HEALTH BENEFITS OF REDUCING AIR TRAFFIC NOISE: EVIDENCE FROM CHANGES IN FLIGHT PATTERNS

Silvia Beghelli
King's College London

Augustin de Coulon
King's College London, CVER(LSE) and IZA

Mary O'Mahony
King's College London, NIESR and ZEW
30 Aldwych, WC2B 4BG London, UK
Tel: +44(207)8484425
Email: Mary.OMahony@kcl.ac.uk

Abstract

This paper investigates health externalities generated by air transportation. As a source of exogenous variation, we use an unannounced five-month trial that changed early morning patterns of aircraft landings at London Heathrow airport. Our measure of health is prescribed medication usage for conditions known to be aggravated by noise. Compared to the control regions, we observe a significant and substantial decrease in prescribed drugs for respiratory and central nervous system conditions in the areas subjected to reduction in air traffic. Our findings suggest therefore a causal influence of air traffic noise on health conditions.

JEL Classification: I10, Q5, R4.

Keywords: noise pollution, environmental effects, health.

Word count: 9,581

I. INTRODUCTION

Pollution has well known economic consequences, affecting the health status of workers and so their productivity and well-being. The study of the relationship between environmental exposure and its adverse health effects is a well-documented field in environmental, epidemiological and medical research. This paper contributes to the still limited, but growing, field of studies using exogenous variation to investigate the causal effect of transportation services on health (see Cesur et al. 2017, Deryugina et al. 2016 for recent examples, and Graff Zivin & Neidell 2013 for a review). We present new evidence on the health impact of airports as major sources of pollution (Wolfe et al. 2017, Schlenker & Walker 2016), specifically focusing on noise pollution. We consider regions exposed to a change in patterns of plane landings around a global aviation hub, located within a large metropolitan area, London Heathrow airport. We make use of a trial implemented over five months (between November 2012 and March 2013) that redirected approaching aircraft to reduce early morning noise in designated areas. As a health indicator we use drugs prescribed by medical doctors. We focus on three broad types of diseases that, as suggested by the medical literature, are aggravated by noise pollution: central nervous, respiratory and cardiovascular conditions.

Our main contribution is establishing new and concrete results linking air transportation noise to medical conditions in a causal framework. We do so by exploiting unique context and data. First, the nature of the trial surmounts avoidance behaviours - people may rationally avoid places exposed to increased pollution - that plague earlier literature. This trial had the critical and unique feature of occurring at daybreak, between 4.30am and 6.00am, when targeted residents are most likely to be at home and therefore exposed to the full impact of the changed flight paths. Second, by using data on medicines prescribed by doctors to their patients, we can assess diagnosed health conditions, rather than relying on self-reported health conditions. Finally by quantifying the health impact of air traffic noise caused by airports, we contribute to recent literature trying to credibly estimate the impacts of transport congestion locations on health outcomes using natural experiments. This literature consid-

ered air pollution generated by airports (Schlenker & Walker 2016), ports (Moretti & Neidell 2011), and tollbooths and traffic congestion (Currie & Walker 2011) as well as a paper on noise pollution near Zurich airport (Boes et al. 2013).

Our main results are that during the trial we observe a decrease in monthly prescription expenditure on central nervous system and respiratory medication by 5.8% and 3.3% respectively[1]. These results are more pronounced for areas most affected by the changing flight patterns. We test the main results by checking whether similar prescription changes happened for other diseases known to be unrelated to noise pollution (infections and musculoskeletal conditions); we cannot detect any significant changes over the same period. The results are also robust to changes in the time periods covered and choice of control group. Our results therefore suggest a causal link between air traffic noise and health, which has financial implications for health spending. A permanent reduction in early morning air traffic are estimated to save around £5 millions per year from prescribed medicines in respiratory and central nervous conditions, in the areas most affected by reductions in air traffic.

This paper is structured as follows. The following section gives background information on airports, noise pollution and health. Section III describes the institutional setting and the empirical strategy. Section IV describes the data and Section V presents and discusses the results. Section VI concludes.

II. AIRPORTS, NOISE AND HEALTH

Communities located near major airports such as Heathrow suffer from exposure to noise pollution, which can have negative impacts on their health. In the UK the Civil Aviation Authority (CAA) on behalf of the Department for Transport produces noise contours maps to estimate the size of the areas subject to different noise levels (Lee et al. 2014). As a standard, noise contours are plotted at levels

1. In Britain medical prescriptions are subsidised by the National Health Service (NHS) and arise from visits to physicians, known as general practitioner (GP) doctors. This is in contrast to the reimbursement systems that occur in countries with medical insurance schemes.

from 57 to 72 dB[2], in 3 dB steps. Additional steps from 48 to 57 dB are added for night contours due to the higher sensitivity of people during sleep hours. Traditionally, the 57 dB level represents the starting point of significant community annoyance. For Heathrow airport, Lee et al. (2014) calculated that in 2013 about 266,000 and 421,000 people were exposed to 57 $L_{Aeq,16hour}$ during the day and 48 $L_{Aeq,\ 8hour}$ during the night respectively. The large number of residents affected is due to the proximity of Heathrow to a highly urbanised area. In fact, Heathrow lies within the boundaries of Greater London (an unusual location for a major international hub)[3].

There is strong evidence that noise, defined as undesirable sound, impinges on human health. Among its adverse effects, we focus on those non-auditory ones - i.e. those health effects other than tinnitus and hearing loss, triggered by environmental noise. In their recent review, Basner et al. (2014) identified four main outcomes from excessive noise: sleep disturbance, annoyance, cognitive impairment and cardiovascular disease. People react to various levels of noise when it interferes with sleep or daily activities. They experience a range of effects of varying severity, from exhaustion and stress-related symptoms to anger and displeasure. The human body can respond through direct and indirect pathways to acute exposure to noise. The latter refers to the path from perceived nuisance to emotional stress reactions. The direct pathway consists of the autonomic physiological stress triggered by the interaction between the central auditory system and the central nervous system. Even at low noise levels, this is considered to be the prevalent mechanism in sleeping individuals (Basner et al. 2014). Observations on chronically exposed populations show an effect on the metabolism and the deterioration of the cardiovascular system (Basner et al. 2014). Sleep disturbance is regarded as the most harmful effect of environmental noise exposure. Occasional incidents as low as dB 33 L_{Amax} at night can induce

2. Noise exposure is measured in decibels (dB), a logarithm scale that ranks noise pressure levels. When noise varies over time, the $L_{Aeq,T}$ is the equivalent average continuous sound which would contain the same sound energy as the time varying noise for a given period T. When noise has instantaneous effects, such as sleep disturbance due to aircraft, it is better measured as a maximum value during the time period (L_{Amax}).

3. The initial location was chosen for military purpose during WWII, without foreseeing its expansion into one of the world's top four busiest civilian airports (from "The History of Heathrow", The Independent, 1st March 2011.)

various physiological reactions during sleep, such as tachycardia, body movements and awakenings (Basner et al. 2014). There is conflicting evidence on the size of these effects, which vary according to whether the study considers elderly, children or people with existing conditions (van Kamp & Davies 2013).

Noise source is a fundamental contributor in reaction to noise. Different sources hold different acoustic characteristics: frequency, sound level, duration, intensity and psychoacoustic measures. For instance, at the same average night noise level, aircraft noise is found to trigger a higher level of annoyance than other transportation noise (European Commission, 2004).

Studies on noise effects date back to the 1970s (Ando et al. 1975). Initially, laboratory settings were promoted, followed by field experiments with a focus on airports (Cohen et al. 1981, Chen & Chen 1993, Evans et al. 1995). These found harmful effects of noise on cognitive ability and on blood pressure. Since then qualitative research played an increasingly important role in documenting individuals' reaction to noise. There are many epidemiological studies drawing on large administrative sources of health outcomes to investigate the effects of noise on health. Examples include (Tzivian et al. 2015) who reviewed studies on the mental health effects of exposure to noise pollution and reported a positive association with anxiety, depression and impaired activities of daily living, among other outcomes. Hansell et al. (2013) focused on the Heathrow airport region specifically. They found that exposure to higher noise levels increased mortality and the prevalence of strokes, coronary heart disease and cardiovascular disease for both hospital admissions and mortality.

Although these cross-sectional studies control for some of the confounding factors that could be associated with the relevant outcomes, such as socio-economic status and individual overall health conditions, they do not unequivocally determine causation between environmental factors and health. For example, they assume that exposure to noise happens mainly at the individual's home address. However, a large proportion of the population spend most of their day outside their home, thus raising problems of exposure bias. In response, economists have adopted quasi-experimental techniques to tackle some of these issues (Graff Zivin & Neidell 2013).

A few recent papers exploit exogenous shocks to emissions to estimate the related health effects. However, these typically focus on air pollution levels (Currie & Walker 2011, Beatty & Shimshack 2011, Schlenker & Walker 2016, Halliday et al. 2018 among others) rather than noise. An exception is Boes et al. (2013) who attempt to measure exposure to noise around Zurich airport by using individual fixed-effects and a change in flight routes over a number of years. They found that daytime exposure to an increase in aircraft noise significantly affects self-reported health problems. Our paper uses actual medicines prescribed for conditions aggravated by noise during sensitive sleeping hours, in a framework conducing to a causal interpretation by comparing exposures to changed flight patterns between treated and control groups.

III. METHODOLOGY

III.a. Identification strategy

In order to address its noise externalities, Heathrow airport explores ways to reduce these through a number of adjustments and measures. For instance, it encourages the use of quieter planes especially during sensitive hours, promotes quieter operating procedures, and working with local communities it provides individual home insulation (Heathrow Airport Limited 2013). The Early Morning Arrivals Trial (EMAT) in 2012 and 2013 was introduced to provide noise respite to specific communities affected by landings at Heathrow airport.

Our analysis focuses on this intervention. During five months, from 5^{th} November 2012 to 31^{st} March 2013, Heathrow airport ran the trial in collaboration with the noise pressure group HACAN (Heathrow Association for the Control of Aircraft Noise), British Airways and NATS (formerly National Air Traffic Services). The main feature of the trial was the identification of four pairs of exclusion zones (two to the east and two to the west of Heathrow), which were designed to be free of aircraft movements during the night and early morning in alternate weeks for the duration of the trial, redirecting the night flights to other areas. The trial implemented a weekly switch between these two sets of exclusion

zones, which we term 'odd' and 'even' weeks. A commissioned report (Tucker et al. 2013) evaluated the outcome of the trial but did not provide the exact flights paths for affected areas. We therefore rely on graphics produced by the report to illustrate the distribution of flights across the affected zones - shown in Appendix A.

Night quota restrictions reduce landings at Heathrow between 11:30 pm and 6:00 am. However airlines, responding to travellers preferences for early morning landings, allocate nearly all those landing slots between 4:30 am to 6:00 am. This pattern translates into one aircraft landing every four to ten minutes during those crucial 90 minutes when sleep is likely to be disrupted. In addition these early morning landings are typically transcontinental large bodied jets which are noisier than the average aircraft landing at other times of the day.

Our data on prescriptions are available on a monthly basis only, therefore we use early morning flights per month. This has the advantage of picking up most of the prescription changes in any one month of the trial, as patients often consult their doctors with a delay. The nature of the trial means that residents will have experienced reduced or no noise in two weeks in a month but may have increased noise in alternate weeks. There are a number of reasons why this exposure does not cancel out in aggregate, allowing us to identify the impact of the trial on prescriptions.

The first is climate related. Aircraft have to land into the wind when its speed exceeds 5 knots; in South East England 70% of the year the wind direction is west to east. This little known pattern implies that, as opposed to a more regular alternation between landing from the west and the east, more than 70% of planes typically land flying over central London (from the east). So relative to the pre-trial flight patterns, areas to the east experienced more of a reduction in the odd weeks than an increase in the even weeks. Areas further away to the east[4] and to the west in contrast experienced an increase in air traffic for each month during the trial. Indeed the trial report shows that landings from the west

4. When landing, planes have to join a direct line or corridor from the runway, which at Heathrow runs horizontally east to west. Because of the wind prevalence, planes in the odd weeks cannot always land from the west so they have to join the corridors further away from Heathrow to the east during the trial in order to avoid the areas subject to a reduction.

increased and from the east decreased relative to the same five month period a year earlier (Tucker et al. 2013).

The second factor relates to population density. The areas showing a reduction in air traffic overall are densely populated, largely residential areas. Indeed, it covers a large part of metropolitan London stretching to the east of the city (a distance of about 20 miles to the east of Heathrow). This is obvious from Figure 1 (discussed later) which shows the density of general medical practices for the areas covered by the trial. Likewise, areas that experienced increased exposure to landings are less densely populated. The trial report estimated that 138,000 people to the west and over a million to the east of Heathrow experienced a respite during the trial (Tucker et al. 2013).

Finally, it is likely that a complete respite from night noise has stronger impacts on health than an increase in noise from an already noisy environment. This draws on the idea that people may become habituated to noise levels. Although such an effect is not precisely estimated in the medical literature, there is a consensus that it is an important consideration and is very likely to be picked up by our data. Therefore the combination of the wind direction bias, differential population densities and any asymmetric reaction to noise enable us to identify the trial impacts on monthly prescribed medicines.

A visual inspection of the flight tracks comparing our baseline time span (November 2011 to March 2012) to the trial period (November 2012 to March 2013) - see Figures A.1. and A.2. in the Appendix - suggests five geographical zones in the Greater London (GL) area experiencing varied exposure as a result of the trial. These are our 'treated' regions, drawn as trapeziums on Figure 1). We labelled them as follows: GLW1 and GLW2 to the west of Heathrow and GLE1, GLE2 and GLE3 to the east of the airport. The average height of the areas is 10 miles and the average width is 5 miles. The Figures suggest considerable variation in the exposure to early morning aircraft noise for affected sub-populations. The regions called GLE1 and GLE2 were almost free of flights during their exclusion weeks; it was over-flown only slightly more than normal on the other weeks, so overall experienced a reduction in early morning aircraft noise. Similarly there were areas overflown more overall, which were mostly located to the west of Heathrow GLW1, GLW2.

As control group, we chose all medical practises located in two rectangles situated north and south of Heathrow, lying outside of the approach path corridor. The residents in those areas remained unaffected by changes in air traffic throughout the trial period.

{Insert Figure 1 here}

Residential sorting did not seem to be an issue within this setting thanks to two inherent attributes of the trial. The first is suggested by the name of the trial: the Early Morning Arrivals Trial. We assume that most people are at home between 4.30am and 6.00am and are in light sleep hours where deep sleep is infrequent[5]. Secondly, no advance notification about the start of the trial was given to residents (Tucker et al. 2013). The organisations involved decided to communicate the implementation of the change only after the first week of the on-going trial, and then to collect feedback from residents through media and meetings. Therefore, it is unlikely that people relocated due to this unexpected temporary change.

III.b. Empirical specification

The goal of this research is to assess the impact of changes in aircraft emissions on health conditions for those people living underneath flight paths. To simultaneously isolate causal effects of the flight changes and control for confounding factors, we explored GP prescribing differences between communities that experienced the flight change and communities that did not, outside and during the trial. The empirical design adopted here is a standard difference-in-differences (DD) approach. The strong assumption that needs to hold for this model to be valid is the so-called parallel paths assumption: non-affected regions provide information on the expected health outcome trends for affected regions, had changes not occurred. This is discussed further in Section V below.

The epidemiological literature on the detrimental impact of noise pollution on health suggest to

5. Night sleep is divided in a series of cycles made of Rapid Eyes Movements (REM) and non REM episodes. During last cycles before daybreak, REM periods significantly increase which implies shallow sleep, see Klemm (2011).

focus on medical conditions related to central nervous, respiratory and cardiovascular systems. The adverse health consequences are measured by monthly spending on prescriptions for three therapeutic classes. This comprises medications to aid circulation and breathing, and for the central nervous system includes anti-depressants and drugs to treat insomnia.

We estimate the multiple time period DD regression model taking the following form:

$$\ln(SPENDING_{it}^{j}) = \delta_j(TRIAL_t \times TREAT_i) + \gamma^j{}_k + \lambda^j{}_t + \sum_s X_{sit}\beta_s^j + \varepsilon_{it}^j, \qquad (1)$$

where $SPENDING_{it}^{j}$ is the total spending on prescription medicines for one of the three conditions of interest (j) per thousand patients in each practice (i), and month (t). The causal effect of the trial on medication spending is captured by the coefficient δ of the interaction term, with $TRIAL_t$ taking value 1 for the trial months (November 2012 to March 2013) and 0 for the baseline months (November 2011 to March 2012) and $TREAT_i$ taking value 1 for treated practices and 0 for control practices. The model includes region effects (γ_k), where the region k which contains practice i can be broad or more narrowly defined geographical areas as explained below, and monthly time effects (λ_t). X_{it} represents a series of s controls including index of multiple deprivation (IMD) scores to account for socio-economic levels; practice proportions of patients by gender and age; the practice proportion of GPs by age and GPs who qualified in countries other than the UK and finally the number of GPs per thousand patients. The last term, ε_{it}^j represents an idiosyncratic disturbance term.

{Insert Table 1 here}

Table 1 summarises the list of variables. We estimated the model in equation (1) for different macro-regions: first all areas grouped together, then regions GLE1, GLE2, GLE3 and GLW1 individually[6]. In the first case, we estimated the overall effect of the trial. The remaining estimates show the effect by smaller geographical areas that from a visual inspection seemed to experience consistently distinct air

6. GL stands for Greater London, then E is east, W is west. GLW2 is not estimated separately due to the low number of practices in this region.

traffic changes. The analysis of these variations is discussed in Section V

IV. DATA

Monthly general practice prescriptions were drawn from the Health and Social Care Information Centre (HSCIC) for the period from November 2011 to October 2013[7]. The aim was to capture conditions induced and exacerbated by environmental exposure that are treated by medications rather than in emergency rooms or hospital visits. The key variables for our analysis are the practice code (unit of observation) and its postcode, the medication identifier, the month of prescription and the Net Ingredient Cost (NIC - the basic cost of a drug that adjusts for the size/quantity of the medication). We matched the practice postcodes data with the trial regions. The locations of all practices within the Heathrow airport trial areas are shown in Figure 1. Each medication lies within a specific therapeutic class, called BNF (British National Formulary) chapter. The three categories selected for our analysis are central nervous, respiratory and cardiovascular systems. In addition we extracted data for infections and muscoloskeletal and joint diseases to use as placebo conditions in order to test the robustness of our results.

The logarithm transformation of the practice spending per thousand patients is the main outcome used in our analysis. It summarises information on monthly expenditure by practice aggregated at medication category level. In the publicly funded British health system (NHS), this adjusted measure of practice spending corresponds to prescribed medications consumed in countries where health systems relies on private medical insurances. The practice postcode was used to match the practice with the six trial regions (five treatment trapeziums to the east and west and one control - the aggregated areas to the north and south of Heathrow). We assumed that people tend to register with one of the practices closer to their home[8]. Therefore, we expect GP prescribing to be a good measure of medica-

7. The datasets are released under the terms of the Open Government Licence and can be downloaded freely online at: GP practice prescribing data - Presentation level, https://data.gov.uk/dataset/prescribing-by-gp-practice-presentation-level.

8. This idea was confirmed by a recent study on the trade-off between practice quality and patient distance in

tion spending for patients living within the same trial region of the practice. Using GIS (Geographic Information System) tools (QGIS software, Google Earth and Maps Engine) we geocoded the practices' location in order to assign them to the trial areas.

We included the Index of Multiple Deprivation (IMD) data to control for local socio-economic levels; this is a multidimensional composite index including dimensions related to income, employment, health, education and crime.[9] We matched all practice's postcodes with the respective lower layer super output areas (LSOAs), which are socio-geographical areas with an average of 1,500 residents. IMD data at LSOA level for 2011 are published on the data.gov.uk website.[10] Our dataset reported a minimum IMD score of 0.99 for the least deprived areas and a maximum IMD score of 66.21 for the most deprived areas. We gather yearly information on practice characteristics by using General Practice Workforce data. It contains patients headcount and its breakdown by age and gender as well as the number of GPs, their age, gender and country of qualification[11].

As discussed in the previous Section, in South East England wind is predominantly westerly. This is especially important when looking at landing planes at Heathrow since above 5 knots they need to land into the wind regardless of the scheduled landing direction. Introducing a monthly wind switch variable that returned the monthly proportion of nights when wind speed exceeded this threshold does not change the results because of collinearity with the month dummies. Besides the general wind prevalence, the upper panel on Figure 2 shows that March 2013 (a month that falls in the trial period) dramatically deviated from the usual pattern. The number of nights when planes came from the west of Heathrow (i.e. wind blew from east) outweighed the number of nights with planes landing from the east. This contradicted the westerly wind direction prevalence. We addressed this issue in the following

---

England (Santos et al. 2017).

9. The four constituent nations of the UK have each developed their own index of multiple deprivation (IMD). These have been developed to identify small area concentrations of deprivation, and are based on methodology developed at the University of Oxford Social Disadvantage Research Centre (Noble et al. 2006).

10. Freely accessible data provided under the Open Government licence.

11. See https://www.england.nhs.uk/gp/gpfv/workforce/

section dropping March 2013 and exploring the effects on GP spending of a reduced four-month trial period (November 2012 to February 2013 only).

{Insert Table 2 here}

Table 2 reports the summary statistics for the data we used, broken down by control and treatment groups. Overall we are able to use 802 practices for which we can link the prescribing data to the variables listed in Table 1. The practices excluded are specialist clinics, hospitals and out-of-hours services that do not have a patient list. Overall we dropped around 24% of providers, which is similar to other studies using the same data (Rowlingson et al. 2013).

V. RESULTS AND DISCUSSION

V.a. Landing Patterns

As discussed in section 3, the trial implemented a weekly switch between two sets of air traffic exclusion zones, which we term 'odd' and 'even' weeks below. The aim was to provide early morning noise respite to the population affected by landings at Heathrow airport. A very detailed report on the flight patterns during the trial is available (Tucker et al. 2013); here we visually summarise the main findings. In Figures A.1. and A.2., the top panel of both figures represents the map of all landing tracks during the five-month period in the year before the trial. The second and the third panels show the aircraft tracks of planes landing at Heathrow on odd and even weeks during the trial. Since data on medication spending is available in the form of monthly datasets, we aggregated the second and third panels and interpreted the trial as a monthly event comprising a combination of alternated weekly changes. Below we describe how these monthly events are different for each region of interest.

The control regions (outlined above and below the airport on the maps in Figures A.1. and A.2.) included those regions that were not affected by changes implemented during the trial. The GLE1 area (see Figure A.1.) experienced an overall notable reduction in air traffic on odd weeks and a slight increase on even weeks of the trial, with a reduction overall in each month of the trial. Similarly

GLE2 (see Figure A.1.), an area generally subject to heavy early morning air traffic, saw some increase in traffic on the odd weeks and an important drop on the even weeks. These are the two regions most affected by the trial. The last region to the east of Heathrow is GLE3 (see Figure A.1.); if we distinguished the northern from the southern region, the latter experienced an overall increase in air traffic and specifically a sharp increase in traffic on even weeks. From the second and third panels of Figure A.2. we can see that the GLW1 area was characterised by a serious increase in air traffic on the odd weeks and a decrease on the other weeks, implying an overall increase in early morning air traffic. The GLW2 area (see Figure A.2.) saw a drastic reduction of air traffic on odd weeks and almost no change on even weeks. However interesting this area might be, it contains only six GP practices in a mainly rural region.

These are the broad regions identified by the trial final report. However, we assume that the level of variation occurred at a lower regional dimension. Our observations are at the practice level but the environmental quality may be common to groups of practices. This is supported by the fact that noise and air pollution levels vary at a refined level. Maps of noise contours provided by the Civil Aviation Authority draw a picture of how much variation there is from one street to a few streets apart. This suggests using a geographical unit smaller than the broad regions but larger than practice level. We use the Middle Layer Super Output Areas, MSOAs, in which environmental quality is likely to be more homogeneous[12] (Lee et al. 2014). Our unit of observation (practices) is smaller than the MSOAs which could bias our standard-errors, as documented by (Moulton 1986). Failure to take account of this clustering dimension could lead to a downward bias of the standard errors. The main specification, reported below, controlled for these potential common group variations by adopting cluster-robust standard errors - the number of clusters (MSOAs) is large (between 227 to 444 - see Table 2).

We checked for possible standard error bias and calculated the intraclass correlation coefficients

12. MSOAs enclose between 5,000 to 15,000 residents, with an average of 7,700 population as of Census 2011. Each MSOA includes a minimum of one and a maximum of seven practices.

(ICC) of errors and covariate (i.e. $TRIAL_t \times TREAT_i$, the main regressor of interest)[13]. In fact, the correct standard error can be biased by a quantity which depends on the magnitude of those coefficients, on the number of clusters and on the size of the clusters[14]. We obtained very small ICC of covariate (0.073) and zero ICC of errors. This suggests standard error bias may not be a major concern. However we decided to maintain the more conservative cluster adjusted standard errors, rather than the commonly used robust adjustment. These are the main results reported in the paper but later we discuss in detail a series of alternative specifications and corrections to standard errors.

Besides the regional variations due to the trial, we need to keep in mind that wind speed and wind direction affect the landing provenance regardless of the planned schedule. In other words, ideally during the trial there should have been a regular weekly switch between planes landing from the east (i.e. over London) and planes landing from the west (i.e. over Reading). The reality however departs from the forecast due to changing atmospheric conditions. When wind speed is above 5 knots, planes always land into the wind. As we have already mentioned, in South East England on average wind is westerly 70% of the year. We therefore expect more robust results for the three areas to the east of Heathrow - GLE1, GLE2 and GLE3 - as for these regions there was a significant reduction during weeks when they experienced respite (see Figure 2). This westerly preference of planes landing over London was observed during the first four months of the trial.

To summarise, the trial included four broad areas where we can investigate the impacts on medical prescriptions of changes in air traffic during early mornings for five months. What can we expect to be the relationship between the variations in population exposure to noise and air pollution and monthly medication spending? The impact will depend crucially on the population density of the affected areas. Those areas where there appeared to be a significant reduction in air traffic during the trial, GLE1 and

13. This can be done using the `loneway` command in STATA (StataCorp 2014).

14. The so-called Moulton factor, which tells how much larger the corrected standard error would be compared to an unadjusted standard error. With unbalanced group sizes, this is given by: $\frac{SE(\hat{\beta}_1)}{\bar{SE}(\hat{\beta}_1)} = \left(1 + \left[\frac{V(n_g)}{\bar{n}} + \bar{n} - 1\right]\rho_\varepsilon \rho_x\right)^{\frac{1}{2}}$, where n_g is the size of group g; $V(n_g)$ is the variance of group sizes, $\bar{n}$ is the average group size and ρ_ε and ρ_x are the ICC of errors ε and covariate x, respectively.

GLE2, were in fact the most densely populated, as illustrated in Figure 1 by their high GP practice density. Therefore we might expect an overall reduction in medical prescriptions. Our GP practice data are at a much more refined geographical level and so the regressions will ascertain if significant reductions can be detected. The next subsection discusses the results we obtained.

V.b. Effect of the Trial by Health Condition and Region

Our analyses focus on the effects of the trial on central nervous, respiratory and cardiovascular, system ailments. The previous literature showed that these conditions are associated with noise pollution exposure.

An investigation of the parallel paths assumption is given by Figures A.3., A.4., A.5. and A.6. where trends of monthly spending by thousand patients are adjusted by percent of female patients, percent of old patients (85+ years old) and IMD scores of the small socio-geographical areas. They show the patterns of medication spending on control and several treatment groups and generally suggest no differences in trends. Therefore, we take this as supporting evidence that the parallel paths assumption holds.

{Insert Table 3 here}

Table 3 summarises regression estimates using equation (1) by health condition for the whole sample for the main variable of interest, $TRIAL_t \times TREAT_i$, which is a trial indicator equals to 1 for all practices within treated areas and during the five months of the trial and to 0 for the same five months one year earlier[15].

The first column of Table 3 shows the results for the central nervous system, a therapeutic class related to the treatment of sleep loss, concentration deficits and other stress-related diseases. The estimate is significantly negative overall for the regions involved in the trial. This condition showed

15. Full regression results are available in Appendix Tables A.1.-A.3. The regression analysis was repeated for each broad treatment region and included all the atmospheric, socio-economic, GP and patient controls listed in Table 1.

the greatest reduction in spending of 5.8% during the trial. Column 2 of Table 3 reports the results for respiratory system conditions. The five-month trial reduced the spending on respiratory medication by 3.3%.

The final column of Table 3 shows the estimates for cardiovascular system medication spending. This indicates that the trial had no overall significant effect on all regions involved in the flight-path variations. As shown in the Appendix Tables, the coefficient estimates are significantly positive around 7.2% for GLE3 and only slightly significant for GLE1 and GLE2. The weak results here probably reflect the more long term nature of these conditions that make it difficult to identify impacts from short term changes as in our trial.

On average a negative effect on central nervous and respiratory system conditions seem to dominate. The explicit purpose of the systematic flight-paths variations set up by Heathrow airport was to reduce the population exposed to high noise pollution levels during sensitive hours. The results from Table 3 for all regions seem to confirm an overall decrease in medication spending caused by the trial.

The trial final report documented the comments received by local communities after the trial was conducted (Tucker et al. 2013). The response was mixed, residents outside the areas of predictable respite expressed vocal complaints of increased air traffic and annoyance. However, other communities perceived a decrease in early morning noise and positively assessed the trial. Therefore it is worthwhile focusing on the regional results in more detail. These are given in Table 4, where we concentrate on the more robust estimates for the central nervous systems and respiratory conditions.

{Insert Table 4 here}

The GLE1 area reported significant effects mainly for the nervous system class. In fact, there are negative changes in GP spending of 7.7% for central nervous system conditions. The GLE2 region was characterised by a marked decrease in air traffic during its respite weeks and it produced the clearest picture. The almost complete reduction in landing aircraft prevailed over the increase in flights in alternate weeks. In fact during the trial, monthly GP spending decreased significantly by 10.5% for central nervous conditions and by 6.8% for respiratory conditions (this region also shows a decrease

in GP spending on cardiovascular conditions, as shown in the Appendix). Evidently, the results for the GLE2 area indicate that residents benefited from the weekly respite during early morning hours. It appears that two weeks per month of air traffic suspension were enough to reduce monthly prescription spending on all conditions.

For GLE3 as a whole we found a 4.7% significant increase for those medicines related to the central nervous conditions. From the maps in Figures A.1. and A.2. we can see that the change differently affected the northern and the southern part of GLE3. To investigate the effect of the trial on the two regions of GLE3 we separately estimated the model for the two areas. The results - not reported here - showed that prescribing practices in the northern part drove the change, in contrast to our expectations that the southern part experienced the most increase in medication spending. The two main concerns are the reduction in the number of observations and in the areas extension. Having smaller regions opened the issue of patient sorting. In fact residents of one side of the region could easily be registered with a GP on the other side, with a maximum distance from the southern to the northern part of 10 miles. This division also resulted in small numbers of practices, sixteen for GLE3 north and just five practices for GLE3 south.

For the GLW1 region, the coefficient estimates are positive as expected due to an overall increase in air traffic. However, they are not statistically significant. As previously discussed and as shown in the Figure 2, we know that wind is predominantly westerly which implies that the majority of the flights landed over the three other areas. This, combined with the sparse population density and low number of practices in this region, could explain the lack of significant results.

To conclude, our estimates suggest the decreases in air traffic noise were responsible for the health effects. The identification of these effects was aided by the fact that the groups with the higher number of practices, hence more densely populated, and the higher percentage of landing aircraft happened to be the two regions that experienced an important reduction in air traffic during the trial.

V.c. Robustness Tests

We introduced a number of robustness tests to further investigate our main results; these are summarised in Table 5. The top panel reports the baseline coefficient estimates from Table 3. Panel 1 reports the estimates of the coefficient δ of equation (1) with heteroskedasticity-robust standard errors. As expected the standard errors are lower, raising the significance relative to the variant with MSOA clusters. In panel 2 of Table 5 we changed the cluster dimension to a more aggregate level, the four trial zones: GLE1, GLE2, GLE3 and GLW1. The significance levels are comparable to the previous panel with larger standard errors. Therefore the results are robust to alternative error term variance corrections.

{Insert Table 5 here}

For each outcome group we repeated the analysis including all the 24 months of available data from November 11 to October 13. This has the advantage of including months when flight paths returned to normal operation but has the drawback of including seasonal variation unrelated to the trial. We found smaller coefficients with similar levels of significance (panel 3). The second panel of Figure 2 shows the well known seasonal pattern of flights with the majority of landings in the summer months. As the trial was during the off season it seems preferable to compare landings with the same period one year earlier.

The structure of DD panel data raises concerns over serial correlation. The literature does not give unequivocal guidance over the resolution of this potential problem. One reference paper by Bertrand et al. (2004) highlighted that within the DD setting the combined presence of long time series and the use of the period of treatment indicator imposes very little variation over time, potentially leading to serious issues of serial correlation. A common solution is to aggregate the observations across time periods. Therefore we average across all five months for the year before the trial and all 5 months during the trial period, equivalent to using two cross sections.

We estimated equation (1) with this new two-period set up and we obtained the coefficient esti-

mates for the regressor of interest $TRIAL_t \times TREAT_i$ reported in panel 4. We can see that the size and the direction of the effects did not change, however the significance was affected. With such a large reduction in observations it is difficult to obtain very precise estimates. The less restrictive alternative of adding a time trend to equation (1) does not substantially affect the nervous coefficient, although it does impact on the significance of the respiratory coefficient (panel 5). An intermediate approach is to include area time trends as these allow for region specific shocks. In this case the nervous coefficient is larger and highly significant but the value of the respiratory coefficient drops.

As we mentioned earlier, March 2013 showed an unusual wind direction pattern, see Figure 2. To overcome possible issues caused by the easterly wind prevalence in that specific month, we decided to exclude observations for March 2013 and consequently for March 2012. The results in panel 7 of Table 5 suggest that this deviation from the usual wind direction pattern did not significantly affect our original estimates.

We also experimented with alternative regional groupings, given that they are differentially affected by the landing patterns. The results are shown in panels 8 to 11. In panel 8 we include only observations for GLE1 and GLE2, which as previously discussed, and clearly shown in Table 4, reported the most significant results. We estimated the trial coefficients with these two regions grouped together, keeping the same control region and omitting the GLE3 and GLW1 areas. We, therefore, assess the impact of the trial on regions that experienced a visible decrease in air traffic. As expected, the estimates increased in absolute value. GP spending decreased most for central nervous system medication, from 5.8% in the original pooled estimate to 7.6%. For respiratory medication, the overall decrease in GP spending went from 3.3% in the original estimation to 3.9%.

We next report results for all regions to the east of Heathrow, adding GLE3 to the previous specification (panel 9). This confirms the same estimates reported in row 7. The magnitude reduces as we would anticipate considering that in the GLE3 area we see some increase in air traffic during the trial. Keeping observations for GLE3 and GLW1, groups all areas that had an overall increase in air traffic during the trial (panel 10). For these regions, we see a significant increase in cardiovascular medication

spending by around 6%, as well as a positive change of 4.7% for nervous spending. This important result shows that the gains in some areas were, to some extent, counterbalanced by increased spending in regions overflown more heavily during the trial. This lends additional support to our identification strategy that relies on early morning changes in landing patterns. Not only are we observing reduced prescriptions in areas less overflown, but we do observe as well increased spending in those overflown more. Finally we report the results grouping all regions to the west of Heathrow (panel 11). For this specification we retrieve data for the GLW2 area that was excluded from the analysis due to too few practices. The signs remain positive for the three therapeutic classes, but the coefficients are not significant.

We detected a substantial decrease in spending from June 2012 onwards for cardiovascular system diseases. We discovered that in May 2012 the patent of a medicine widely used to control cholesterol levels (atorvastatin) expired inducing a 93% reduction in its price. Consequently the NHS advised GPs to switch to atorvastatin[16]. This change is likely to have been driven by the drop in the medicine price rather than in a decrease in the quantity prescribed. To account for the possibility that the switch to the generic medicine has been differentially adopted in the treated and control groups, we added a further division: cardiovascular diseases spending excluding atorvastatin medicines. Panel 12 of Table 5 shows results for all cardiovascular medicines other than atorvastatin to rule out a possible confounding effect caused by this drug. The coefficient estimate changed in size but remained statistically insignificant, as for the coefficients of the main specification.

In row 13, we test the sensitivity of our results to the chosen control group. We can see on Figure 1 that the two rectangles to the north and the south of the runways have slightly different subpopulations size (the north including more medical practices than the south). We therefore run additional regressions using the north rectangle as control only, excluding practices to the south. The coefficients retain the same significance, with both the coefficients and standard-errors only marginally

16. See http://www.pulsetoday.co.uk/price-of-atorvastatin-plummets-93-as-patent-end.

increased. The difference in coefficients is not statistically significant at any conventional levels.

Finally we ran a series of regressions as placebos using health conditions that were deemed unlikely to be affected by variations in ambient noise. We identified infections and muscoloskeletal and joint diseases as such 'placebos', deemed unaffected by either noise pollution exposure. Panel 14 of Table 5 shows the results of this analysis. The estimates for both therapeutic classes were found to be statistically insignificant, hence providing further support for our identification strategy.

V.d. Impacts on Health Spending

We next investigate the economic significance of our results. Table 6 shows back-of-the-envelope calculations of changes in monthly prescribing costs due to the implementation of the trial by region, which generated an overall decrease in spending by GP practices.

For instance, for the GLE1 and GLE2 regions we found a 7.6% reduction in monthly spending on nervous system conditions per thousand patients (see Table 5). On average a practice based in GLE1 or GLE2 has 6,600 patients and recorded about £1,760 monthly spending per thousand patients (derived from Table 2). From these figures we calculated the monthly change in spending per practice, and we multiplied it by 351 - the total number of the practices within the GLE1 and GLE2 regions (see Table 2). The result of this calculation is shown in Table 6 and adds up to about £310,000 saved in monthly spending for the whole GLE1 and GLE2 regions only for the nervous system therapeutic class.

{Insert Table 6 here}

To put this number in context, we calculated the monthly saving in these regions arising from the substitution to atorvastatin following the expiration of the patent in May 2012, as described above. This suggested about £110,000 savings per month from this one drug alone. Therefore our estimate of the savings from the trial for the entire nervous system class of drugs, £310,000, seems realistic.

We similarly calculated the cost savings for respiratory conditions, which was generally significant but less robust, and added these to the nervous system savings. Looking at all the regions involved in the trial, we calculated a net monthly saving overall of about £420,000. Had the flights reduction

been adopted permanently[17], the NHS would have saved around £5 millions per year in respiratory and nervous system prescribing costs. To put this figure into perspective we can calculate the total annual prescribing spend in the trial area. In 2013 in England the prescribing spend was at £142 per person[18]. Multiplying this by the 403 practices times the average number of patients per practice, we obtain about £410 millions, which consists of an estimate of the annual total prescribing spend in the trial regions. Therefore, the estimated savings account for 1.23% of the total prescribing spending. We should also note that these are likely to be conservative figures since we ruled out all those practices that did not have a patient list (e.g., specialist clinics, out-of-hours services and hospitals - which accounted for about 24% of all practices).

To complete the figure of the induced monetary saving, we should add the reduced costs of GP time due to the likely lower number of visits by patients to request prescriptions. However we do not have sufficient data to estimate this. In addition there are likely to be indirect benefits, such as reduced absenteeism and related gains in productivity. Combining these with the direct reduction in medical spending is likely to lead to much greater savings.

VI. CONCLUSIONS

In this paper we estimate the health externalities generated by noise pollution from aircraft, exploiting a five-month trial that took place around London Heathrow airport from November 2012 to March 2013. The trial involved changes in patterns of aircraft landings during early morning hours (4.30am to 6.00am). Health effects are measured through changes in medication prescribing by GP practice. We find a statistically significant response of monthly medication spending on central nervous and respiratory system conditions to these changes, and weak effects for cardiovascular conditions. We detect significant reductions in prescription spending on central nervous and respiratory conditions in

17. The trial was not made permanent after a well orchestrated campaign by residents who experienced an increase in air traffic in one specific area during the trial.

18. See "Annual prescribing spend per person in the UK", Nuffield Trust at http://www.nuffieldtrust.org.uk/data-and-charts/prescribing-spend-person-uk

the regions that experience a drop in air traffic during the trial. Residents in regions more overflown during the trial have increased their medicines intakes but the effects are weaker. These results are also consistent with the idea that complete respite has a stronger effect than an increase in an already noisy environment.

This quasi-experimental approach suggests a causal impact of aircraft noise pollution on human health. By relying on a quasi-experimental research design (Graff Zivin & Neidell 2013), we complement previous epidemiology based studies that find negative associations between aircraft noise and health around major airports (Clark et al. 2012).

This study also illustrates the benefits of using publicly available data to estimate some of the direct costs from adverse environmental exposure imposed on society, whose costs are often borne by the public health system. Our calculations suggest a sizeable direct impact on GP spending in the areas affected. These estimates do not include the reduced costs of avoided GP visits, the gain in patients well-being, and impacts on individual worker productivity through absenteeism or less effective effort in the workplace. Our findings suggest that small variations to air traffic exposure during critical hours affect health and this could inform environmental policy.

Funding

Financial support from the Economic and Social Research Council (ESRC-KISS DTC) gratefully awarded to Silvia Beghelli.

Acknowledgments

Helpful comments were received by seminar participants at East Anglia, Middlesex, Oxford, Queen Mary and York universities. We especially acknowledge comments from Joshua Angrist, Brian Bell, Osea Giuntella, Mireia Jofre-Bonet, Carol Propper, Nigel Rice and Filipa Sá.

References

Ando, Y., Nakane, Y. & Egawa, J. (1975), 'Effects of aircraft noise on the mental work of pupils', *Journal of Sound and Vibration* **43**, 683–691.

Basner, M., Babisch, W., Davis, A., Brink, M., Clark, C., Janssen, S. & Stansfeld, S. (2014), *Lancet* **383**(9925), 1325–32.

Beatty, T. K. M. & Shimshack, J. P. (2011), 'School buses, diesel emissions, and respiratory health.', *Journal of Health Economics* **30**(5), 987–99.

Bertrand, M., Duflo, E. & Mullainathan, S. (2004), 'How much should we trust differences-in-differences estimates?', *Quarterly Journal of Economics* **119**(1), 249–75.

Boes, S., Nüesch, S. & Stillman, S. (2013), 'Aircraft noise, health, and residential sorting: Evidence from two quasi-experiments', *Health Economics* **22**(9), 1037–1051.

Cesur, R., Tekin, E. & Ulker, A. (2017), 'Air pollution and infant mortality: Evidence from the expansion of natural gas infrastructure', *The Economic Journal* **127**(600), 330–362.

Chen, T. J. & Chen, S. S. (1993), 'Effects of aircraft noise on hearing and auditory pathway function of school-age children', *International Archives of Occupational and Environmental Health* **65**, 107–111.

Clark, C., Crombie, R., Head, J., van Kamp, I., van Kempen, E. & Stansfeld, S. A. (2012), 'Does traffic-related air pollution explain associations of aircraft and road traffic noise exposure on children's health and cognition?', *American Journal of Epidemiology* **176**(4), 327–337.

Cohen, S., Evans, G. W., Krantz, D. S., Stokols, D. & Kelly, S. (1981), 'Aircraft noise and children: Longitudinal and cross-sectional evidence on adaptation to noise and the effectiveness of noise abatement', *Journal of Personality and Social Psychology* **40**, 331–345.

Currie, J. & Walker, R. (2011), 'Traffic Congestion and Infant Health: Evidence from E-ZPass', *American Economic Journal: Applied Economics* **3**(1), 65–90.

Deryugina, T., Heutel, G., Miller, N., Molitor, D. & Reif, J. (2016), The mortality and medical costs of air pollution: Evidence from changes in wind direction, NBER Working Paper w22796, National Bureau of Economic Research Cambridge.

European Commission, (2004), Position paper on dose-effect relationships for night time noise, Technical Report , Working Group on Health and Socio-Economic Aspects.

Evans, G. W., Hygge, S. & Bullinger, M. (1995), 'Chronic noise and psychological stress', *Psychological Science* **6**, 333–338.

Graff Zivin, J. & Neidell, M. (2013), 'Environment, Health and Human Capital', *Journal of Economic Literature* **51**(3), 689–730.

Halliday, T., John, L. & de Paula A. (2018), 'Vog: Using Volcanic Eruptions to Estimate the Health Costs of Particulates', *forthcoming in: The Economic Journal* .

Hansell, A. L. et al. (2013), 'Aircraft noise and cardiovascular disease near Heathrow airport in London: small area study', *British Medical Journal* **347**(7928), 1–10.

Heathrow Airport Limited (2013), A quieter Heathrow, Technical Report May, London.

Klemm, W. R. (2011), 'Why Does REM Sleep Occur? A Wake-Up Hypothesis', *Frontiers in Systems Neuroscience* **5**(73), 1–12.

Lee, J., Cebrian, G., Edmonds, L., Patel, J. & Weston, E. (2014), ERCD REPORT 1401. Noise Exposure Contours for Heathrow Airport 2013, Technical Report October, CAA, London.

Moretti, E. & Neidell, M. (2011), 'Pollution, Health, and Avoidance Behavior: Evidence from the Ports of Los Angeles', *Journal of Human Resources* **46**(1), 154–175.

Moulton, B. R. (1986), 'Random group effects and the precision of regression estimates', *Journal of Econometrics* **32**(3), 385–397.

Noble, M., Wright, G., Smith, G. & Dibben, C. (2006), 'Measuring multiple deprivation at the small-area level', *Environment and Planning A* **38**(1), 169–185.

Rowlingson, B., Diggle, P., Taylor, B. & Lawson, E. (2013), 'Mapping English GP prescribing data: a tool for monitoring health-service inequalities', *BMJ open* **3**(1), 1–10.

Santos, R., Gravelle, H. & Propper, C. (2017), 'Does quality affect patients' choice of doctor? Evidence from England', *The Economic Journal* **127**(600), 445–494.

Schlenker, W. & Walker, W. R. (2016), 'Airports, air pollution, and contemporaneous health', *Review of Economic Studies* **83**(2), 768–809.

Tucker, K., Williams, R. & Leighton, S. (2013), Heathrow Airport - Early Morning Arrival Trial Final report, Technical Report August, Helios.

Tzivian, L., Winkler, A., Dlugaj, M., Schikowski, T., Vossoughi, M., Fuks, K., Weinmayr, G. & Hoffmann, B. (2015), 'Effect of long-term outdoor air pollution and noise on cognitive and psychological functions in adults', *International Journal of Hygiene and Environmental Health* **218**(1), 1–11.

van Kamp, I. & Davies, H. (2013), 'Noise and health in vulnerable groups: A review', *Noise Health* **15**(64), 153–159.

Wolfe, P. J., Kramer, J. L. & Barrett, S. R. (2017), 'Current and future noise impacts of the UK hub airport', *Journal of Air Transport Management* **58**, 91–99.

Table 1: List of variables.

Outcome	
GP spending on prescribed drugs for:	Central nervous systems related, respiratory and cardiovascular conditions, monthly medication per practice
Covariates	
Socio-economic	Index of Multiple Deprivation at LSOA[†] level
GP density	GPs per thousand patients
GP characteristics	Non-UK qualified; females; <30 yrs; 30-49 yrs; 50-64 yrs
Patient characteristics	Females; 4-14 yrs; 45-64 yrs; 65-74 yrs; 75-84 yrs; >85 yrs

[†]LSOA: Lower Layer Super Output Area, socio-geographical area with an average of 1,500 residents.

Table 2: Sample descriptive statistics, monthly averages, Nov 2012 - Mar 2013 (during the trial) and Nov 2011 - Mar 2012 (before the trial).

Variables	Total	Control	GLE1	GLE2	GLE3	GLW1
Number of practices	802	393	197	154	21	31
Number of MSOAs	444	213	120	83	16	14
IMD scores	27.63	23.40	32.88	36.77	15.65	16.87
	(13.77)	(13.27)	(10.46)	(12.08)	(10.20)	(7.79)
GPs per 1,000 patients	0.70	0.68	0.80	0.65	0.60	0.69
	(0.34)	(0.31)	(0.43)	(0.27)	(0.21)	(0.26)
Patients per practice	6,550	6,233	7,074	6,112	7,771	7,717
	(3,912)	(3,820)	(3,897)	(3,648)	(4,105)	(4,370)
% patients:						
- females	49.14	49.46	49.05	48.24	51.24	48.34
	(3.98)	(3.18)	(5.40)	(4.06)	(1.43)	(2.37)
- children (4 to 14 yrs)	10.88	11.14	9.32	11.95	12.56	10.80
	(2.96)	(2.68)	(3.22)	(2.60)	(2.24)	(2.67)
- elderly (over 85)	1.26	1.42	0.92	1.09	2.02	1.49
	(0.80)	(0.81)	(0.52)	(0.84)	(0.82)	(0.80)
Prescribed medicines, spending per 1,000 patients:						
- Central nervous system						
before the trial	1,768	1,747	1,813	1,715	2,099	1,762
	(882)	(632)	(1,324)	(786)	(389)	(734)
during the trial	1,592	1,575	1,618	1,496	2,069	1,734
	(715)	(579)	(987)	(627)	(336)	(717)
- Respiratory system						
before the trial	1,063	1,075	967	1,068	1,381	1,232
	(398)	(381)	(369)	(440)	(244)	(431)
during the trial	1,068	1,068	985	1,065	1,432	1,289
	(401)	(383)	(380)	(427)	(258)	(477)
- Cardiovascular system						
before the trial	1,458	1,538	1,222	1,466	2,003	1,503
	(558)	(547)	(479)	(584)	(429)	(519)
during the trial	957	985	835	953	1,433	1,041
	(350)	(327)	(327)	(362)	(342)	(351)

Notes: Standard deviations in parenthesis. IMD refers to the Index of Multiple Deprivation. MSOA: Middle Layer Super Output Areas, which are geographies with a mean population of around 7,700.

Table 3: Trial effect on medication spending per 1,000 patients by therapeutic class.

All regions	Nervous system (1)	Respiratory (2)	Cardiovascular (3)
$TRIAL \times TREAT$	-0.058***	-0.033*	0.020
	(0.020)	(0.018)	(0.017)
Adjusted R^2	0.655	0.687	0.829
Observations	7832	7834	7834
Clusters	444	444	444
Months	10	10	10

Notes: Every regression includes all control variables. Tables showing the full set of control variables are presented in Appendix A.1.-A.3. Cluster-robust standard errors in parentheses. The clustering dimension is MSOA, where each cluster has a minimum of 1 and a maximum of 7 practices. 10 months correspond to 5 in the baseline period (Nov 2011- Mar 2012) plus 5 in the trial period (Nov 2012- Mar 2013). * $p < .1$, ** $p < .05$, *** $p < .01$

Table 4: Trial effect on medication spending per 1,000 patients by therapeutic class and treatment regions.

Region		Nervous system (1)	Respiratory conditions (2)
GLE1	$TRIAL \times TREAT$	-0.077***	-0.033
		(0.027)	(0.021)
	Adjusted R^2	0.634	0.669
	Observations	5843	5845
	Clusters	333	333
	Months	10	10
GLE2	$TRIAL \times TREAT$	-0.105***	-0.068***
		(0.023)	(0.020)
	Adjusted R^2	0.691	0.662
	Observations	5374	5374
	Clusters	296	296
	Months	10	10
GLE3	$TRIAL \times TREAT$	0.047**	0.001
		(0.022)	(0.024)
	Adjusted R^2	0.675	0.643
	Observations	4110	4110
	Clusters	229	229
	Months	10	10
GLW1	$TRIAL \times TREAT$	0.046	0.009
		(0.036)	(0.023)
	Adjusted R^2	0.665	0.646
	Observations	4205	4205
	Clusters	227	227
	Months	10	10

Notes: Every regression includes all control variables. Cluster-robust standard errors in parentheses.
The clustering dimension is MSOA, where each cluster has a minimum of 1 and a maximum of 7 practices.
10 months correspond to 5 in the baseline period (Nov 2011- Mar 2012) plus 5 in the trial period (Nov 2012- Mar 2013)
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 5: Robustness tests for all regions involved in the trial.

		Nervous	Respiratory	Cardiovascular
Baseline - MSOAs clusters	$TRIAL \times TREAT$	-0.058*** (0.020)	-0.033* (0.018)	0.020 (0.017)
	Adjusted R^2	0.655	0.687	0.829
	N	7832	7834	7834
1. No clusters	$TRIAL \times TREAT$	-0.058*** (0.013)†	-0.033*** (0.011)†	0.020** (0.010)†
	Adjusted R^2	0.829	0.687	0.655
	N	7834	7834	7832
2. Trial zones as clusters	$TRIAL \times TREAT$	-0.058** (0.018)	-0.033** (0.014)	0.020 (0.026)
	Adjusted R^2	0.829	0.687	0.655
	N	7832	7834	7834
3. Full period of 24 months	$TRIAL \times TREAT$	-0.035*** (0.011)	-0.020** (0.009)	0.005 (0.009)
	Adjusted R^2	0.812	0.665	0.655
	N	18802	18804	18801
4. Averaging across time	$TRIAL \times TREAT$	-0.055** (0.022)	-0.028 (0.018)	0.023 (0.019)
	Adjusted R^2	0.611	0.694	0.830
	N	1569	1569	1569
5. With time trend	$TRIAL \times TREAT$	-0.059*** (0.019)	-0.024 (0.017)	-0.023 (0.016)
	Adjusted R^2	0.642	0.681	0.816
	N	7832	7834	7834
6. With area-specific time trends	$TRIAL \times TREAT$	-0.081*** (0.017)	-0.014 (0.014)	-0.164*** (0.017)
	Adjusted R^2	0.641	0.680	0.784
	N	7832	7834	7834
7. Dropping obs for March 2012 and 2013	$TRIAL \times TREAT$	-0.058*** (0.021)	-0.033* (0.019)	0.019 (0.017)
	Adjusted R^2	0.827	0.679	0.646
	N	6267	6267	6266
8. GLE1 & GLE2	$TRIAL \times TREAT$	-0.076*** (0.022)	-0.039** (0.019)	0.013 (0.018)
	Adjusted R^2	0.654	0.678	0.827
	N	7317	7319	7319
9. GLE1, GLE2 & GLE3	$TRIAL \times TREAT$	-0.067*** (0.021)	-0.036** (0.018)	0.017 (0.018)
	Adjusted R^2	0.656	0.683	0.830
	N	7527	7529	7529
10. GLE3 & GLW1	$TRIAL \times TREAT$	0.047* (0.025)	0.006 (0.019)	0.059*** (0.021)
	Adjusted R^2	0.670	0.654	0.801
	N	4415	4415	4415
11. GLW1 & GLW2	$TRIAL \times TREAT$	0.042 (0.033)	0.005 (0.021)	0.038 (0.028)
	Adjusted R^2	0.662	0.645	0.796
	N	4265	4265	4265
12. No atorvastatin	$TRIAL \times TREAT$			-0.002 (0.017)
	Adjusted R^2			0.804
	N			7834
13. Control area: north rectangle only	$TRIAL \times TREAT$	-0.065*** (0.023)	-0.038* (0.021)	0.029 (0.020)
	Adjusted R^2	0.650	0.679	0.825
	N	6727	6769	6769
14. Placebos		Infections	Musculoskeletal	
	$TRIAL \times TREAT$	-0.003 (0.024)	-0.023 (0.019)	
	Adjusted R^2	0.399	0.656	
	N	7833	7831	

Notes: Cluster-robust standard errors in parentheses if not otherwise specified. † Heteroskedasticity-robust standard errors. * $p < .1$, ** $p < .05$ and *** $p < .01$.

Table 6: Monthly change in prescribing costs (GBP) induced by the five-month systematic flight paths variation. GP spending for the cardiovascular therapeutic is omitted since no significant results were detected.

	Overall†	*GLE1 & GLE2*
Practices	403	351
Thousand patients per practice	7.17	6.59
Respiratory conditions	-110,745	-91,817
Nervous system	-309,481	-310,152
Total	-420,226	-401,969

†Results for the MSOAs within the areas of GLE1, GLE2, GLE3 and GLW1.

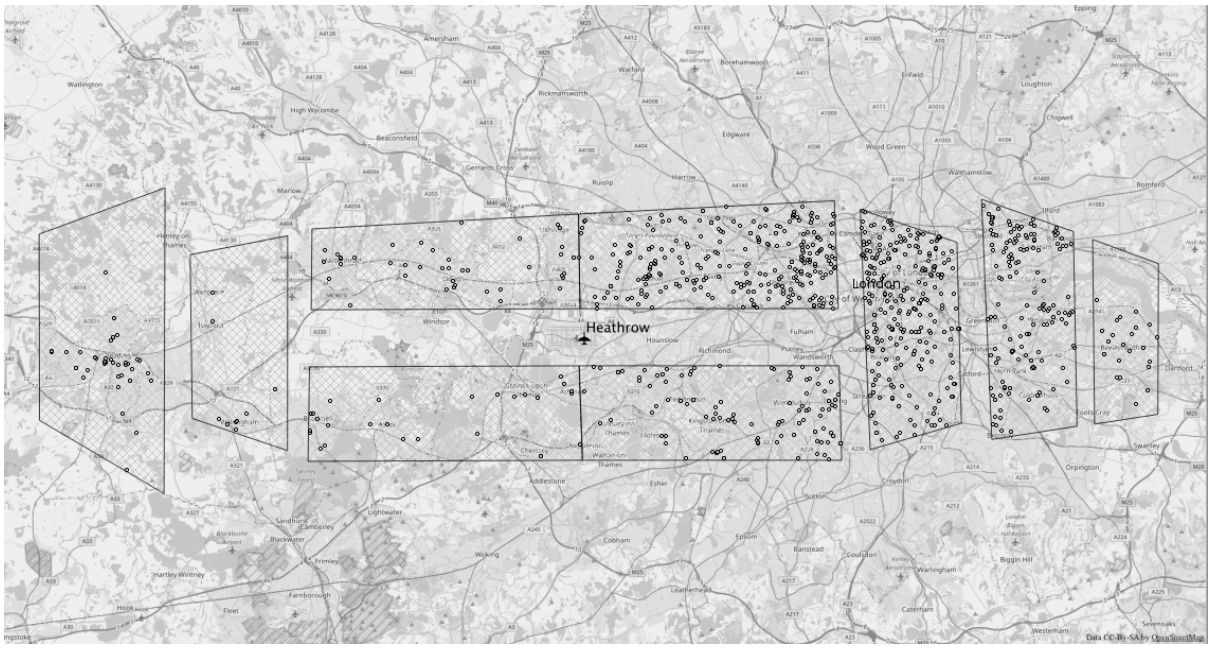


Figure 1: Location of Heathrow airport, GP practices (dots) and trial areas: two control rectangles, north and south - five treated trapeziums, two west and three east of Heathrow.

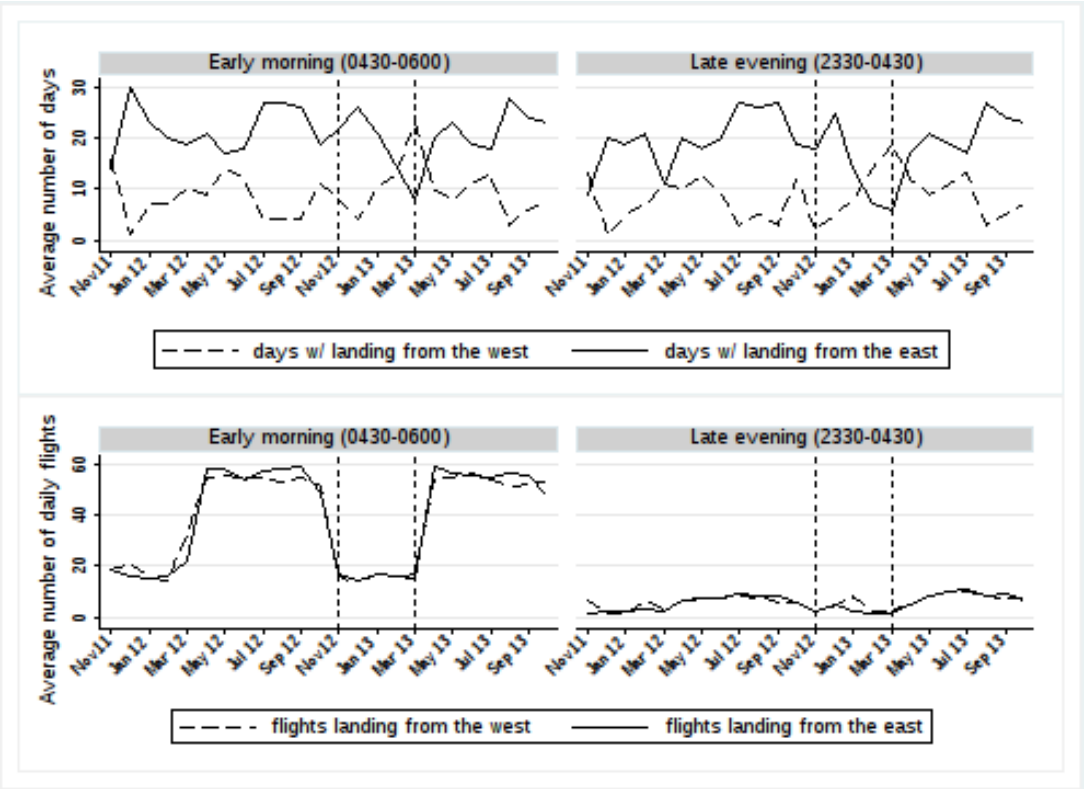


Figure 2: Average monthly number of days and flights per landing direction.

A. APPENDIX

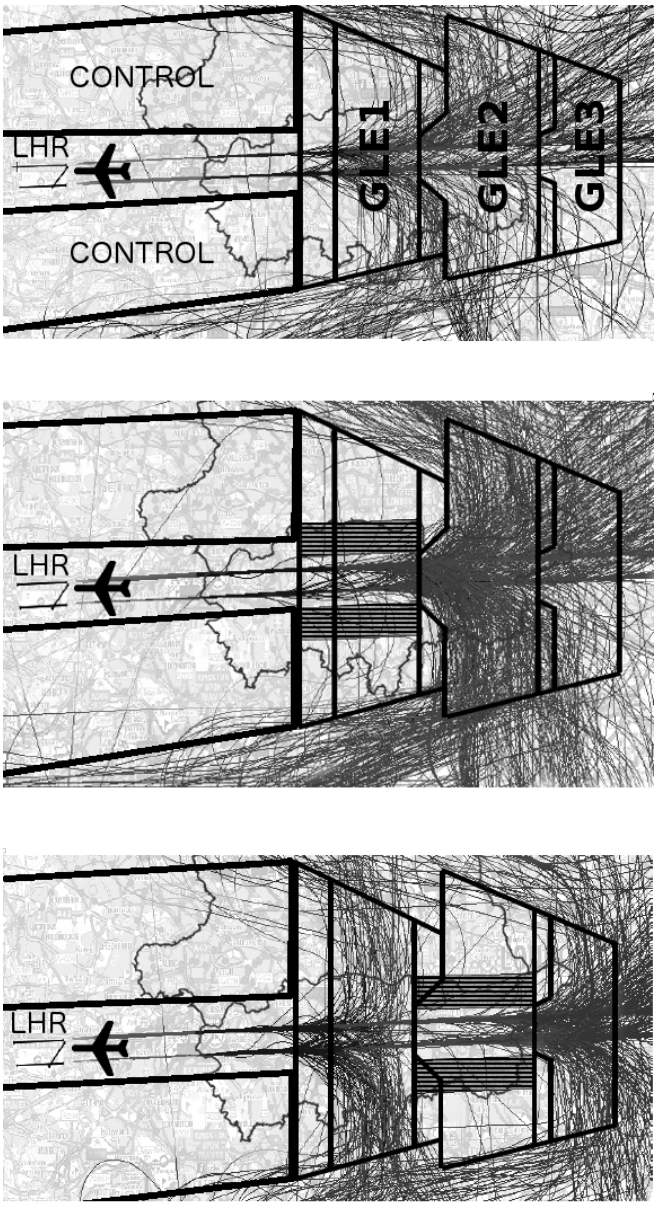


Figure A.1.: Affected areas drawn on flight tracks from the trial report (Tucker et al. 2013). Aircraft tracks for the baseline period 2011/2012 - top panel (∼45 nights) - and for the trial period 2012/2013 - middle (∼44 nights) and bottom (∼41 nights) panels, when aircraft landed from the east and inner and outer exclusion zones (i.e. the shaded areas in the second and third panels) were operative, respectively. London Heathrow airport is labelled as LHR. The maps show five macro-regions involved in the study: the control zones are to the north and south of LHR; and to the east of LHR there are GLE1, GLE2 and GLE3. All areas on these maps are approximative.

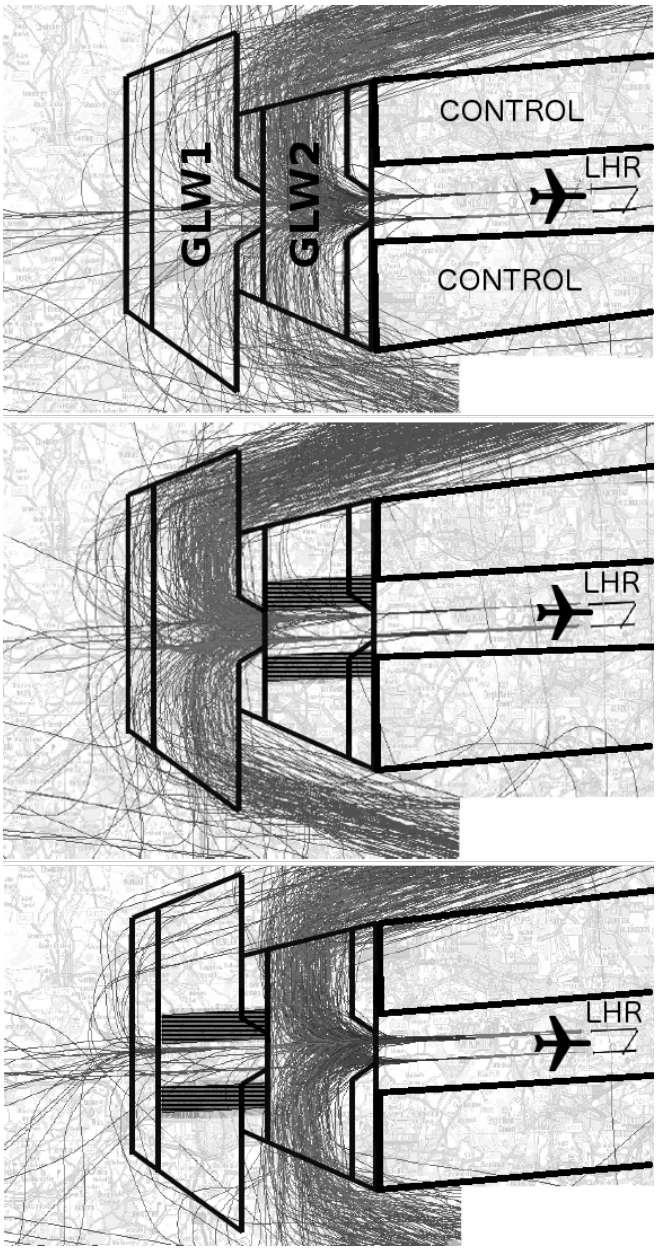
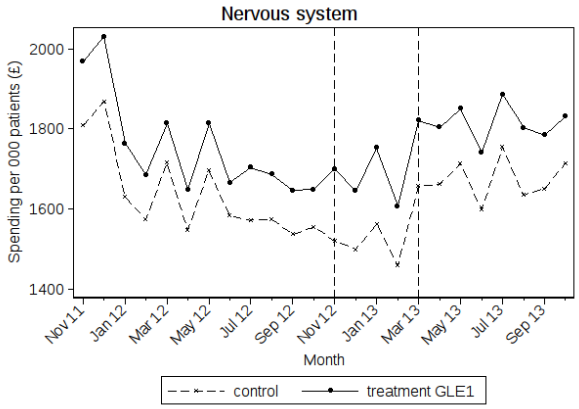
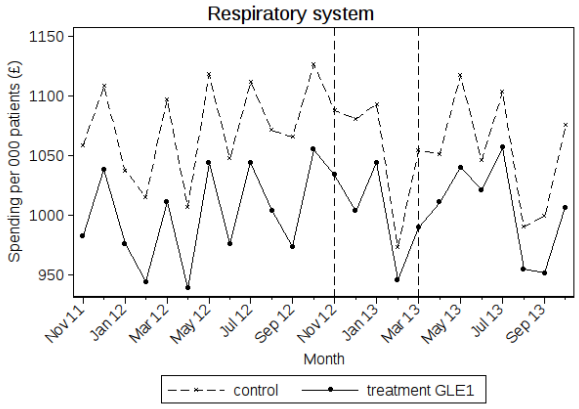


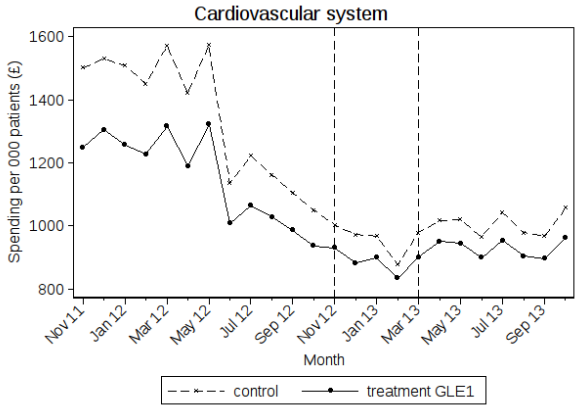
Figure A.2.: Affected areas drawn on flight tracks from the trial report (Tucker et al. 2013). Aircraft tracks for the baseline period 2011/2012 - top panel (∼25 nights) - and for the trial period 2012/2013 - middle (∼25 nights) and bottom (∼25 nights) panels, when aircraft landed from the west and inner and outer exclusion zones (i.e. the shaded areas in the second and third panels) were operative, respectively. London Heathrow airport is labelled as LHR. The maps show four macro-regions involved in the study: the control zones are to the north and south of LHR; and to the west of LHR there are GLW1 and GLW2. All areas on these maps are approximative.



(a) *Nervous system.*

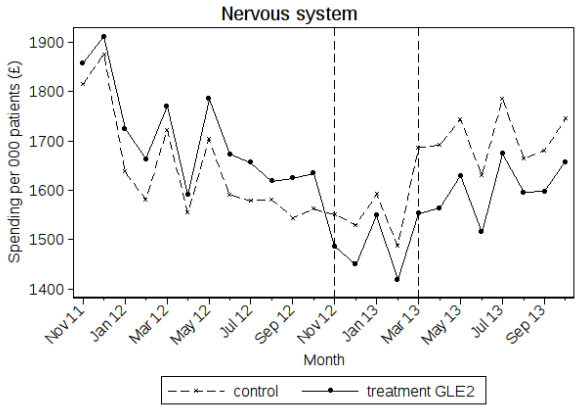


(b) *Respiratory system.*

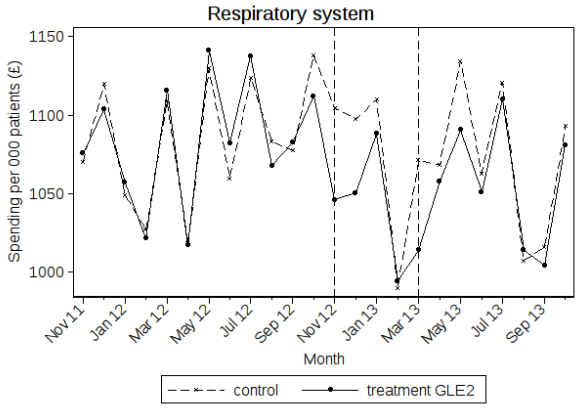


(c) *Cardiovascular system.*

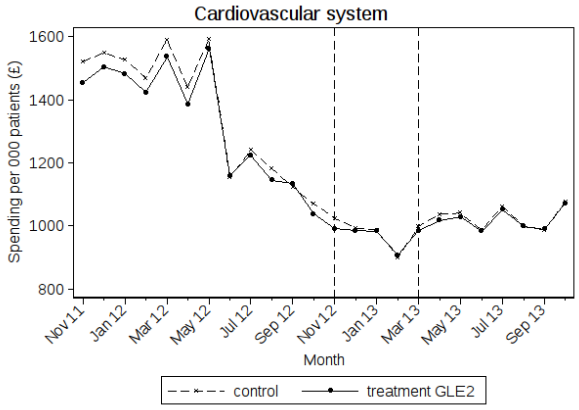
Figure A.3.: GLE1 area. Average monthly practice medication spending related to different categories of medication adjusted by IMD score, percent of female patients and percent of the elderly (85 and above years old). The dashed lines indicate the trial period from November 2012 to March 2013.



(a) *Nervous system.*

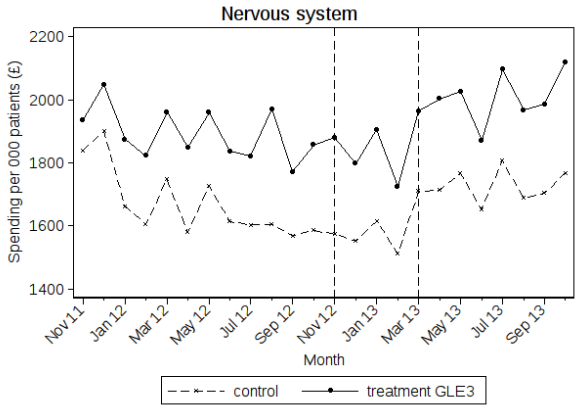


(b) *Respiratory system.*

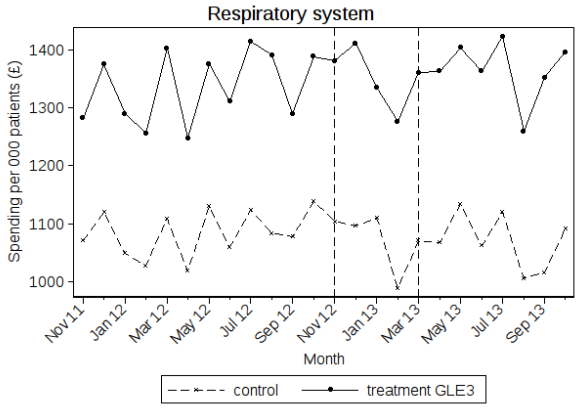


(c) *Cardiovascular system.*

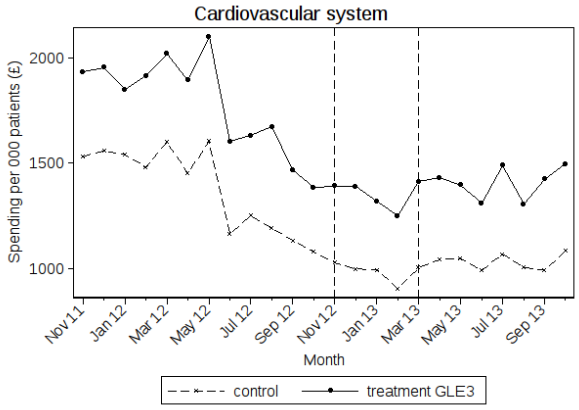
Figure A.4.: GLE2 area. Average monthly practice medication spending related to different categories of medication adjusted by IMD score, percent of female patients and percent of the elderly (85 and above years old). The dashed lines indicate the trial period from November 2012 to March 2013.



(a) *Nervous system.*

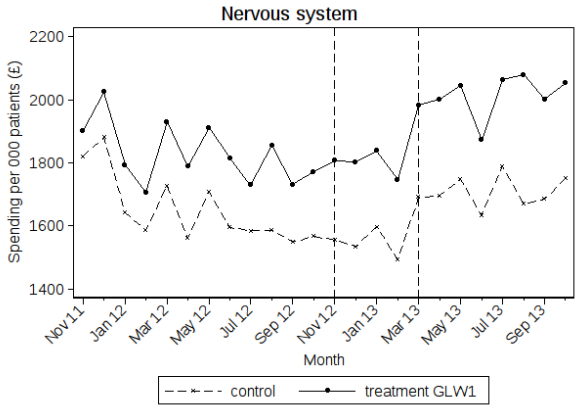


(b) *Respiratory system.*

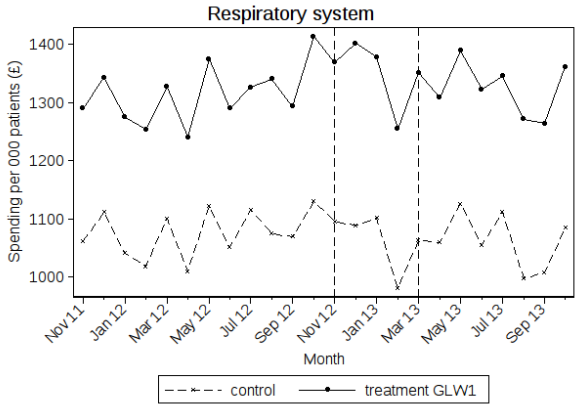


(c) *Cardiovascular system.*

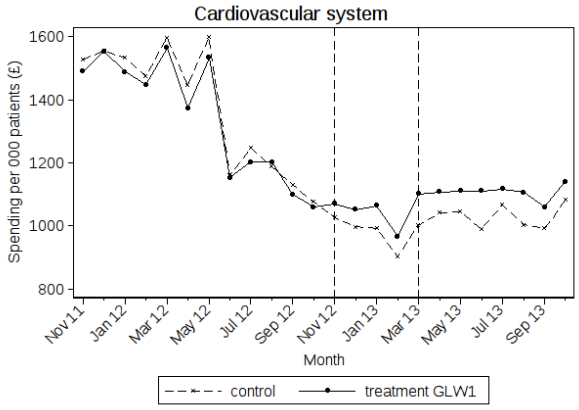
Figure A.5.: GLE3 area. Average monthly practice medication spending related to different categories of medication adjusted by IMD score, percent of female patients and percent of the elderly (85 and above years old). The dashed lines indicate the trial period from November 2012 to March 2013.



(a) *Nervous system.*



(b) *Respiratory system.*



(c) *Cardiovascular system.*

Figure A.6.: GLW1 area. Average monthly practice medication spending related to different categories of medication adjusted by IMD score, percent of female patients and percent of the elderly (85 and above years old). The dashed lines indicate the trial period from November 2012 to March 2013.

Table A.1.: Trial effect on nervous system medication spending per 1,000 patients by treatment regions.

	All	GLE1	GLE2	GLE3	GLW1
$TRIAL \times TREAT$	-0.058***	-0.077***	-0.105***	0.047**	0.046
	(0.020)	(0.027)	(0.023)	(0.022)	(0.036)
treatment region		-0.248	0.144	-0.363*	-0.768***
		(0.177)	(0.126)	(0.212)	(0.186)
GLE3	0.378*				
	(0.226)				
GLE2	-0.007				
	(0.239)				
GLE1	-0.257				
	(0.164)				
GLW1	-0.446***				
	(0.090)				
Dec 2011	0.037***	0.033***	0.035***	0.034***	0.037***
	(0.007)	(0.008)	(0.008)	(0.009)	(0.009)
Jan 2012	-0.100***	-0.108***	-0.106***	-0.109***	-0.108***
	(0.007)	(0.008)	(0.009)	(0.010)	(0.010)
Feb 2012	-0.136***	-0.145***	-0.137***	-0.139***	-0.140***
	(0.007)	(0.008)	(0.009)	(0.010)	(0.010)
Mar 2012	-0.053***	-0.060***	-0.052***	-0.049***	-0.046***
	(0.007)	(0.008)	(0.009)	(0.010)	(0.010)
Nov 2012	-0.161***	-0.171***	-0.159***	-0.161***	-0.159***
	(0.015)	(0.016)	(0.015)	(0.016)	(0.016)
Dec 2012	-0.178***	-0.186***	-0.178***	-0.178***	-0.173***
	(0.015)	(0.015)	(0.014)	(0.015)	(0.016)
Jan 2013	-0.130***	-0.141***	-0.130***	-0.138***	-0.137***
	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)
Feb 2013	-0.203***	-0.211***	-0.203***	-0.206***	-0.203***
	(0.015)	(0.015)	(0.014)	(0.015)	(0.015)
Mar 2013	-0.083***	-0.083***	-0.090***	-0.082***	-0.076***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
IMD score	-0.000	-0.002	0.002	0.002	0.003
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
GPs per 1,000 patients	-0.073	-0.040	-0.003	0.018	-0.008
	(0.099)	(0.118)	(0.050)	(0.052)	(0.051)
Patients females	3.170***	2.603**	4.456***	4.523***	4.893***
	(0.932)	(1.114)	(0.789)	(0.910)	(0.933)
Patients 4-14 years old	-0.799	0.072	1.236	2.670***	2.332**
	(1.506)	(1.770)	(0.927)	(0.994)	(0.999)
Patients 45-64 years old	5.144***	5.522***	1.113	1.455*	2.094**
	(1.761)	(2.035)	(0.824)	(0.809)	(0.880)
Patients 65-74 years old	-3.164	-4.264	4.215	2.578	1.155
	(3.406)	(3.904)	(2.650)	(2.464)	(2.546)
Patients 75-84 years old	1.018	0.705	3.483	2.538	-0.234
	(3.183)	(3.569)	(3.039)	(3.286)	(3.611)
Patients over 85 years old	16.049***	18.110***	6.680	10.632**	16.046***
	(5.331)	(5.413)	(5.153)	(4.884)	(5.814)
GP females	-0.021	-0.034	-0.068	-0.081	-0.078
	(0.053)	(0.060)	(0.049)	(0.056)	(0.054)
GP up to 30 years old	0.377**	0.386**	0.331**	0.297*	0.365*
	(0.171)	(0.189)	(0.145)	(0.151)	(0.186)
GP 30-49 years old	0.178**	0.180**	0.150**	0.156**	0.142**
	(0.072)	(0.083)	(0.069)	(0.071)	(0.068)
GP 50-64 years old	-0.022	-0.044	0.004	-0.011	-0.018
	(0.064)	(0.082)	(0.055)	(0.062)	(0.063)
GP qualified outside UK	0.088*	0.113*	0.080*	0.092*	0.112**
	(0.049)	(0.063)	(0.044)	(0.051)	(0.051)
constant	4.714***	4.903***	4.133***	4.417***	4.214***
	(0.434)	(0.518)	(0.394)	(0.493)	(0.511)
MSOA dummies	✓	✓	✓	✓	✓
Observations	7832	5843	5374	4110	4205
Adjusted R^2	0.655	0.634	0.691	0.675	0.665

Notes: Cluster-robust standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$

Table A.2.: Trial effect on respiratory medication spending per 1,000 patients by treatment regions.

	All	GLE1	GLE2	GLE3	GLW1
$TRIAL \times TREAT$	-0.033*	-0.033	-0.068***	0.001	0.009
	(0.018)	(0.021)	(0.020)	(0.024)	(0.023)
treatment region		0.112	0.388***	-0.576***	-0.523***
		(0.122)	(0.103)	(0.203)	(0.159)
GLE3	0.807***				
	(0.167)				
GLE2	0.541***				
	(0.149)				
GLE1	0.129				
	(0.117)				
GLW1	-0.064				
	(0.082)				
Dec 2011	0.048***	0.052***	0.044***	0.049***	0.048***
	(0.007)	(0.009)	(0.009)	(0.011)	(0.011)
Jan 2012	-0.015**	-0.018*	-0.020**	-0.023*	-0.023**
	(0.007)	(0.009)	(0.010)	(0.012)	(0.011)
Feb 2012	-0.044***	-0.046***	-0.049***	-0.049***	-0.048***
	(0.010)	(0.013)	(0.014)	(0.017)	(0.016)
Mar 2012	0.041***	0.039***	0.041***	0.044***	0.043***
	(0.007)	(0.008)	(0.008)	(0.009)	(0.009)
Nov 2012	0.027**	0.030**	0.022*	0.027*	0.027**
	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)
Dec 2012	0.014	0.012	0.011	0.016	0.017
	(0.012)	(0.013)	(0.012)	(0.013)	(0.013)
Jan 2013	0.037***	0.037***	0.036***	0.032**	0.034**
	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)
Feb 2013	-0.067***	-0.074***	-0.073***	-0.080***	-0.080***
	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)
Mar 2013	-0.008	-0.006	-0.011	-0.002	-0.002
	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)
IMD score	0.001	-0.001	0.006*	0.005	0.005
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
GPs per 1,000 patients	0.012	0.049	-0.028	-0.029	-0.050
	(0.047)	(0.053)	(0.046)	(0.051)	(0.050)
Patients females	2.219***	1.750**	3.464***	3.503***	3.608***
	(0.759)	(0.840)	(0.772)	(0.947)	(0.927)
Patients 4-14 years old	2.087***	2.472***	3.325***	4.051***	4.011***
	(0.732)	(0.762)	(0.814)	(0.845)	(0.828)
Patients 45-64 years old	2.175***	2.185**	0.239	0.593	1.006
	(0.841)	(0.880)	(0.789)	(0.770)	(0.777)
Patients 65-74 years old	3.185	3.125	8.490***	7.757***	6.384***
	(2.150)	(2.509)	(2.017)	(2.203)	(2.285)
Patients 75-84 years old	7.356**	6.782*	6.293*	4.954	4.513
	(3.174)	(3.779)	(3.478)	(3.942)	(3.663)
Patients over 85 years old	-0.261	1.266	-5.651	-2.631	0.138
	(4.782)	(5.101)	(5.406)	(5.820)	(5.741)
GP females	0.019	0.012	-0.021	-0.034	-0.034
	(0.044)	(0.051)	(0.050)	(0.059)	(0.056)
GP up to 30 years old	-0.000	-0.077	0.015	-0.014	0.017
	(0.152)	(0.176)	(0.173)	(0.194)	(0.210)
GP 30-49 years old	0.016	0.034	-0.001	0.029	0.021
	(0.058)	(0.069)	(0.059)	(0.064)	(0.059)
GP 50-64 years old	-0.011	0.012	-0.032	-0.010	-0.005
	(0.050)	(0.059)	(0.051)	(0.055)	(0.055)
GP qualified outside UK	-0.023	0.025	-0.039	0.015	0.038
	(0.052)	(0.052)	(0.056)	(0.062)	(0.061)
constant	4.404***	4.593***	3.720***	4.228***	4.171***
	(0.361)	(0.403)	(0.386)	(0.533)	(0.519)
MSOA dummies	✓	✓	✓	✓	✓
Observations	7834	5845	5374	4110	4205
Adjusted R^2	0.687	0.669	0.662	0.643	0.646

Notes: Cluster-robust standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$

Table A.3.: Trial effect on cardiovascular medication spending per 1,000 patients by treatment regions.

	All	GLE1	GLE2	GLE3	GLW1
$TRIAL \times TREAT$	0.020	0.040*	-0.036*	0.072***	0.048
	(0.017)	(0.022)	(0.020)	(0.019)	(0.031)
treatment region		0.312***	0.470***	-0.487***	-0.684***
		(0.084)	(0.085)	(0.169)	(0.161)
GLE3	0.067				
	(0.108)				
GLE2	0.082				
	(0.108)				
GLE1	0.285***				
	(0.080)				
GLW1	-0.214***				
	(0.059)				
Dec 2011	0.030***	0.033***	0.023***	0.023***	0.024***
	(0.006)	(0.008)	(0.006)	(0.007)	(0.007)
Jan 2012	0.008	0.007	0.009	0.003	0.005
	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)
Feb 2012	-0.026***	-0.027***	-0.033***	-0.034***	-0.035***
	(0.006)	(0.007)	(0.007)	(0.008)	(0.008)
Mar 2012	0.056***	0.057***	0.054***	0.055***	0.055***
	(0.006)	(0.007)	(0.007)	(0.008)	(0.008)
Nov 2012	-0.396***	-0.391***	-0.396***	-0.394***	-0.394***
	(0.012)	(0.013)	(0.012)	(0.012)	(0.012)
Dec 2012	-0.425***	-0.426***	-0.422***	-0.421***	-0.421***
	(0.012)	(0.013)	(0.012)	(0.012)	(0.012)
Jan 2013	-0.420***	-0.421***	-0.422***	-0.427***	-0.423***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Feb 2013	-0.510***	-0.513***	-0.515***	-0.520***	-0.520***
	(0.012)	(0.013)	(0.012)	(0.013)	(0.013)
Mar 2013	-0.410***	-0.410***	-0.414***	-0.411***	-0.409***
	(0.014)	(0.014)	(0.014)	(0.015)	(0.015)
IMD score	0.004	0.004	0.005	0.006*	0.007*
	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)
GPs per 1,000 patients	-0.098*	-0.092	-0.006	-0.007	-0.031
	(0.056)	(0.065)	(0.046)	(0.054)	(0.053)
Patients females	0.324	0.047	0.885*	0.764	0.823
	(0.534)	(0.625)	(0.492)	(0.558)	(0.587)
Patients 4-14 years old	2.084***	2.232***	2.922***	3.249***	2.988***
	(0.777)	(0.858)	(0.631)	(0.657)	(0.718)
Patients 45-64 years old	2.984***	2.982***	1.397**	1.610***	2.158***
	(0.789)	(0.907)	(0.539)	(0.567)	(0.590)
Patients 65-74 years old	4.830***	5.057**	7.570***	7.088***	6.211***
	(1.797)	(2.115)	(1.728)	(1.853)	(1.801)
Patients 75-84 years old	11.361***	11.759***	12.113***	12.474***	11.139***
	(2.317)	(2.469)	(2.700)	(2.825)	(3.094)
Patients over 85 years old	-4.597	-5.529	-7.395*	-7.247*	-4.696
	(3.507)	(3.880)	(3.843)	(4.381)	(4.889)
GP females	-0.072*	-0.107***	-0.076*	-0.114**	-0.110**
	(0.037)	(0.041)	(0.041)	(0.046)	(0.044)
GP up to 30 years old	-0.026	-0.000	-0.114	-0.037	0.017
	(0.111)	(0.139)	(0.117)	(0.136)	(0.143)
GP 30-49 years old	0.004	0.069	-0.060	0.005	-0.005
	(0.048)	(0.054)	(0.049)	(0.053)	(0.056)
GP 50-64 years old	-0.051	-0.034	-0.019	0.021	0.029
	(0.041)	(0.052)	(0.041)	(0.046)	(0.044)
GP qualified outside UK	0.008	0.035	-0.014	0.029	0.061
	(0.041)	(0.043)	(0.044)	(0.051)	(0.052)
constant	5.253***	5.278***	5.034***	5.396***	5.365***
	(0.254)	(0.300)	(0.265)	(0.334)	(0.339)
MSOA dummies	✓	✓	✓	✓	✓
Observations	7834	5845	5374	4110	4205
Adjusted R^2	0.829	0.824	0.809	0.801	0.793

Notes: Cluster-robust standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$