ODSN85053***	-0.7359	***
GP_ODSN83019	-0.1232	***	***GP_ODSN82104***	-0.7444	***
GP_ODSN82094	-0.13	***	***GP_ODSN82664***	-0.7463	***
GP_ODSN82054	-0.1359	***	***GP_ODSN83014***	-0.7545	***
GP_ODSN82033	-0.1378	***	***GP_ODSN85058***	-0.7698	***
GP_ODSN82092	-0.1454	***	***GP_ODSN83010***	-0.7779	***
GP_ODSN84630	-0.1577	***	***GP_ODSN82091***	-0.7817	***
GP_ODSN85014	-0.1856	***	***GP_ODSN83013***	-0.8114	***
GP_ODSN82082	-0.186	***	***GP_ODSN83022***	-0.8355	***
GP_ODSN83001	-0.255	***	***GP_ODSN82066***	-0.85	***
GP_ODSN82022	-0.2588	***	***GP_ODSN85052***	-0.8738	***
GP_ODSN82648	-0.2723	***	***GP_ODSN85633***	-0.8783	***
GP_ODSN82046	-0.3156	***	***GP_ODSN82034***	-0.9122	***
GP_ODSN82058	-0.3218	***	***GP_ODSN82668***	-0.9214	***
GP_ODSN82670	-0.3239	***	***GP_ODSN82050***	-0.9485	***
GP_ODSN85059	-0.3257	***	***GP_ODSN82116***	-0.957	***
GP_ODSN82669	-0.3336	***	***GP_ODSN85648***	-0.9595	***
GP_ODSN82662	-0.3465	***	***GP_ODSN85021***	-0.9721	***
GP_ODSN82101	-0.3635	***	***GP_ODSN85025***	-0.9729	***
GP_ODSN85054	-0.375	***	***GP_ODSN82065***	-0.9956	***
GP_ODSN82093	-0.4071	***	***GP_ODSN84018***	-1	***
GP_ODSN82090	-0.4093	***	***GP_ODSN82665***	-1.002	***
GP_ODSN82103	-0.4133	***	***GP_ODSN83628***	-1.012	***
GP_ODSN82035	-0.4216	***	***GP_ODSY02162***	-1.116	***
GP_ODSN82084	-0.4317	***	***GP_ODSN85629***	-1.151	***
GP_ODSN85019	-0.4381	***	***GP_ODSN85634***	-1.175	***
GP_ODSN85003	-0.4459	***	***GP_ODSN85015***	-1.176	***
GP_ODSN85024	-0.4496	***	***GP_ODSN83018***	-1.185	***
GP_ODSN85040	-0.4507	***	***GP_ODSN83020***	-1.221	***

GP_ODSN83024	-1.224	***	***GP_ODSN83610***	-2.767	***
GP_ODSN83021	-1.231	***	***GP_ODSN83007***	-2.838	***
GP_ODSN82089	-1.232	***	***GP_ODSN83637***	-3.008	***
GP_ODSN83035	-1.238	***	***GP_ODSN84626***	-3.13	***
GP_ODSN82051	-1.245	***	***GP_ODSN83060***	-3.316	***
GP_ODSN82651	-1.262	***	***GP_ODSN83026***	-3.358	***
GP_ODSN82039	-1.293	***	***GP_ODSN83622***	-3.445	***
GP_ODSN82003	-1.331	***	***GP_ODSY00110***	-3.607	***
GP_ODSY02510	-1.339	***	***GP_ODSN83620***	-3.678	***
GP_ODSN85617	-1.341	***	***GP_ODSN83619***	-3.697	***
GP_ODSN83609	-1.387	***	***GP_ODSN83604***	-3.718	***
GP_ODSN83031	-1.447	***	***GP_ODSN84036***	-4.07	***
GP_ODSN82095	-1.476	***	***GP_ODSY02511***	-4.586	***
GP_ODSN82650	-1.519	***	***GP_ODSN83638***	-5.517	***
GP_ODSN82081	-1.564	***	***GP_ODSN84012***	-5.706	***
GP_ODSN84618	-1.581	***	***GP_ODSN84017***	-6.064	***
GP_ODSN82107	-1.631	***	***GP_ODSN84014***	-6.474	***
GP_ODSN83012	-1.631	***	***GP_ODSN84013***	-6.903	***
GP_ODSN83603	-1.643	***	***GP_ODSN84024***	-7.126	***
GP_ODSN82002	-1.696	***	***GP_ODSN84005***	-7.187	***
GP_ODSN83633	-1.699	***	***GP_ODSN84021***	-7.203	***
GP_ODSN82106	-1.745	***	***GP_ODSN84613***	-7.266	***
GP_ODSN83053	-1.8	***	***GP_ODSN84037***	-7.346	***
GP_ODSN83054	-1.869	***	***GP_ODSN84611***	-7.713	***
GP_ODSN84006	-2.012	***	***GP_ODSY02610***	-8.061	***
GP_ODSN83635	-2.08	***	***GP_ODSN84008***	-8.074	***
GP_ODSN82036	-2.081	***	***GP_ODSN84614***	-8.11	***
GP_ODSN83025	-2.092	***	***GP_ODSN84625***	-8.118	***
GP_ODSN83624	-2.117	***	***GP_ODSN84617***	-8.167	***
GP_ODSN83614	-2.143	***	***Ph_ODSFD635***	9.82	***
GP_ODSN82004	-2.153	***	***Ph_ODSFVN86***	8.771	***
GP_ODSN83621	-2.2	***	***Ph_ODSFR460***	8.634	***
GP_ODSN82109	-2.215	***	***Ph_ODSFDV38***	8.611	***
GP_ODSN82621	-2.278	***	***Ph_ODSFKM99***	4.201	***
GP_ODSN85619	-2.295	***	***Ph_ODSFJV69***	4.12	***
GP_ODSN83017	-2.37	***	***Ph_ODSFV731***	3.922	***
GP_ODSN83611	-2.372	***	***Ph_ODSFN264***	3.831	***
GP_ODSN83050	-2.413	***	***Ph_ODSFPX42***	3.793	***
GP_ODSN83015	-2.466	***	***Ph_ODSFAN25***	3.416	***
GP_ODSN83023	-2.594	***	***Ph_ODSFK182***	3.27	***
GP_ODSN83027	-2.622	***	***Ph_ODSFGX53***	3.22	***
GP_ODSN83006	-2.624	***	***Ph_ODSFWG19***	2.974	***
GP_ODSN83003	-2.629	***	***Ph_ODSFV122***	2.874	***
GP_ODSN83047	-2.638	***	***Ph_ODSFF360***	2.818	***
GP_ODSN83043	-2.675	***	***Ph_ODSFC972***	2.322	***
GP_ODSN83605	-2.752	***	***Ph_ODSFNL49***	2.136	***

Ph_ODSFAQ22	2.096	***
Ph_ODSFLM92	1.906	***
Ph_ODSFC368	1.85	***
Ph_ODSFLP52	1.614	***
Ph_ODSFCT19	1.605	***
Ph_ODSFN923	1.598	***
Ph_ODSFY215	1.581	***
Ph_ODSFQT68	1.54	***
Ph_ODSFF015	1.536	***
Ph_ODSFQM51	1.513	***
Ph_ODSFM378	1.502	***
Ph_ODSFYL72	1.472	***
Ph_ODSFF987	1.459	***
Ph_ODSFL617	1.378	***
Ph_ODSFXX14	1.351	***
Ph_ODSFG049	1.339	***
Ph_ODSFM249	1.251	***
Ph_ODSFMR00	1.21	***
Ph_ODSFJV77	1.182	***
Ph_ODSFLV18	1.128	***
Ph_ODSFQX26	1.08	***
Ph_ODSFK272	1.068	***
Ph_ODSFVR87	1.065	***
Ph_ODSFH645	1.048	***
Ph_ODSFYV68	0.9912	***
Ph_ODSFMD34	0.9766	***
Ph_ODSFL048	0.8211	***
Ph_ODSFF769	0.7671	***
Ph_ODSFEV79	0.7241	***
Ph_ODSFGN97	0.6708	***
Ph_ODSFHK81	0.6592	***
Ph_ODSFXV39	0.5319	***
Ph_ODSFK207	0.4793	***
Ph_ODSFMC89	0.41	***
Ph_ODSFH796	0.4063	***
Ph_ODSFFM98	0.2077	***
Ph_ODSFEN99	0.1202	***
Ph_ODSFVF47	0.08484	***
Ph_ODSFT423	0.07707	***
Ph_ODSFP502	0.07159	***
Ph_ODSFJJ30	0.04597	***
Ph_ODSFRK88	0.01338	*
Ph_ODSFHM91	-0.12	***