# Geo Health: A Unified Dashboard for Prescribing, Prevalence and Demographic Data

GeoHealth is a model for a public health dashboard set up for the purpose of presenting and connecting medical and social data from diverse sources. The current version is focused on England using public health and socio-demographic information available at the level of General Practice from around 8,000 surgeries. Their datasets, which were collected and prepared as part of this project, can be queried to produce a geo-spatial representation of the prevalence of diseases, risk factors and social deprivation on the national, regional and local levels.

The goal of this project is to support the monitoring of disease and risk factors on a national level enabling policy makers in public health policies and decision-making regarding treatment and prevention. NHS England and Public Health England released the data through the [General practice data hub](https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/general-practice-data-hub) and [National General Practice Profiles](https://fingertips.phe.org.uk/profile/general-practice/page-options/map-ao-4) and built long-standing dashboards used to monitor [drugs prescribing behaviour](https://openprescribing.net/) and the [prevalence or disease and risk factors](https://healthierlives.phe.org.uk/). Geo-Health, which is still in a development stage, was inspired by these efforts and should not be consulted as an alternative. The purpose is to suggest the development of a unified visualization platform, connect health, demographic and social [deprivation](http://dclgapps.communities.gov.uk/imd/iod_index.html) data, and support the development of machine learning, statistical and epidemiological tools.

A screenshot of a cell phone

Description automatically generated

### The bulk of the data for this project is the prescribing data, made available for all practices in England for each month (2018-2019). These datasets can support Drug Utilization Research ([x](https://www.researchgate.net/publication/301252985_Introduction_to_drug_utilization_research)) to quantify, understand and evaluate the processes of prescribing, dispensing and consumption of medicines and for the testing of interventions to enhance the quality of these processes thus improving the use of medicines in populations. The dashboard allows users to compare drug utilization (DU) across different geographical areas ([x)](https://www.researchgate.net/publication/301252984_Comparison_of_drug_utilization_across_different_geographical_areas), relate DU to the morbidity profiles of these areas ([x](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2323523/)) and examine the therapeutic effects of specific drug groups ([x](https://pubmed.ncbi.nlm.nih.gov/27659408/?from_term=%28%28morbidity%5BTitle%5D%29+AND+%28prescribing%5BTitle%5D%29%29+OR+%28drug+utilization%5BTitle%5D%29&from_sort=&from_pos=3)) and review the impact of prescribing policies such as the UK 5 Year Antimicrobial Resistance Strategy ([x](https://www.gov.uk/government/publications/uk-5-year-antimicrobial-resistance-strategy-2013-to-2018)).

Alongside prescribing the dashboard allows for querying the prevalence of disease ([x](https://academic.oup.com/jpubhealth/article/33/1/108/1543312))

and life-style factors such smoking ([x](https://bmjopen.bmj.com/content/4/7/e005217.short)) and obesity. The prevalence data is sourced from the Quality and Outcomes Framework (QOF) established to underpin indicators on quality of care. Since reporting is voluntary and due to variation in case finding, code definition and recording between practices, the QOF registers may not correspond to epidemiological standards (Technical annex, 2015-16, p. 9); In some studies they were identified as significantly underestimating morbidity of certain diseases as modelled or measured in health surveys ([x](https://bjcardio.co.uk/2005/11/hypertension-its-detection-prevalence-control-and-treatment-in-a-quality-driven-british-general-practice/),[x](https://www.sciencedirect.com/science/article/abs/pii/S0165032715001792),[x](https://www.bmj.com/content/337/bmj.a2030.full)). These factors make it difficult to interpret year-on-year changes or assess the actual prevalence of indications using QoF figures. Yet the discrepancies between reported and expected prevalence can be useful to identify geographical variation and under-diagnosis of prevalent and life-threatening conditions ([x](https://link.springer.com/article/10.1186/1478-7954-5-8#Sec8), [x](https://bmccardiovascdisord.biomedcentral.com/articles/10.1186/1471-2261-11-12)).

Users interested in the complex relations between prescribing and prevalence can select drug and indication, zoom in on specific areas or practices, NHS regions, sub-regions or Clinical Commissioning Group (CCG), that form the local unit for healthcare delivery in England. The plots in figure 1 connect the prevalence of diabetes with the quantity of anti-diabetic drugs (a,c) and the prevalence of depression (age 18+) with the quantity of antidepressants (b,d) prescribed in each practice. Plots a and b, drawn for Leeds, Bradford and Wakefield present prevalence and prescribing at a regional level whereas plots b and d show the data for a single CCG (Barnet, North London). Prevalence is marked by points radii while their colour indexes the quantity of drugs prescribed. The visualization of both allows the user to identify spots that merit examination for high prescribing levels or for under-diagnosis of the morbidity of interest.

A close up of a map

Description automatically generated

The efforts to contain the spread of Covid-19 offers another example for the potential use of Geo Health to inform prevention and containment policies. The NHS calls everyone to stay at home to help stop the spread of coronavirus but an emphasis is placed on people affected by certain chronic conditions, age 70 or older and significant overweight (BMI>40) identified as risk groups ([x](https://www.nhs.uk/conditions/coronavirus-covid-19/people-at-higher-risk-from-coronavirus/), [x](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30566-3/fulltext)). While they are advised to adhere to stricter and more cautious their geographical variation may inform preventive programs on the regional level and used to update the convention infection model. The variation in these conditions is demonstrated in figure 1, showing the patients registers for Coronary Heart Disease (a), Diabetes Mellitus (b), obesity (c) and chronic obstructive pulmonary disease (COPD; d) plotted using the prevalence data provided by the QoF. The patterns vary but show a general picture of increased morbidity in the north of England and coastal areas as compared to the London region.

A close up of a map

Description automatically generated

An area raising a concern can be explored in detail by combining medical and demographic data. The gender and age composition of practices was identified as strongly impacting the prevalence of conditions diagnosed and the drugs that are likely to be used ([x](https://www.bmj.com/content/311/7011/991?ijkey=56a152db34196e83aaea519e50ab62a7843fdd6a&keytype2=tf_ipsecsha)). While morbidity from cardiovascular disease, neurological diseases and diabetes increases with age, the greatest burden from mental health and substance use disorders is in the younger ages ([x](https://www.gov.uk/government/publications/health-profile-for-england-2018/chapter-3-trends-in-morbidity-and-risk-factors#trends-in-risk-factors)). Studies of the prevalence of coronavirus infections show that males experience higher case fatality rates than females ([x](https://www.sciencedirect.com/science/article/pii/S1201971220301363), [x](https://www.worldometers.info/coronavirus/coronavirus-age-sex-demographics/)), a pattern that is reflected in epidemiological data from the 2002–2003 SARS epidemic, recent Middle East respiratory syndrome outbreak and laboratory studies ([x](https://www.mendeley.com/viewer/?fileId=24e1d9e2-4c26-b6bf-57c5-cf881626a52e&documentId=ea90b4bb-3ea6-3317-a1c4-0f7687fa5c0d)).

To simplify the demonstration the next figure will focus on age and prevalence alone.

Figure 2 compares the prevalence (%) for obesity (points color) and for patients aged 70 or above (by circles’ size) for practices in the areas of Greater Manchester (a) and the North Midlands and between two cities there: Nottingham (b) and Stoke-On-Trent (c). The resulting plot suggests local and regional variation in demographic and morbidity patterns. In Nottingham the population that is both aging and obese appears on the city’s periphery while in Stoke they are located closer to the city center. The variation observed suggests a geographical distribution of infection risks that could inform containment programs in the city, region and sub-region levels.

A picture containing food

Description automatically generated

The monitoring of diseases and health factors can also be informed by the socio-economic profile of its area. Social deprivation was found as a major determinant of the increased prevalence of diverse conditions including hypertension, epilepsy, coronary heart disease, lung diseases, diabetes and various types of cancer ([x](https://www.bmj.com/content/337/bmj.a2030.full), [x](https://onlinelibrary.wiley.com/doi/full/10.1111/epi.12763), [x](https://academic.oup.com/jpubhealth/article/33/1/108/1543312), [x](https://www.sciencedirect.com/science/article/abs/pii/S0277953608005248), [x](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)00195-6/fulltext), [x](https://journals.sagepub.com/doi/full/10.1177/014107680710000613)) and higher levels or prescribing of antibiotics and drugs for various mental conditions ([x](https://academic.oup.com/fampra/article/22/1/37/440452), [x](https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(18)30053-1/fulltext)). Deprivation scores are calculated using small-area census variables whose values are normalised or transformed, standardised to local, regional or national means, and combined according to a weighting scheme ([x](https://ij-healthgeographics.biomedcentral.com/articles/10.1186/1476-072X-13-22)). The dashboard uses the Index of Multiple Deprivation (IMD) which is the official measure of relative deprivation in England and is calculated as an area-level weighed sum of measures of income (22.5%), employment (22.5%), education, skills and training (13.5%), health and disability (13.5%), crime (9.3%), barriers to housing and services (9.3%) and the living environment (9.3%) ([x](https://www.gov.uk/government/publications/english-indices-of-deprivation-2015-technical-report)).

Deprivation indices are usually sensitive to urban–rural differences ([x](https://ij-healthgeographics.biomedcentral.com/articles/10.1186/1476-072X-13-22)) but may be useful for comparing differences between and within cities to identify risk hotspots that call for focused attention. David McCoy professor of Global Public Health at Queen Mary University of London [criticized](https://www.theguardian.com/commentisfree/2020/apr/10/modelling-pandemic-politicians-decisions-science?CMP=Share_iOSApp_Other) the UK government for adhering uncritically to scientific advisors without considering the disproportionate effects of the virus on older people and men, and the “collateral damage of control measures will disproportionately affect the poorer and more marginalised segments of society.” Efforts to tackle these effects, towards, during and following the epidemics may be informed by exploring the geography of health in tandem with social deprivation.

Figure 4 explores COPD prevalence in the West-Midland region, focusing on Birmingham and the nearby city Wolverhampton. COPD prevalence is shown alone (a) and correlated with the index for health and disability (b) or index of barriers to housing and services (c). The former measures the risk of premature death and the impairment of quality of life through poor physical or mental health while the latter include accessibility in terms of road distance of shops, GPs, schools and Post Offices and affordable housing that includes household overcrowding and homelessness ([x](https://link.springer.com/article/10.1057/rt.2009.7)). While COPD prevalence does not exceed 5% of the population their normalization in relation to deprivation measures may point at inter-city locations where Covid-19 risks could be enhanced due to housing conditions, access to services or the health status of the population there.

A picture containing text, photo, snow, table

Description automatically generated

The dashboard is written in Python (version 3.7) using the [ipywidgets](https://ipywidgets.readthedocs.io/en/latest/) and [bokeh](https://docs.bokeh.org/en/latest/index.html) libraries for data visualization and user interactions. The data is held in a [PostgreSQL](https://www.postgresql.org/) database hosted on [AWS](https://aws.amazon.com/rds/) in a private account that is kept open with a read-only access to all users. Though not available as a web application, the dashboard can be downloaded and run in a [Jupyter Notebook](https://jupyter.org/) which is a convenient research environment for data science. The [code](https://github.com/ronyarmon/geo_health) and data is freely available to support data exploration, connect to additional data source, and develop user interfaces if desired. The package is built using a modular architecture to allow analysts, developers and data scientists to query of the data using their favourite BI software, build statistical and machine learning models, add datasets and selection menus, adapt the dashboard to data from other countries or improve the user interface.