The Role of Non-Profits in Public Health Service Provision:

Evidence from 25,338 heterogeneous procurement datasets

Charles Rahal† and John Mohan‡

[[1]](#footnote-1)

†Leverhulme Centre for Demographic Science and Nuffield College, University of Oxford

‡Third Sector Research Centre, University of Birmingham

1st July, 2021

**Abstract**

The role and extent of external suppliers across statutory health insurance procurement processes varies widely and is a source of political contention throughout the modern world. We comprehensively examine the role of non-profit organisations in public health procurement within publicly funded healthcare which runs parallel to private provision in a ‘two-tier’ system. We build a unique ‘Big Data’ based pipeline which scrapes tens of thousands of heterogeneous accounting datasets from across a commissioning hierarchy; from centralised commissioning (NHS England), decentralised Clinical Commissioning Groups, and various types of local NHS Trusts. These datasets provide granular information on every element of procurement at the micro-level (where the value of a transaction is greater than twenty-five thousand pounds), mandated by transparency requirements introduced by David Cameron in 2010. We develop data scientific tools to parse these datasets and reconcile suppliers with institutional registers, with a focus on those registered on the Charity Commission for England and Wales (CCEW). The processed dataset contains over four hundred and forty-five billion pounds worth of commissioning across over 1.9 million rows of clean data. We estimate that approximately 1% at each level of procurement comes from institutions listed on the CCEW: a number relatively consistent across time, despite annual contractual patterns coinciding with the financial year. We show only a slight regional variation and analyse the ‘North-South’ divide. Linking to the International Classification of Non-profit Organizations (ICNPO) and undertaking natural language processing on charity objectives, we show a ranged involvement of multiple different types of charity classification, with more payments going to the ‘Social Services’ aggregate, but the highest cumulative values of payments going to the ‘Health’ aggregate. We analyse the distribution of contracts to charities across various sizes and ages, from grassroots to ‘Super Major’ non-profits, and analyse how these concentrations have varied over time. We conclude with a discussion of the sector more generally in relation to the effects of the controversial Health and Social Care Act of 2012 and the integration of the free market and volunteerism, otherwise known as the ‘Big Society’.

**Keywords:** *Healthcare Supply, National Health Service, Data Visualisation, Big Data*

# Introduction

Following a decade of austere policies combined with an unprecedented public health crisis in the form of a viral pandemic, publicly funded health systems have never felt more important nor more precarious. Having tried various (‘New Public Management') reforms, such as outsourcing, private finance initiatives, introducing co-payments, and relaxations in rules governing income variation (Ferlie *et al*., 2007), most governments are increasingly reliant on external providers to provide healthcare. Non-profits[[2]](#footnote-2) in particular produce a ‘social capital’ which is vital to economic and political maintenance of such systems (Fukuyama, 1995), and substantial concern and interest surrounds the development of social capital as a resource for economic development, regeneration, and empowerment (Smallbone et al., 2001). However, data on the scale, structure, and composition of the non-profit sector is often piecemeal at best, leading to what has been termed ‘open season for non-profit theory building’ (Salamon and Anheier, 1998). This is despite the fact that 19.4 million people are estimated to volunteer through a group, club, or organisation in during 2018/2019 in the UK, with nearly a million employed by the Third Sector (Hornung et al. 2020).

The English National Health Service (an entity distinct from NHS Scotland, HSC Northern Ireland and NHS Wales, ran by devolved governments as of 1999) is controlled by the Department of Health and Social Care (DHSC), which takes political responsibility for one of the world's largest employers. At present, the Department of Health passes the majority of the healthcare budget to NHS England. At the time of writing, NHS England is responsible for overseeing the commissioning of NHS services — commissioning some services itself — while allocating the majority of its budget to Clinical Commissioning Groups (with the allocations being contingent on socio-economic and demographic characteristics within each commissioning area). These ‘CCGs’ procure from a multitude of different organisations, such as large NHS Trusts (which provide hospitals and community services), General Practitioners, and from external organisations such as charities and various elements of the private sector.

The NHS has a long-standing relationship with external suppliers, such as with regards to the provision of pharmaceuticals and the construction of hospitals (Greener, 2011). This relationship must be viewed in conjunction with an increasing proportion of income being received by the various elements of the NHS from non-NHS sources, such as from car parking, the renting of retail space on hospital grounds, and international medical tourists (Exworthy and Lafond, 2020). A general overview of spending across the NHS is provided in Figure 1. Panel A shows that in 2018/2019, 100 of the 222 NHS Trusts were operating in deficit, with the average Trust operating at -£2.402m. Of the 36 countries tracked by the Organisation for Economic Co-operation and Development (OECD), Great Britain had the 8th lowest share of healthcare spending as a percent of GDP, and the 19th spend per capita (Panel B). Panel C shows differences in the financing of the NHS over time and in relation to different government regimes. Spending on healthcare as a percent of GDP (measured as a percent of Public Expenditure Statistical Analysis) is still to return to its pre-financial crisis allocations (a maximum of 7.29% in 2010). Finally, Panel D shows the distribution of funding across all CCGs in both absolute and per-capita terms: the average allocation in the 2019/2020 financial year was £358.271m, and the average per-capita value was £1,284.

[Figure 1 around here]

Resource allocation was delegated to NHS England (an arms-length public body) by the intensely controversial Health and Social Care Act 2012. A white paper entitled ‘Equality and Excellence: Liberating the NHS’ was published within two months of the Conservative-Liberal Democrat coalition coming into power (in 2010), despite David Cameron claiming in a pre-election speech that ‘there will be no more of the tiresome, meddlesome, top-down re-structures that have dominated the last decade of the NHS’. Led by Andrew Lansley, the then Secretary of State for Health, it proposed the removal of Strategic Health Authorities (SHAs) and Primary Care Trusts (PCTs) with CCGs and General Practitioner (GP) led commissioning consortia. Another controversial aspect of this policy-change was the enabling of NHS Trusts to earn up to 49% of their income from non-NHS sources. Leys and Player (2011) saw this as further evidence of the persistent erosion of a publicly funded health service. Later dubbed ‘Dr Lansley's Monster' by Delamothe and Godlee (2011) in a British Medical Journal editorial, the Act represented the most significant change within the institution since its inauguration in 1948, leading Sir David Nicholson — the NHS Chief Executive at the time — to infamously describe the change as so large ‘that you can actually see it from space’. It gained Royal Ascent on March 27th, 2012, coming into effect in April 2013. It had the primary intention of creating a diverse provider market for healthcare where NHS, voluntary and private sector organisations compete for contracts. It represents part of a wider political shift which devolved power from Whitehall to local communities, inherently positioning the market as the main mechanism for delivering public services.

It allows consortia to commission services from any licensed provider under the ‘Any Qualified Provider’ contractual system. It was intended that providers (be they from the NHS, the private, or non-profit sector) will compete on a ‘level playing field’ for NHS contracts, reducing barriers to entry while supposedly simultaneously offering considerable opportunities to the voluntary and community sector. While some progress appears to have been made in recent years regarding ‘silos’ of commissioning, a lot more work is required to eliminate piecemeal and uncoordinated efforts in healthcare commissioning. For example, Lord Carter of Coles' final report (Coles, 2016) sets out how non-specialist Acute Trusts can reduce unwarranted variation in productivity and efficiency across every area, with the objective of saving billions of pounds in commissioning efficiency every year.

Multiple bodies of theory are relevant in the formation of our understanding of the presence, scale, and composition of non-profits in the role of public healthcare supply. The most dominant, perhaps, is Market Failure Theory, which notes the inherent limitation of the state to supply sufficient quantities of ‘public goods’ (Weisbrod, 1977). Trust Theory is relevant as consumers of healthcare may face ‘information asymmetries’ surrounding their care and rely on a removed set of impartial institutions due to the inherent complexity of the problem (Krashinsky, 1986). Traditional Welfare State Theory might lead us to expect that the more extensive the state provision of public health, the smaller the non-profit sector would become. Salamon (1995) talks of the ‘paradigm of conflict’ with regards to Interdependence Theory, where potential sources of conflict are also potentially elements of partnership, while Salamon and Anheier (1998) argue for the usefulness of the term ‘social origins’ to describe the complex ‘embeddedness’ arising from constrained by prior patterns of historical development.

Prior to the implementation of the Health and Social Care Act 2012, the voluntary sector represented a substantial provider of health and social care and support, supplying a multitude of services to PCTs. As a sector, it is known for diversity and flexibility, and varies in composition from large multinational organisations (‘mega-charities’) to small community groups run largely by volunteers. Healthcare supply by the voluntary sector ranges from specialist clinical provision, where it often dominates the market — such as the hospice movement (Field and Johnson, 1992) — to disease-specific advice services (such as Diabetes UK) and general wellness support and advocacy. The sector is recognised for an ability to connect with ‘hard to reach’ (Flanagan and Hancock, 2010) parts of the community that the statutory sector finds difficult to integrate, often filling the gaps left by statutory provision. The notion of ‘hard to reach’ is a slightly contested yet important concept, commonly used within the spheres of social care and health, especially in within policy discourse surrounding health and social inequalities, and it is often ‘under the radar’ (Mohan, 2012a) organisations which undertake this role.

The statutory sector has historically worked with larger voluntary organisations because the procurement and administrative costs of dealing with multiple small organisations can be prohibitive and entails high risk (Baines et al. 2009). There were initial concerns that the new system would allow larger organisations to crowd out the smaller ones due to their ability to invest, take risks, and carry forward losses, leading to a lack of representation of the sector. Pre-Health and Social Care Act work by the National Council for Voluntary Organisations (NCVO) estimated that just under a quarter of England’s voluntary and community organisations were involved in the supply of adult health and/or social care and support services (Clark *et al*., 2010), with £3.39bn being allocated by the statutory sector. More recent estimates from the NCVO (Hornung *et al*., 2020) are that £15.7bn of the sector's £53.5bn (29.35%) came from ‘Government’ in 2017/2018. The remaining funding comes from individual donations, private sector contributions, National Lottery grants or otherwise internally generated incomes.

Parallel to these legislative and policy changes, the UK government was one of the first to realise the power of ‘Open Data’ as early as in 1998 when the Cabinet Ofﬁce published a key policy document entitled ‘Crown Copyright in the Information Age’. The UK leads the world in transparency, regularly topping international rankings such as the Open Data Baraometre (The Web Foundation, 2016): an important ambition to realize given that roughly one in every three pounds is spent by the public sector is spent on procurement (Davis *et al*. 2018). The specific research design in this paper is made possible through the uniquely progressive transparency policies outlined in a letter sent by David Cameron to Secretaries of State on May 31, 2010 (Cameron, 2010). This letter had the objective of revolutionising transparency in government by ‘opening up data’. Later guidance from HM Treasury (2010) confirmed that this agenda applied to all parts of central government as defined by the Office for National Statistics including, critically, all levels of the NHS commissioning chain from centralised government departments, Clinical Commissioning Groups, and finally individual NHS trusts.[[3]](#footnote-3)

Specifically, the letter and guidance required the publication of information on expenditure over £25,000, with rows of data including information required on the supplier, the date of transaction, the transaction value, and many other auxiliary fields. The data should be published by working day 15 after each month end and should not contain information on payments made which would compromise personal or national security (in accordance with the Data Protection Act of 2018 which enacts the General Data Protection Regulation into UK law). All data is provided under the Open Government Licence v3.0 (The National Archives, 2001), unless otherwise stated.

In terms of identifying manually entered and approximate supplier names on individual institutional registers, the task of matching *n* string literals to a unique resource identifier likely dates to Newcombe *et al.* (1959). The Company, ORganization and Firm name Unifier (CORFU) approach of Alvarez-Rodriguez *et al.* (2015) reconciles the suppliers of public procurement with supplier names in Australia between 2004-2012 (containing 77,526 unique names across 430,188payments). In similar work, Mynarz *et al.* (2015) outlines the different, yet interrelated tasks of data extraction, presenting a ‘Public Contracts Ontology’ at both a national and EU-level. The MOLDEAS project – which developed a linked pan-European e-procurement platform (outlined in Alvarez-Rodriguez *et al*., 2012) – addressed the matching task using techniques such as spreading activation and resource description framework (RDF) classification. The most similar work in terms of method and application which this work builds on are Rahal (2018) and Rahal (2019). The former develops an open-source methodological pipeline with regards to analysing data made available by the Department for Communities and Local Government (2015) — specifically in terms of reconciliation of supplier data against multiple registers — with applications to the third sector. The latter develops the *centgovspend* toolkit[[4]](#footnote-4) which analyses central government procurement data and provides three applications of what such data makes possible in the form of analysing procurement across Standard Industry Classifiers, stratification across those who supply government, and an application to board interlock across suppliers. In terms of pure application, the Institute for Government (Davis *et al*. 2018) uses data from the Spend Network to provide a broad, macro-level overview as to what the UK government procures. We believe that the current paper offers innovation in advance of all of the aforementioned, and despite the presently uncertain future of Clinical Commissioning Groups, believe that our approach offers a significant and substantive insight into the potential role of the Voluntary and Community Sector moving forward.

We merge these important and related concepts; the effect of a contentious policy change in the form of the Health and Social Act, advances in transparency regimes, and the provision of ‘Open’ public data, and the recent development of computational tools and approaches to match such data to institutional registers to draw conclusions on our substantive question of interest: what role does the non-profit voluntary sector play in the post-Health and Social Care Act public healthcare regime?

# Data and Methods

## 2.1 Raw Procurement Data

In addition to NHS England, every CCG and NHS Trust publishes their expenditure data in a different place, using a different format, and at different points in time (of which we consider 191 and 225 respectively, although there is substantial longitudinal fluidity). The challenge of harmonizing these unstructured data cannot be understated, and our complex pipeline for curating a final dataset for analysis is visualised in Figure 2. Sir Tim Berners-Lee (inventor of the World Wide Web and founder of the Open Data Institute) developed the ‘Five Stars of Openness’ rating system for open data transparency. Based on his criteria, most of the procurement datasets which we analyse would receive either a 1\*; “make your stuff available on the Web (whatever format)”, 2\*; “make it available as structured data (e.g., Excel instead of image scan of a table)”, or 3\*; “non-proprietary format (e.g. CSV instead of excel)” ratings at best (Berners-Lee, 2009).

We manually curated a look-up table for all 191 CCGs and all 225 NHS Trusts as of April 1st, 2020. Data is typically provided from around the time of the introduction of the transparency requirements, or the introduction of the Clinical Commissioning Group (as in existence in April 2020). The first step of the NHSSpend code library then sequentially attempts to scrape data for each of the 416 institutions, in addition to the data hosted by NHS England on its own (national) spending. Data was available in the public domain in *some* format for 189 CCGs and 205 Trusts, and for 68 months for NHS England. The data originates predominantly from each individual CCG and Trust website, but was also seldom hosted on data.gov.uk, and very rarely incorporated bulk downloads hosted on non-governmental organization websites because of previous Freedom of Information requests.

[Figure 2 around here]

A small number of Trusts did not publish data, and some publish the data in a format which is in a machine-readable format, but which is not easily read by a machine (and hence requires manual manipulation). Some Trusts regularly overlook the need to maintain links either on their homepage, or on data.gov.uk, leading to several ‘404 Not Found’ style errors. The files are provided almost exclusively as .pdf, .csv and .xls or .xlsx format. A small fraction was provided in.xlsb or.xlsm format. In total, our data curation process curates a total of 25,338 raw procurement datasets (12,709 CCG files, 12,561 NHS Trust files and 68 files from NHS England) which get passed as inputs into our parsing module. We implement a custom pdf-parsing module which is based — but substantially expands **—** on the *pdfminer3* library (with a specific purpose on procurement pdfs). The purpose of this is to convert the raw Portable Document Formats (.pdf files) into machine readble (‘flat’ or ‘tabular’) files. If multiple files are provided by the same organisation for the same month (containing the same data), we use only the more ‘open’ file-type which corresponds to the highest value based on the Five Stars of Openness’ rating system in order to eliminate duplicates. The official guidance HM Treasury (2010, p.10) states:

‘The files are to be uploaded in CSV file format. Microsoft Excel files should be converted to CSV. The CSV file must have precisely one header line with field names exactly as in the example file supplied.'

However, huge variation exists not only in blank lines at the top of machine-readable files (with missing data across various rows), but primarily in the language used to define fields of interest. We use manually curated lists and dictionaries to clean each file sequentially. A list of length 547 removes ambiguous and potentially conflicting headers. A dictionary with 162 keys replaces ambiguous headers to standardize into six key fields which are retained, where the standardised headers are; ‘transactionnumber’, ‘expensetype’, ‘expensearea’, ‘supplier’, ‘amount’, and ‘date’. Each row of each file must have a minimum of data in three key fields in order to be included in our database: a date of payment, the value of the payment, and the name of the supplier. Our parsing module detects and drops fields where large proportions of rows or columns are null, detects cumulative 'total' cell summations, and undertakes other such similar functionality in order to ensure validity of the parsed file. Table 1 shows the distribution of the file types across NHS Trusts, CCGs, and NHS England, in accordance with the Berners-Lee ratings. A surprising observation to note here from this data curation exercise is the number of files which have seemingly been converted *out of* .csv or .xls style spreadsheets and *into* .pdfs. Another is the seeming lack of desire to meet the Berners Lee criteria, where the barriers to saving a file as a 3\* format in comparison to a 2\* format are non-existent.

[Table One around here]

We crawl through the files one by one with our custom parser, and then clean the merged database for certain attributes. First, we drop information on payments under £25,000 for the small number of institutions which provide it, in order to standardize across all CCGs and Trusts. We drop any invoice amounts which are not integers or floats. We drop any payments which have dates which cannot be rationally coerced into a standardised dd-mm-yyyy format. Supplier names must be strings greater than length three (to avoid ambiguity). We check for the presence of eight substrings which indicate a redacted supplier name (indicated as 'redacted' in the supplier field): these indicate redacted payments which were not redacted prior to publication (but perhaps should have been), of which we find a total of 256 payments worth a total of over £14m. We find one payment which has the supplier name of ‘Various' (worth £54,996). We also drop rows where suppliers are purely numeric (but not alphanumeric). The cleaned process results (as seen in Table 1) in a total of 1,909,925 rows (646,938, 661,226 and 601,761 for CCGs, Trusts and NHS England respectively) of procurement data worth a total of £445,863m (£204111m, £86932m and £154821 for CCGs, Trusts and NHS England respectively). Note, importantly, that while all three of our substantive databases contain same number of rows (with shares of 33.387%, 34.62%, and 31.51%), the distribution of payment values across the three datasets is less evenly distributed (45.78%, 19.5% and 34.72%), commensurate with the hierarchical commission process.

## 2.2 Registers for Record Linkage

Multiple registers for record linkage are essential to understand the direction of payments and to eliminate false positive matches. The task is to map these messy, approximately entered supplier names where (*M*) multiple fuzzy strings can relate to one distinct identity (*N*): an *M* **→** *N* problem, where *M*>*N*.[[5]](#footnote-5) We use three key institutional registers. building on the framework of Rahal (2018). An auxiliary function of our library first downloads and processes the full list of unique companies presently active on Companies House (the United Kingdom's registrar of companies), as indexed by their ‘Free Company Data Product’. This — as with Charity Commission and NHS Digital Data — is fixed to versions online as of 1 April 2020, in order to provide consistency with the supplier names in the procurement data. This currently comprises 4,490,685 unique strings, and if a normalized version of the supplier string (of which there are 4,489,992, see Section 2.3) is duplicated, both are dropped to avoid any ambiguity. A similar procedure is followed for the Charity Commission for England and Wales, which is the non-ministerial government department that regulates registered charities in England and Wales and maintains the Central Register of Charities. The Register contains 469,691 unique raw supplier names and 463,733 normalized ones (discussed below), where each registered charity can have multiple names. This corresponds to 327,547 unique registration numbers. A final function obtains data from NHS Digital, the national provider of information, data and IT systems for commissioners, analysts and clinicians in health and social care in England, particularly those involved with the National Health Service.[[6]](#footnote-6) For the NHS Digital data, there are 45,739 unique raw entries, and 45,560 normalized ones.

The algorithmic procedure discussed in Section 2.3 describes how the reconciliation takes place between the unique raw suppliers and these institutional entities in the absence of a unique identifier within the data. As indicated in Section 2.3 and later in Figure 4, there is a substantial degree of overlap with multiple institutions legitimately appearing on all three indexes: not only is there is no unique resource identifier within the data to link out to institutional suppliers, but companies, charities and public bodies all frequently fall within multiple classification boundaries. For example, there are 51,575 normalised unique names on both of the Companies House and the Charity Commission, 1,408 on both of the Companies House and NHS Digital, and 1,280 on both of the Charity Commission and NHS Digital registers. A total of 371 normalised unique names occur on all three registers.

## 2.3 Reconciliation Approach

We take the set of unique names across the three registers (*N*=*n*1∪*n*2∪*n3*), where *n*1-*n*3 represent the three institutional registers. We then normalize this set (*N*) in order to enhance our reconciliation process. Our normalization function cases all characters within a supplier string, and uses a manually curated dictionary with 43 keys to expand abbreviations (such as ‘LTD’ into ‘LIMITED’) and removes all non-alphanumeric values. It then strips any residual ‘white space’. We then ingest both the set of (cleaned) raw and normalized institutional registrar entrants into a purpose built Elasticsearch indices. Once this data is ingested, we pass through as function inputs unique set of supplier names, a variant of which is normalized of length 42,998 and 19,060 respectively. Matches take place via Elasticsearch using Lucene's Practical Scoring Function for multi-term queries. We write a customised wrapper which customises Elasticsearch DSL, an existing high-level library whose aim is to help with writing and running queries against Elasticsearch, itself built on top of the official low-level client called elasticsearch-py. Lucene combines Boolean model (BM) of Information Retrieval with Vector Space Model (VSM) of Information Retrieval: documents ‘approved’ by BM are scored by VSM. Given a query term (*q*: in our case an NHS supplier), a ‘document’ (*d*: an entry on the institutional registers), and terms which form part of a multi-term query (*t*: sub-strings of supplier names split on white space), the Practical Scoring Function is defined as (Gormley and Tong, 2015):

where; coord(*q*, *d*) is a score factor based on the count of query terms found in the specified document, queryNorm(*q*) is a normalizing factor used to make scores between queries comparable, tf(*t* in *d*) correlates to the term's frequency, idf(*t*) stands for Inverse Document Frequency, t.getBoost() is a search time boost of term *t* in the query *q* as specified in the query text, encapsulates boost and length factors such as Field boost and lengthNorm. This represents an efficient package that collects matching documents and scores them as it goes. Following the calculation of the score(*q,* *d*) ∀ *q* ∈ *N* and *d* ∈ *M*, we extract the five highest scoring institutional register entries for each unique NHS supplier, both raw and normalised. For each of these five scores (*s*qk, where *k* = {1, ...,5}), of length |*q*| and |*s*qk| respectively, we compute the (more computationally intensive) Levenshtein distance (Levenshtein, 1966), which is given by :

where is the indicator function equal to 0 when and equal to 1 otherwise, and is the distance between the first *i* characters of *q* and the first *j* characters of . Figure 3 shows various subfigures which plot the relationship between the scores of the best and second best raw and normalized Elasticsearch supplier returns and their Levenshtein distances. Of particular interest, it shows the importance of stripping trailing whitespace and a strong positive correlation between the raw and normalised strings when using the multi-term based Practical Scoring Function. Following the creation of these arrays of scoring matrices, we undertake an extensive double-blind manual verification process of matching *all* NHS suppliers to the institutional entries, representing a supervised hybrid of the ‘automated\_safematch’ and the ‘manual\_verification’ methods as detailed in Rahal (2019).

In addition to the three main institutional registers which we reconcile our supplier names against, our manual verification exercise also allows us to identify ‘individual doctors’ and ‘individual named persons’ (which were perhaps erroneously not redacted at source). All residual unmatched entities are assigned a ‘No Match’ tag in our database. We take three further additional steps to clean the reconciled output. If an institution is matched both to the NHS Digital and the Charity Commission of England and Wales registers and is either a current or former NHS Trust, Clinical Commissioning group or Primary Care Trust, we manually set the assignment to NHS Digital exclusively. Secondly, we manually add one entry to the NHS Digital register entitled ‘NHS Supply Chain’, an operating model introduced following the report by Carter (2016) which aimed to introduce efficiency and productivity into the NHS. These payments (artificially mapped to the NHS Digital) are separate to the related Supply Chain Coordination Limited (SCCL), a limited company wholly owned by the Secretary of State for Health and Social Care. Finally, all historic payments to NHS Litigation are manually mapped to the renamed NHS Resolution entity (renamed as of April 2017).

## 2.4 Reconciled Dataset

The institutions with the highest value of received payments from the automated match procedure are ‘Sheffield Teaching Hospitals NHS Foundation Trust’, ‘HM Revenue and Customs’, and ‘Guys and St Thomas NHS Foundation Trust’ from CCGs, Trusts and NHS England respectively. The appearance of the former Trust is not surprising: it is one of the UK’s largest, busiest and most successful NHS Foundation Trusts, providing a full range of hospital and community services for people in Sheffield, as well as specialist care for patients from further afield (while managing five of Yorkshire’s best known teaching hospitals). Guys and St Thomas includes two of London's most famous teaching hospitals. The occurrence of these two types (the two trusts and HMRC) of entities (both reconciled to NHS Digital) as high value recipients is also indicative of another fact: that entities registered on NHS Digital (and the entities subsequently mapped to them) are comprised of two types of organisation: core components of the NHS itself, as well as Health authorities and support (executive) agencies. A visualisation of the reconstructed dataset with supplier reconciliation can be seen in Figure 4. An important caveat here is that with specific regards to the Charity and to a large subset of the highest value Company suppliers, we not only undertook a significant manual verification exercise pertaining to the superficial validity of the multiple approximate string matches, but a much deeper manual examination\database audit of whether the matches were pertaining to the indicated recipients within the Charity Commission (specifically) and Companies House datasets. Details of this are available in the supplementary material.

[Insert Figure 4 around here]

One of the main generative oversights of our approach can be seen in Figure 4 Panel A, which shows the distribution of matches across all six registers combined across both the CCG and NHS Trust datasets. It shows that our approach is able to map over 90.5% of the total value of payments in our database. In terms of unmapped payments (allocated to the ‘No Match’ category), the three most commonly seen supplier strings relate to either ministerial departments or executive agencies which are not listed on NHS Digital; ‘Inland Revenue CIS’ (a Construction Industry Scheme for contractors), the National Offender Management Service and the Ministry of Justice.

One immediate point of interest from Figure 4 Panel A is the large aggregate value of payments (20.6%) going to entities registered on the Companies House register. While this number is substantially larger than the 6.1-7.3% quoted in national government accounts as going to ‘Expenditure on non-NHS bodies’ (as from the Department of Health, and then more recently the Department of Health and Social Care, informing a fairly consensus opinion in modern media), this figure alone does not represent a ‘framed’ privatisation of the English National Health Service (Powell and Miller, 2014). This is for two reasons. The first is that these numbers in Panel A do not represent terminal spending: money from NHS England in our database flows to CCGs, and then on to NHS Trusts. The second is that a large number of high value recipients which map exclusively to Companies House but not NHS Digital are entities such as NHS Professionals (a limited company owned by the Department of Health set up to manage temporary staff banks on behalf of more than 55 NHS Trusts) and NHS Property Services (a limited company also owned by the Department of Health which owns around 3,600 National Health Service facilities at the time of writing).

# Procurement from the Community and Voluntary Sector

Our databases contain 36,639 (23,842, 9,008, and 3,789 for CCGs, Trusts and NHS England respectively) payments (with a minimum value of £25,000) to entities matched onto the CCEW, worth a total of £3.860bn (£1.980bn, £0.556bn, and £1.324bn respectively). While the proportion of suppliers from the VCS varies widely across the three hierarchical layers of procurement, the average value of payment to such a supplier is lower (£105k) compared to the entire concatenated database (£233k). Sections 3.1-3.4 provide four different avenues of analysis of the charities mapped to the. Section 3.1 analyses which non-profits provide which ­­­parts of the NHS Commissioning pipeline, and their characteristics. Section 3.2 categorises these based on either their manually curated International Classification of Non-profit Organizations (ICNPO) number, their ‘class’ as provided by the CCEW, or an analysis of the free-text description of the charity. Section 3.3 shows a geographical variation, and Section 4.4 provides temporal analysis.

## 3.1 Charity characteristics­­­

The most frequent suppliers which are mapped to each of our three procurement databases can be seen in Table 2. The appearance of Nuffield Health near the top of the rankings across each of the three databases is unsurprising, as it is not just the largest healthcare charity within the United Kingdom: it is generally regarded as the largest in absolute (in addition to Cancer Research UK and the National Trust)[[7]](#footnote-7).

[Insert Table 2 around here]

Similarly, St. Andrews Healthcare is a large independent charity which provides psychiatric services, and Hospice UK is a national charity working for those experiencing dying, death and bereavement. RAPT (the Rehabilitation of Addicted prisoners Trust) is now otherwise known as the Forward Trust following a relaunch in 2017 after a merger with Blue Sky. Some prominent charities which rank highly on this list focus on specific geographical areas (discussed in Section 3.3. Figure 5 shows four different comparisons of registration dates between those institutions on the CCEW which supply the NHS, and those that do not. It shows that in comparison to the entire CCEW, those NHS-supplying organisations are comparatively younger on average, with mean ages of 24.82 years, in comparison to 27.41 for the entire Commission. However, in comparison to the class of charities which identify as being for the ‘All Advancement of Health’ classification, most are relatively the same age.

[Insert Figure 5 around here]

Figure 6 shows another characteristic of NHS supplying organisations: their cumulative income both before and after the implementation of the Health and Social Care Act. Across the three tiers, those charities who supply the NHS are substantially larger, even when trimming the mean to consider the skewing effect of large mega-charities such as Nuffield Health. NHS England appears to procure from slightly larger other charities, which might be expected as NHS England commissions more from national, as opposed to potentially smaller, regional suppliers. In terms of the definition of Davis and Hornung (2018), we also decompose non-profits into ‘Micro’ (<£10k), ‘Small’ (£10k to £100k), ‘Medium’ (£100k to £1m), ‘Large’ (£1m to £10m), ‘Major’ (£10m-£100m) and ‘Super Major’ (>£100m). This is shown in Table 3, and — consistent with Figure 6 — shows an overwhelming majority of payments going to the larger, already established charities when either considering income received in the financial year of 2011 (before the implementation of the Act) and in the financial year of 2018. While only approximately 0.059% of all charities registered on the CCEW constitute ‘supermajor’ status in either of our two specified years of analysis, we estimate that about 1.894% of the charities which supply the NHS fall into this category: receiving upwards of 21% of the total number of payments, and upwards of 29% in terms of value of procurement.

[Insert Figure 5 around here]

[Insert Table 3 around here]

## 3.2 Activities of charity suppliers

Having outlined the stylistic characteristics of the suppliers in comparison to the NHS, we now explore which roles it is that they undertake across three separable strategies; their manually curated International Classification of Non-profit Organizations (ICNPO) number, their ‘class’ as provided by the CCEW, or an analysis of the free-text description of the charity. The International Classification of Nonprofit Organisations — designed by the Center for Civil Society Studies at Johns Hopkins University — is based on the International Standard Industrial Classification system embodied in national income accounting practice, modified to accommodate key components of the non-profit sector overlooked by the existing SIC system (Salamon and Anheier, 1992; Salamon and Anheier, 1996). We utilise the expansion designed by the National Council for Voluntary Organisations in the United Kingdom, which adds a small number of appropriate new categories[[8]](#footnote-8). With respect to Section 3.1, the NCVO themselves (Hornung *et al*., 2020) estimate that the ICNPO grouping of ‘Health’ (10%) has the largest proportion of large, major and super-major organisations compared to 4% of those in other sectors. Analysis by ICNPO number can be seen in Table 4. As might have been expected given the nature of the commissioning, and consistent with the other two modes of analysis in this section, it is those organisations within the Health classifications (3100, 3200, 3300 and 3400) and the umbrella Social Sciences categorisation (4100) which receive the largest number of payments and the cumulative highest amounts. Charities across almost every ICNPO receive payments, showing the vast diversity and the range of contributions made by the sector in general.

[Insert Table 4 around here]

There are two main differences in analysis concerning ICNPO categories and ‘classes’ provided in the raw CCEW data. The first is that ICNPO numbers are typically manually curated by researchers, while classes are self-reported upon registration. The second is that while ICNPO’s are unique, charities can register as multiple classes. This non-unique distribution is seen in Table 5. In addition to the ‘Advancement Of Health Or Saving Of Lives’ category, the majority of payments are made to service providers, and those which provide services to children, young people, and the differently abled. Figure 7 provides a visualisation of word frequency counts (after cleaning of stopwords and stemming each word) of the self-reported purposes of each charity submitted to the CCEW from charities supplying each of the CCGs, NHS Trusts and NHS England. Naturally, the terms such as ‘health’, ‘care’, ‘sick’ and ‘suffer’ appear frequently in NHS supplying non-profits, but not in the distribution of CCEW purposes overall. Such an approach also informs the relative procurement strategies in terms of substantive areas across the hierarchy, made possible through analysing the comparative frequency of stemmed terms such as ‘disable’. Further work in this trajectory would consider dynamic topic models to see how the involvement of non-profits in the NHS changes over time, which is something we consider from a slightly different perspective in Section 3.4.

[Insert Figure 7 around here]

[Insert Table 5 around here]

## 3.3 Geographical variation in the role of non-profits

Figures 8-9 show the geographical variation in the role of non-profits in healthcare supply for Trusts and CCGs for which we have more than 100 payments (over £25,000). While the Office for National Statistics provides a Shapefile for the catchment areas corresponding to Commissioning Groups, NHS England does not contain information on catchment areas by Trust.[[9]](#footnote-9) Figure 8 shows a substantial variation (0.1%-14.5%) in the percentage of payments by CCGs in procurement from the non-profit sector. This variation is far less when considering cumulative payment value (0.1%-4.0%). Considering geo-spatial variation in Trust procurement from the CCEW, there is a much greater variation in the distribution of amounts procured across all Trusts (0.0%-20.3%). The Trust which procured the most from a non-profit organisation in our database was Derbyshire Healthcare NHS Trust, which allocated 11.6% of its entire cumulative procurement on services from Phoenix House (a non-profit which provides supported accommodation for people with mental health issues). The provision of registered addresses for Trust headquarters allows us to spatially join each Trust into Nomenclature for Territorial Statistics (NUTS) regions. This aggregation highlights the national stability in the figures overall: while some specific NHS institutions might vary in their procurement, the average procurement from non-profits across the nine NUTS varies only between 0.96%-1.791% in terms of number of payments, and 0.374%-0.97% in terms of cumulative payment value. While the co-ordinates of an exact boundary which determines a ‘North-South divide’ in the UK is contentious[[10]](#footnote-10), we estimate that the procurement of services and supplies from non-profits is ever so slightly higher (lower) in the South (North), at 0.66% (0.62%) by value and 1.45% (1.37%) by count.

[Insert Figure 8 around here]

[Insert Figure 9 around here]

## 3.4 Temporal variation in the role of non-profits

The intention of the Health and Social Care Act 2012 was to open up the markets for supply to ‘external’ organisations, including, but not limited to the non-profit sector. One particular preliminary observation to make regards the cyclicality (or lack of) in terms of financial year. Payments across each ICNPO number are relatively consistent across procuring organisations and over time. The exception to this is the procurement from non-profits with the ICNPO number 3200 (‘Nursing Homes’), which receives 1.93% of all payments from CCGs in the month of April in comparison to a mere 0.274% on average for the remaining eleven months of the year, indicating the annual contractual nature of these services supplied at this level of the procurement chain.

[Figure 10 around here]

In terms of absolute change over time, Figure 10 shows this at both a high and low frequency of analysis. Despite some minor variation in the number of payments made by CCGs which involve procurement from non-profits (ranging from 4.7% to 2.9% per annum, as shown in Figure 10, Panel A.), there is a resounding level of consistency in the amount of payments made across the hierarchy of procurement. Clinical Commissioning Groups, for example, have at least 0.9% of their procurement on goods and services form the non-profit sector, and at most 1.1% (as shown in Figure 10, Panel H.). In addition to the annual contractual pattern mentioned above which can also be seen in Figure 10 Panel F., two other patterns emerge from the data. The first is the slight upward trend in the number of payments made to non-profits by NHS England (Figure 10 Panel A. and D.), moving from 0.4% in 2013 to 1.3% in 2019. The second is the ever so slight downward trend in the magnitude of payment value made by NHS Trusts to the non-profit sector. Figure 11 shows a form of concentration ratio: the ratio of the value of payments going to the top five, ten and twenty non-profit organisations which are receiving the largest value of payments per individual rolling window. Considering our combined dataset, this ratio has decreased from 0.425 to 0.283 for the five highest value, from 0.493 to 0.366 for the ten highest value, and 0.609 to 0.446 for the twenty highest value recipients between the years of 2013 and 2019.

[Figure 11 around here]

# Discussion

While there is some heterogeneity of procurement at the institutional (i.e. individual CCG and Trust) level, that there exists such little regional variation in the utilisation of non-profits may be considered surprising given that substantial regional variation exists in multiple other highly related aspects of health and society[[11]](#footnote-11). For example, considering the absolute differences in CCG allocations (and allocations per capital) across commissioning areas, we might otherwise have expected more substantial differences in levels of involvement of non-profits at the regional level, with huge and well documented disparities in regional health inequalities occurring throughout England (Ellis and Fry, 2010). Clifford et al. (2013) show important variations between organisations and localities in government funding of third-sector activity, and data from the Birch and Whittam (2006) emphasize the importance of regional socio-economic development specifically with regard to social embeddedness and social capital.

In addition to showing how the distribution of institutions supplying the NHS clusters across well defined terms of size, (with a focus on ‘Super-Major’ fee-charging healthcare charities), we show how these concentrations have varied over time, which is critically important given the commentaries and concerns surrounding the Health and Social Care Act of 2012. The reduction in these concentrations provide support in favour of the ‘opening up’ of the healthcare market and alleviate concerns that it would allow some small number of institutions to consistently consolidate their power. We are also able to contribute to the debate surrounding whether big charities are becoming more dominant (Backus and Clifford, 2012): there is no evidence that the biggest charities account for a growing share of total charity income received from the NHS over the period of analysis.

# Conclusion

An implicit feature of this study is that procurement and registration data have made it more of a study of ‘mega-charities’ rather than ‘micro-philanthropy’. Automating observational ‘Big Data’ analysis in the field is subject to substantial challenges; a widely quoted – albeit slightly outdated[[12]](#footnote-12) – estimate of the number of voluntary organisations in the UK is somewhere between 600,000 and 900,00 (MacGillivray et al. 2001), which is up to three times more than those currently registered by the Charity Commission for England and Wales. The dataset created herein however could be used to better map out ‘local hotspots’ (Lindsey, 2013, Mohan 2012a, Conservative Party, 2008) and substantiate or contest Iain Duncan Smith’s reference to so-called ‘charity deserts’. Such further analysis is even more important given that only approximately 45% of the institutions listed on the Charity Commission supply address data. Moreover, the methodological approach and the data provision described provide an avenue for academic research to be conducted until public units such as the Government Digital Service and NHS Digital make services such as the Spend Comparison Service complete, comprehensive, and fully open access. While the United Kingdom generally leads the world in terms of transparency in public processes, more work is needed to evaluate the evolution in response to policy changes such as not just the Health and Social Care Act, but also recommendations such as those by Lord Carter of Coles. That is without even beginning to consider the debate on the role of private, profit-making institutions.

# References

Alvarez-Rodriguez, J. M., Vafopoulos, M., and Llorensm, J., (2015), ‘Enabling Policy Making Processes by Unifying and Reconciling Corporate Names in Public Procurement Data. The CORFU Technique’, *Computer Standards and Interfaces* (**1**), pp. 28-38.

Backus, P. and Clifford, M. (2012), ‘Are big charities becoming more dominant? Cross‐sectional and longitudinal perspectives’, *Royal Statistical Society A: Statistics in Society*, 176 (**3**), pp. 761-776.

Baines, S., Bull, M., Woolrych, R. (2009), ‘Engaging Third Sector Suppliers to the NHS’, Paper presented at the Institute for Small Businesses and Entrepreneurship Conference, 3-6 November, Manchester, Manchester Metropolitan University.

Berners-Lee, T., (2016), 'Linked Data', published on https://www.w3.org/, accessed 3rd May, 2020.

Birch, K. and Whittam, G., (2008), 'The Third Sector and the Regional Development of Social Capital', *Regional Studies*, 42(**3**), 437-450.

Buchan, I.E., Kontopantelis, E., Sperrin, M., Chandola, T. and Dolan, T., (2017), ‘North-South disparities in English mortality 1965–2015: longitudinal population study’, *J Epidemiol Community Health*, 71, pp. 928-936.

Cameron, D., (2010), ‘Letter to Government Departments on Opening up Data’, published on <https://www.gov.uk/news>, accessed on 3rd May, 2020.

Clark J, Kane D, Wilder, K. and Wilton, J., (2010), ‘UK Civil Society Almanac’, *National Council for Voluntary Organisations*.

Clifford, D., Geyne-Rahme, F., & Mohan, J., (2013), ‘Variations between Organisations and Localities in Government Funding of Third-sector Activity: Evidence from the National Survey of Third-sector Organisations in England’. *Urban Studies*, 50(**5**), 959–976.

Coles, P., (2016), ‘Operational productivity and performance in English NHS acute hospitals: Unwarranted Variations’, An independent report for the Department of Health by Lord Carter of Coles.

Conservative Party, (2008), ‘A Stronger Society: Voluntary Action in the 21st Century’, Conservative Party, London.

Davis, N., Chan, O., Cheung, A., Freeguard, G. and Norris, E., (2018), ‘Government procurement: The scale and nature of contracting in the UK’, *Institute for Government*, (December).

Davis, J. and Hornung L., (2018), ‘Britains Biggest Charities: Key Features’, NCVO Research Briefing, Nov 2018.

Delamothe, T. and Godlee, F., (2011), ‘Dr Lansley's monster’, *BMJ,* **342**.

DCLG (2014). Transparency Code for Smaller Authorities, published on [https://www.gov.uk/](https://www.gov.uk/government/publications/), accessed on 3rd May, 2020.

DCLG (2015). The Local Government Transparency Code, published on <https://www.gov.uk/>, accessed 3rd May, 2020.

Dorling, D., (2007), ‘The North-South Divide - Where is the line?’, Social and Spatial Inequalities working paper, University of Sheffield.

Ellis, A., and Fry, R., (2010), ‘Regional health inequalities in England’, *Regional Trends*, 42, pp. 60–79.

Exworthy, M. and Lafond, S., (2020), ‘New development: Commercialization of the English National Health Service: a necessity in times of financial austerity?’, *Public Money & Management*, 0 (**0**), pp. 1-4.

Ferlie, E., Lynn, L., and Pollitt, C., (2007), ‘The Oxford Handbook of Public Management’. Oxford Handbooks in Business and Management. Oxford University Press.

Field, D. and Johnson, I., (1992), ‘Volunteers in the British hospice movement’. *The Sociological Review* (**40**), pp. 198-217.

Fukuyama, F., (1995), 'Trust: The Social Virtues and the Creation of Prosperity', The Free Press, New York.

Flanagan, S. and Hancock, B. (2010), ‘Reaching the hard to reach: lessons learned from the VCS (voluntary and community sector). A qualitative study’, *BMC Health Services Research*, 10 (**92**).

Garnick, D.W., Luft, H.S., Robinson, J.C., Tetreault, J. (1987), ‘Appropriate measures of hospital market Areas’, Health Services Research, 22 (**1**), pp. 69-89.

Gormley, C. and Tong, Z., (2015), ‘Elasticsearch: The Definitive Guide’, O'Reilly.

Greener, I. (2011), ‘The case study as history: `Ideology, class and the National Health Service' by Rudolf Klein’, In: *Shaping Health Policy Case Study Methods and Analysis*, pp.77-94.

HM Treasury (2010). ‘Transparency Publication of Spend over £25,000’, Published on <https://www.gov.uk/>, accessed 3rd May 2020

Hornung, L., Chan, O., Jochum, V., Lawson, McGarvey, M. and Rooney, K., (2020), ‘UK Civil Society Almanac’. National Council for Voluntary Organisations.

Jack, I., (1987), ‘Before the oil ran out: Britain 1977-1987’, London.

Krashinsky, M., (1986), 'Transaction costs and a theory of the nonprofit organization'. In S. Rose-Ackerman (ed.), 'The Economics of Nonprofit Institutions', Oxford University Press, Oxford and New York.

Leys, C. and Player, S. (2011). ‘The Plot Against the NHS’, 1st Edition, Merlin Press.

Levenshtein, V. I. (1966), ‘Binary codes capable of correcting deletions, insertions, and reversals’, *Soviet Physics Doklady*, **10**, pp. 707-710.

Lord Carter of Coles, (2016), ‘Operational productivity and performance in English NHS acute hospitals: Unwarranted variations’. An independent report for the Department of Health.

Mohan, J., (2012a), ‘The idea of a ‘charity desert’: challenges in and lessons from mapping the distribution of charities and charitable resources’, in *Will Our Philanthropy Give Us a better Society*, Centre for Charitable Giving and Philanthropy (Alliance Publishing Trust, London), pp. 51-56.

Mohan, J., (2012b), ‘Entering the lists: What can we learn about the voluntary sector in England from listings produced by local infrastructure bodies?’, *Voluntary Sector Review*, 3 (**2**), 197-215.

Mynarz, J., Svátek, V., and Di Noia, T., (2015), ‘Matchmaking Public Procurement Linked Open Data’, In: Debruyne C. et al. (eds) ‘*On the Move to Meaningful Internet Systems*’: OTM 2015 Conferences. OTM 2015. Lecture Notes in Computer Science, vol 9415. Springer, Cham.

Newcombe, H., Kennedy, J., Axford, S., and James, A., (1959), ‘Automatic linkage of vital records’, *Science*, 130 (**3381**), pp. 954-959.

O’Sullivan, J.W., Stevens, S., Oke, J. et al. (2018), ‘Practice variation in the use of tests in UK primary care: a retrospective analysis of 16 million tests performed over 3.3 million patient years in 2015/16’, *BMC Medicine* 16 (**229**), pp. 1-9.

Powell, M. and Miller, R., (2014), ‘Framing privatisation in the English National Health Service’, *Journal of Social Policy*, 43 (**3**), pp. 575-594.

Rahal, C., (2018), ‘The Keys to Unlocking Public Payments Data’, *Kyklos*, 71 (**2**), pp. 310-337.

Rahal, C. (2019). ‘Tools for Transparency in Central Government Spending’, *International Journal of Population Data Science*, 4 (**1**), pp. 1-10.

Rose, L., (2013), ‘Exploring local Hotspots and deserts: investigating the local distribution of charitable resources’, **Voluntary Sector Review**, 4 (**1**), pp. 95-116.

Salamon, L. M., (1995), 'Partners in Public Service: Government-Nonprofit Relations in the Modern Welfare State', The Johns Hopkins University Press, Baltimore, Maryland.

Salamon, L. and Anheier, H., (1992), ‘In search of the non-profit sector II: The international classification of non-profit organisations: a framework for analysis’, *Voluntas:* *International Journal of Voluntary and Nonprofit Organizations*, 3:3, pp. 267–309.

Salamon, L. and Anheier, H., (1998), ‘Social Origins of Civil Society: Explaining the Nonprofit Sector Cross-Nationally’, *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, 9:3, pp. 213–248.

Salamon, L. and Anheier, H., (1996), ‘The International Classification of Nonprofit Organizations: ICNPO’, Revision 1, Working Papers of the Johns Hopkins Comparative Nonprofit Sector Project, no. 19. Baltimore: The Johns Hopkins Institute for Policy Studies.

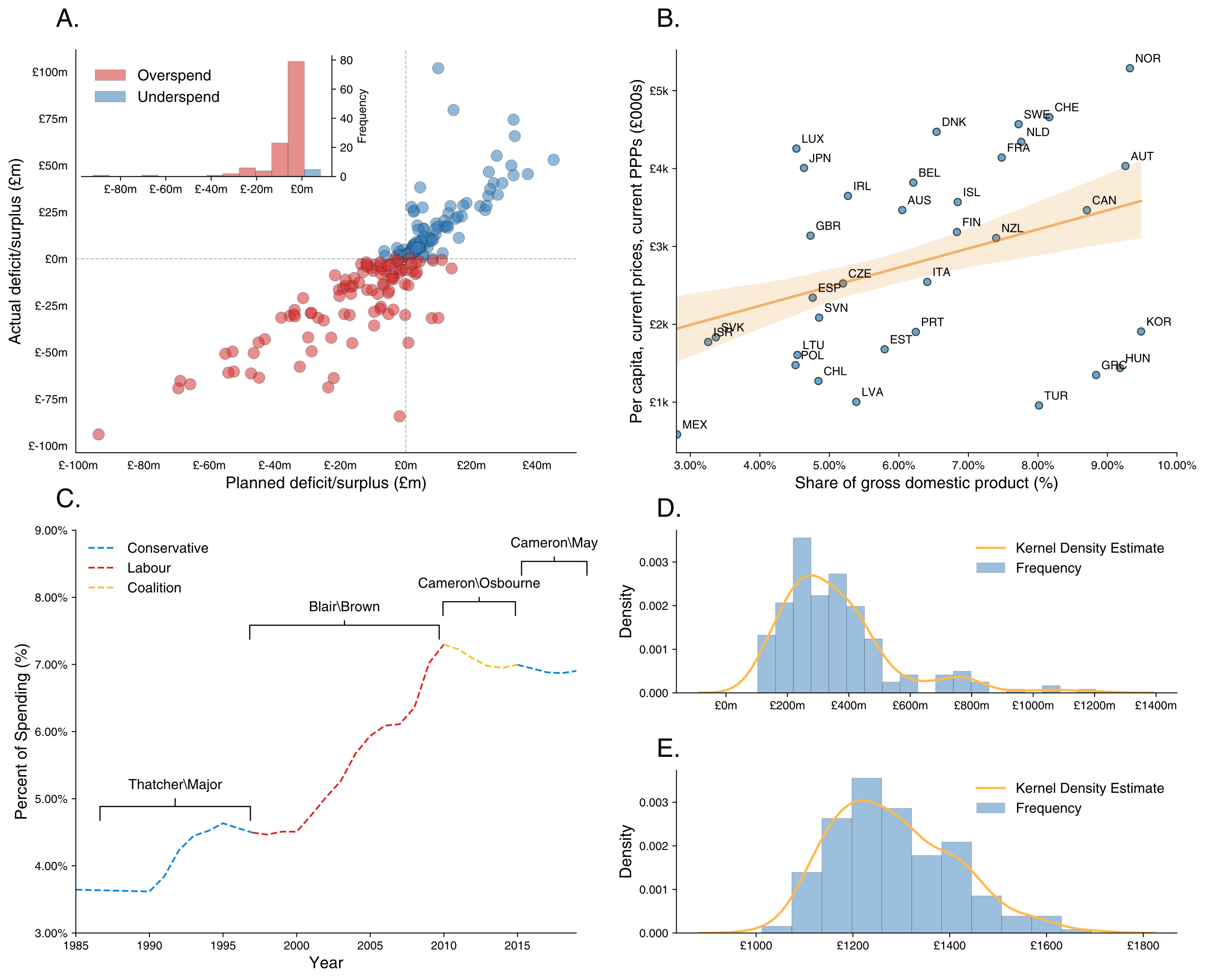
Smallbone, D., Evans, F­.,­ Ekanem, I., and Butters, S., (2001), ‘Researching Social Enterprise’, Small Business Service, London.

The National Archives, (2001), ‘Open Government License’, available at <http://www.nationalarchives.gov.uk>, accessed 3rd May 2020.

The Web Foundation, (2016), ‘The Open Data Barometer’, available at https://opendatabarometer.org/4thedition/, (4th Edition), accessed on 3rd May 2020.

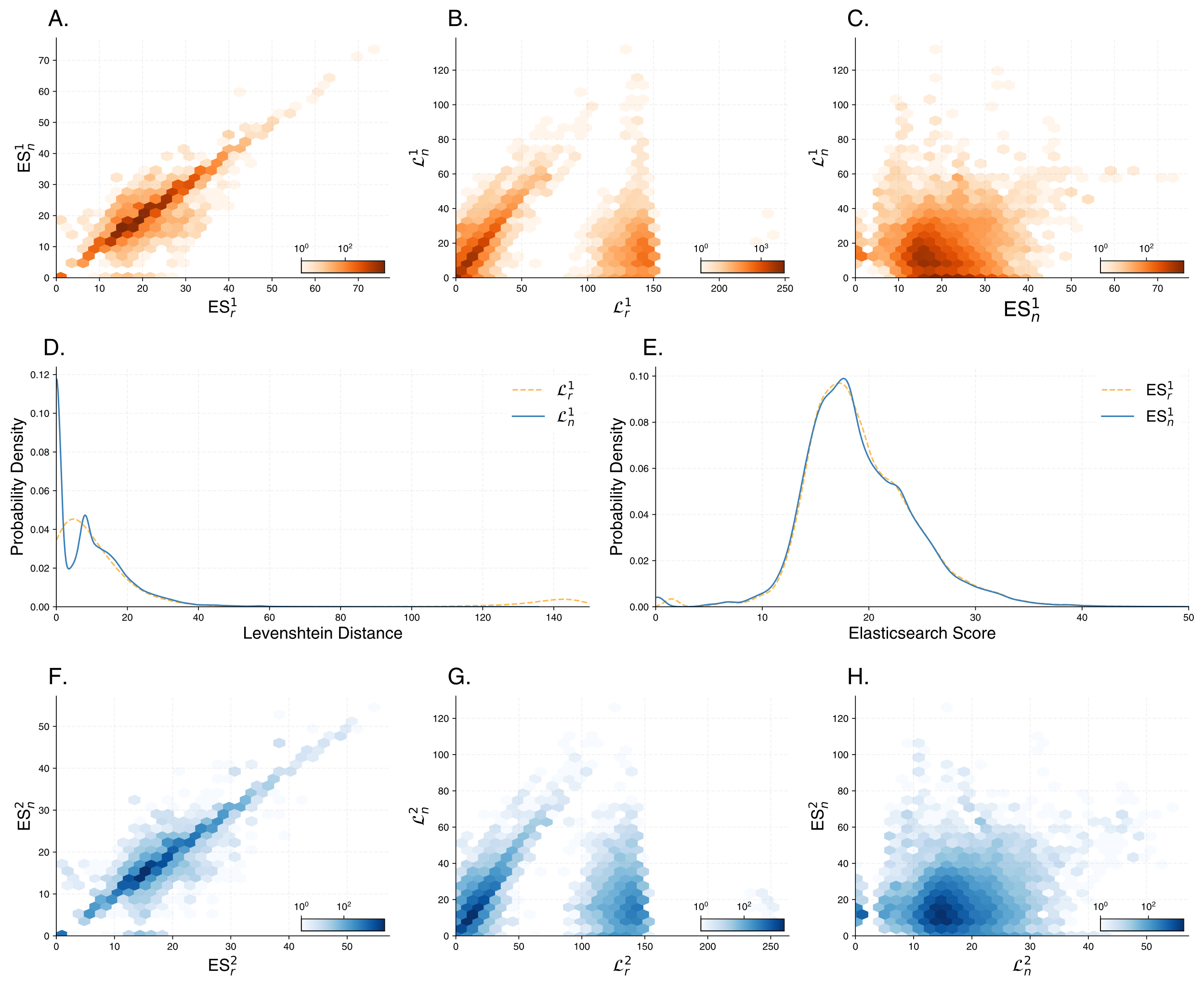
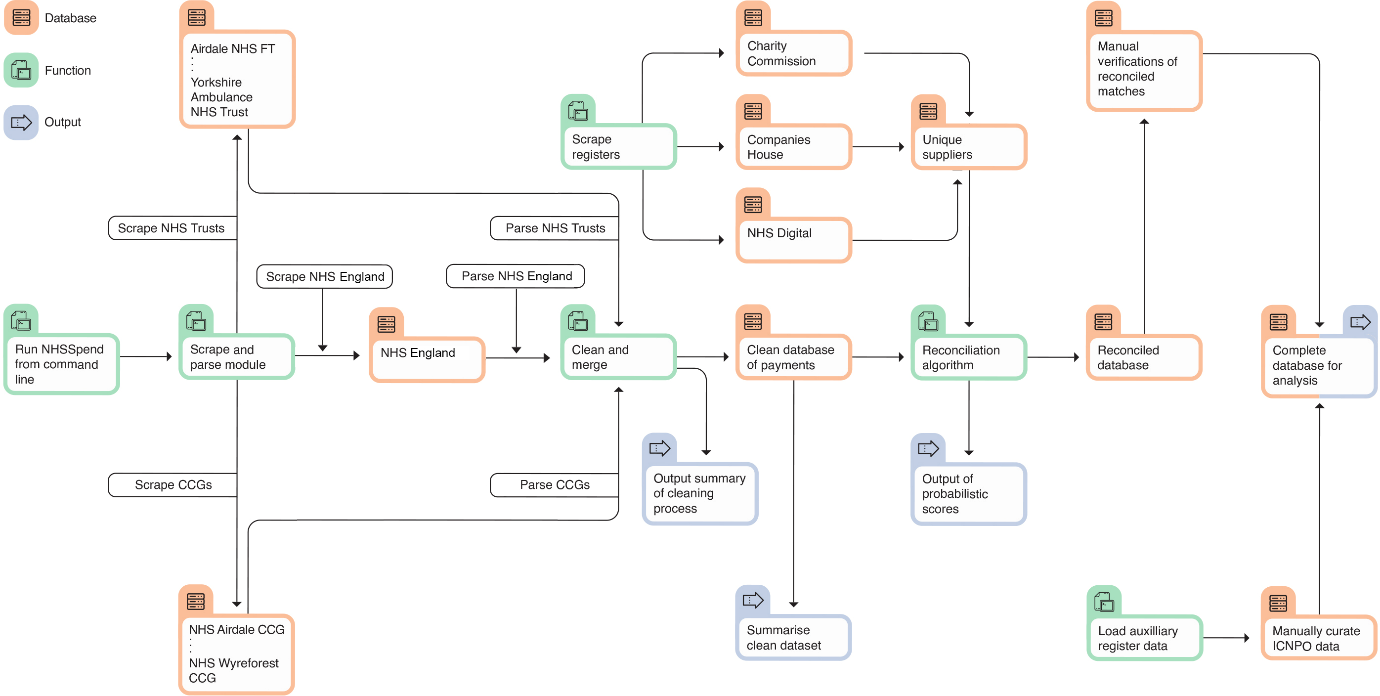
Weisbrod, B., (1977), 'The Voluntary Nonprofit Sector', Lexington Books, Lexington, Massachusetts.

Wells, J. C., (1982), ‘Accents of English: The British Isles’, Cambridge, Cambridge University Press.



**Figure 1: NHS Spending at the trust, CCG, national and international level**. Data for A. comes from NHS Improvement. Data for B. comes from the OECD. Data for C. comes from HMT Public Expenditure Statistical Analyses (PESA). Data for D. comes from NHS England.

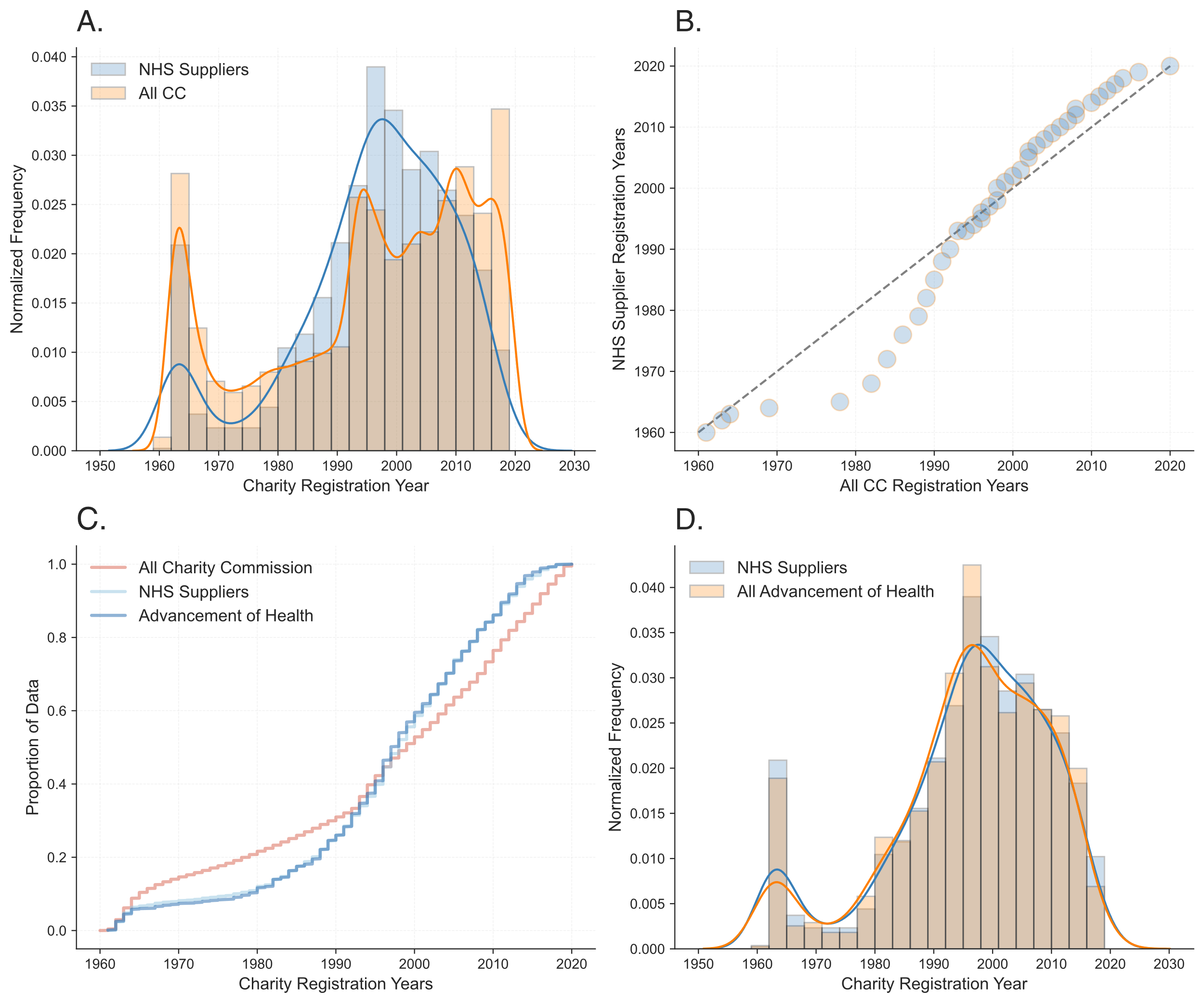
**Figure 2: Schematic detailing the operation of NHSSpend.** Green nodes identify functions and processes, orange nodes identify the creation of a database, and a blue node denotes the creation of an output of analysis on a database.



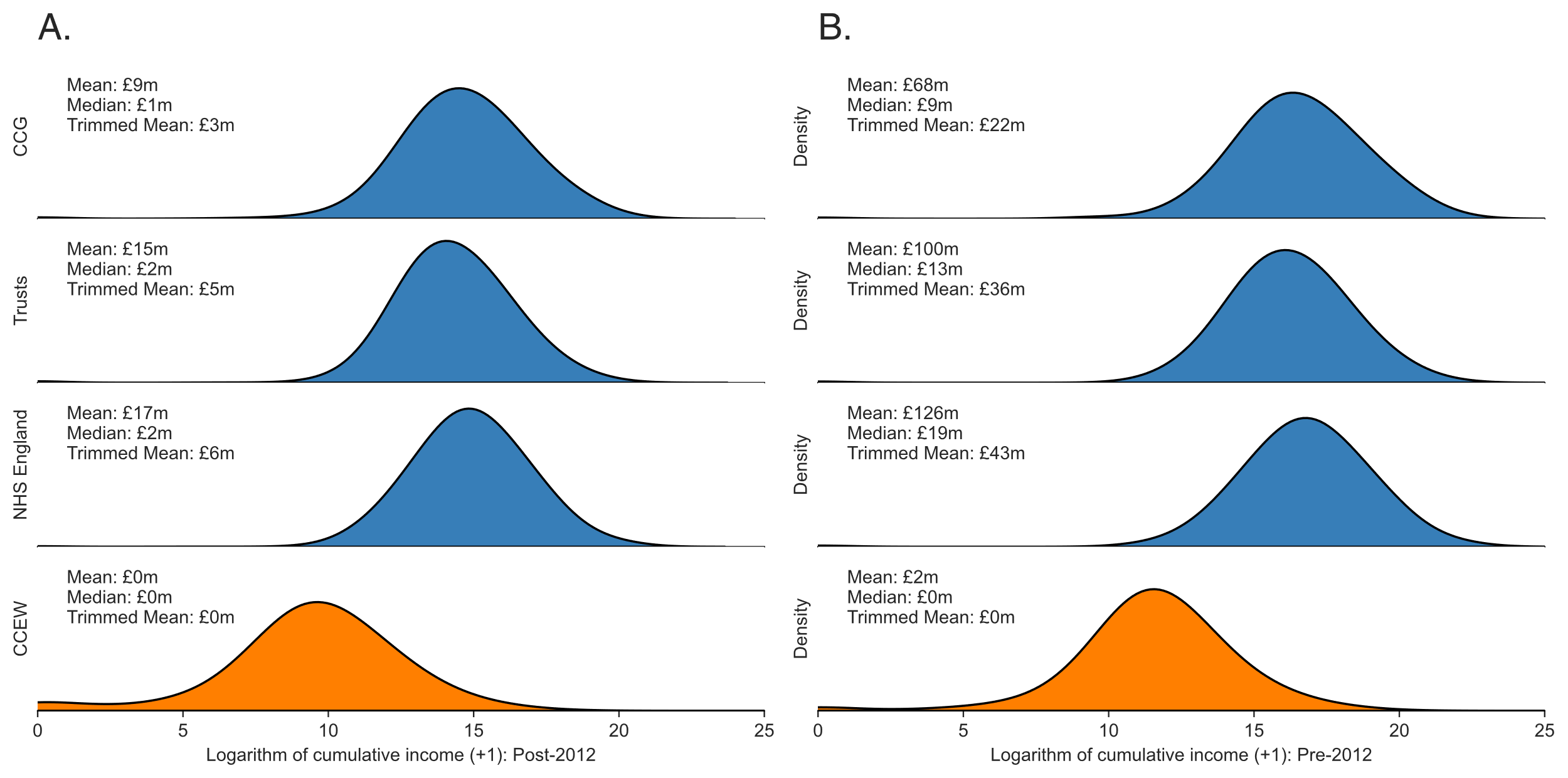
**Figure 3: Distribution of Reconciliation Scores**. ES refers to Elasticsearch, and Lev (L) to Levenshtein. The n subscript refers to scores on normalised entries, and r to raw entries. Subscripts 1 and 2 refer to the best and second-best matches respectively.



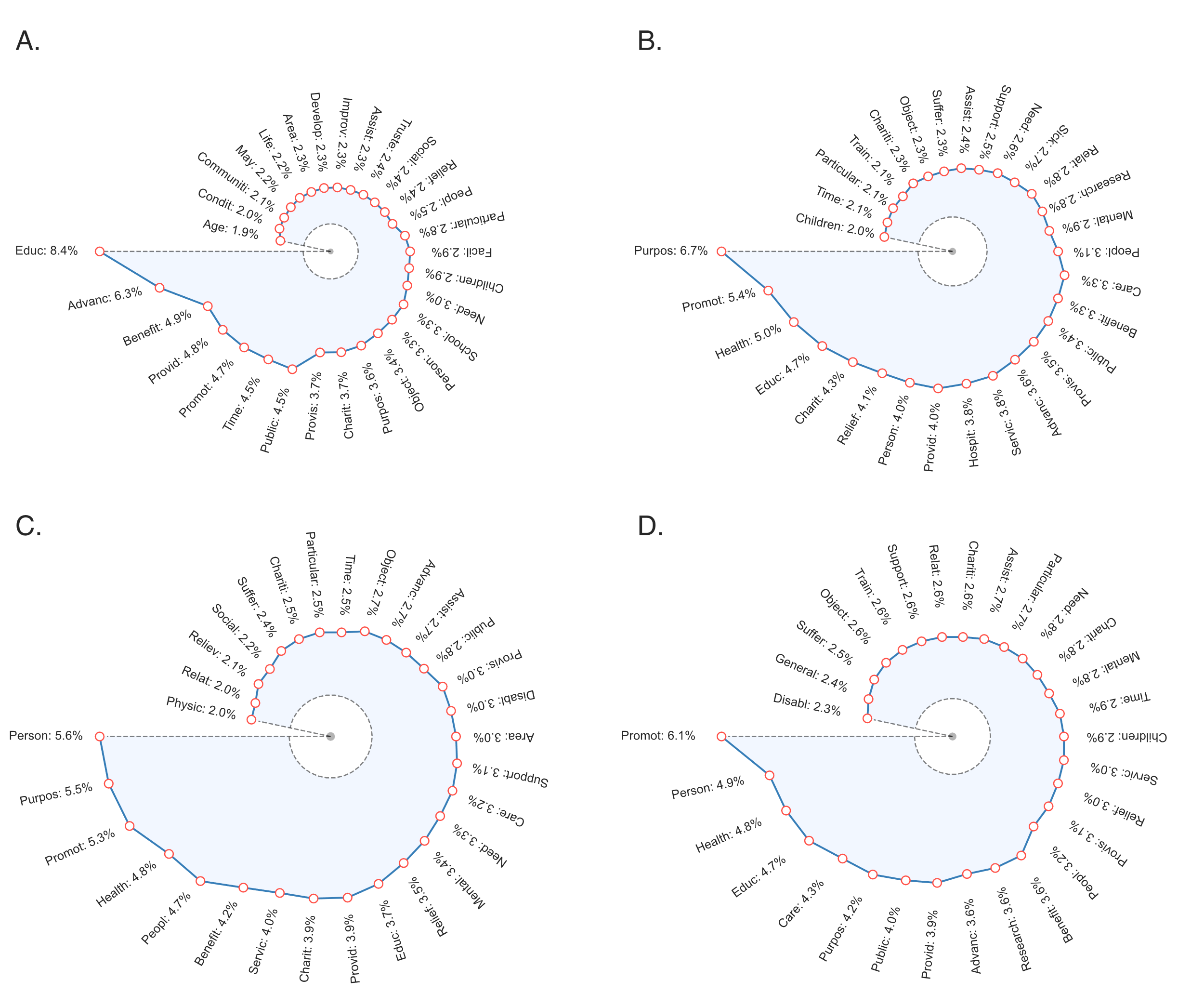
**Figure 4: Match distribution across institution type.** The doughnut chart (Panel A.) shows the cumulative value of all payments made to one specific sector (including where an institution features on more than one register). The Venn diagram (Panel B.) outlines the issue of institutional overlap between the named suppliers in our merged dataset. The two bar charts (Panel C.) document the unique mapping to institutions which are mapped to one of the six categories but not more.



**Figure 5: Analysing Charity Registration Dates:** Panel A. plots the distribution of years of years of registration of all entries on the CCEW, and the subset of those which supply the NHS. Subfigure B. is a Q-Q plot of the same two arrays. Panel C is an empirical cumulative distribution which also considers all charities which report as the ‘Advancement of Health’ class, which is also visualised in contrast to all NHS suppliers in Panel D.



**Figure 6: Income Distributions Across NHS Suppliers.** The figures show the logarithm of cumulative income as detailed on the Charity Commission for England and Wales, both post-2012 (Panel A.) and pre-2012 (Panel B.). The logarithm of cumulative income +1 is taken to deal with zero values on the CCEW. The trimmed mean is calculated by slicing 10% off either side of the distribution.

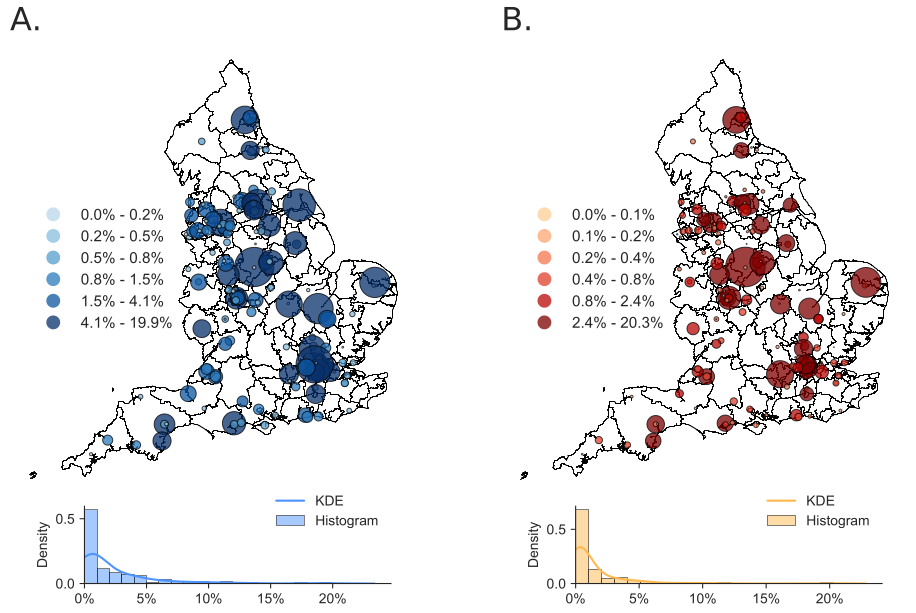


**Figure 7: Analysing supplier ‘Objects’.** The ﬁrst panel (A.) visualises the frequency distribution of cleaned, stemmed words found in the descriptive ambitions of all charities registered and active on the Charity Commission for England and Wales. Panels B.-D. visualise this for just the subset of charities which supply each of NHS Trusts, Clinical Commissioning Groups, and NHS England, respectively.

Map

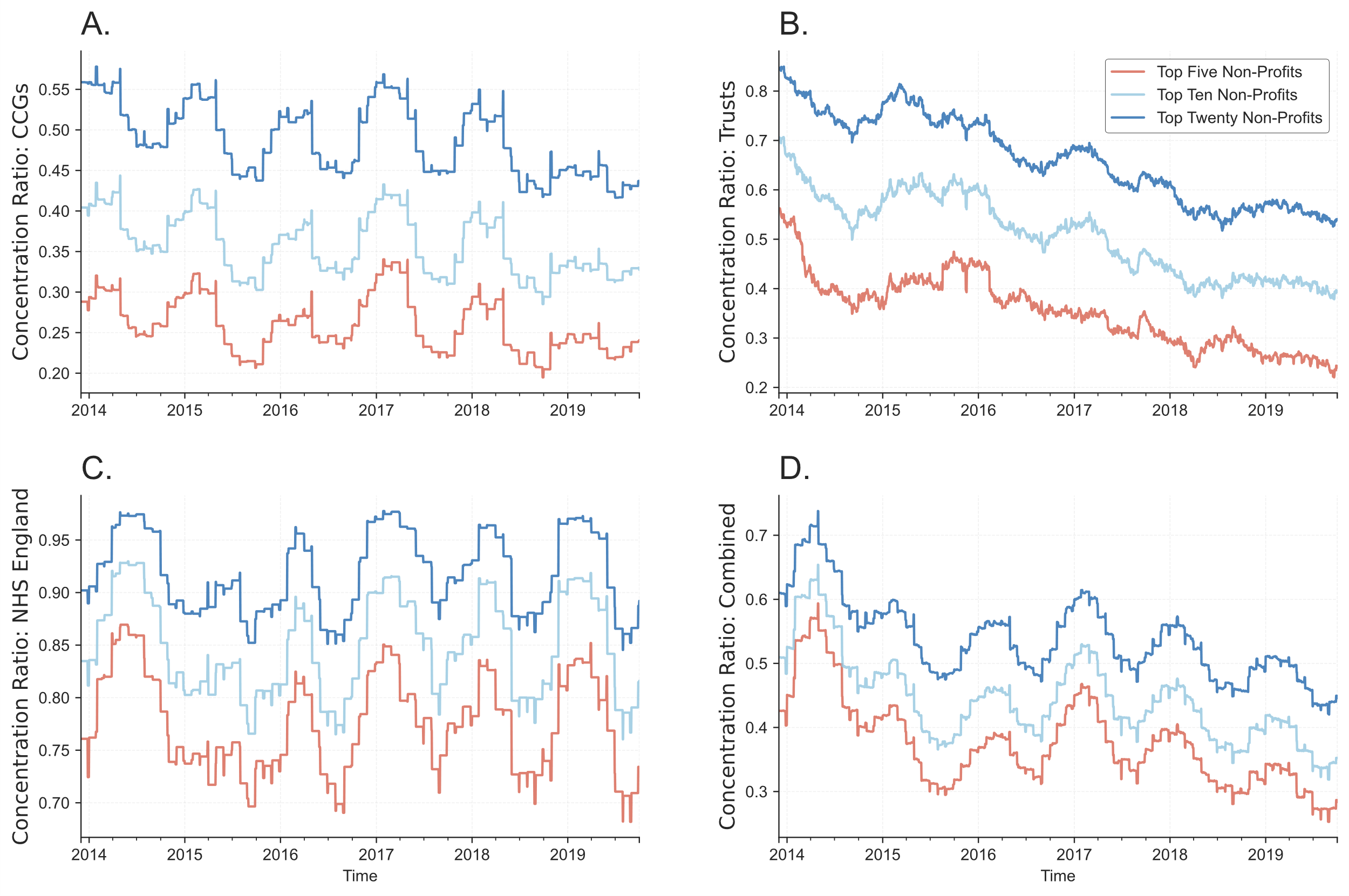
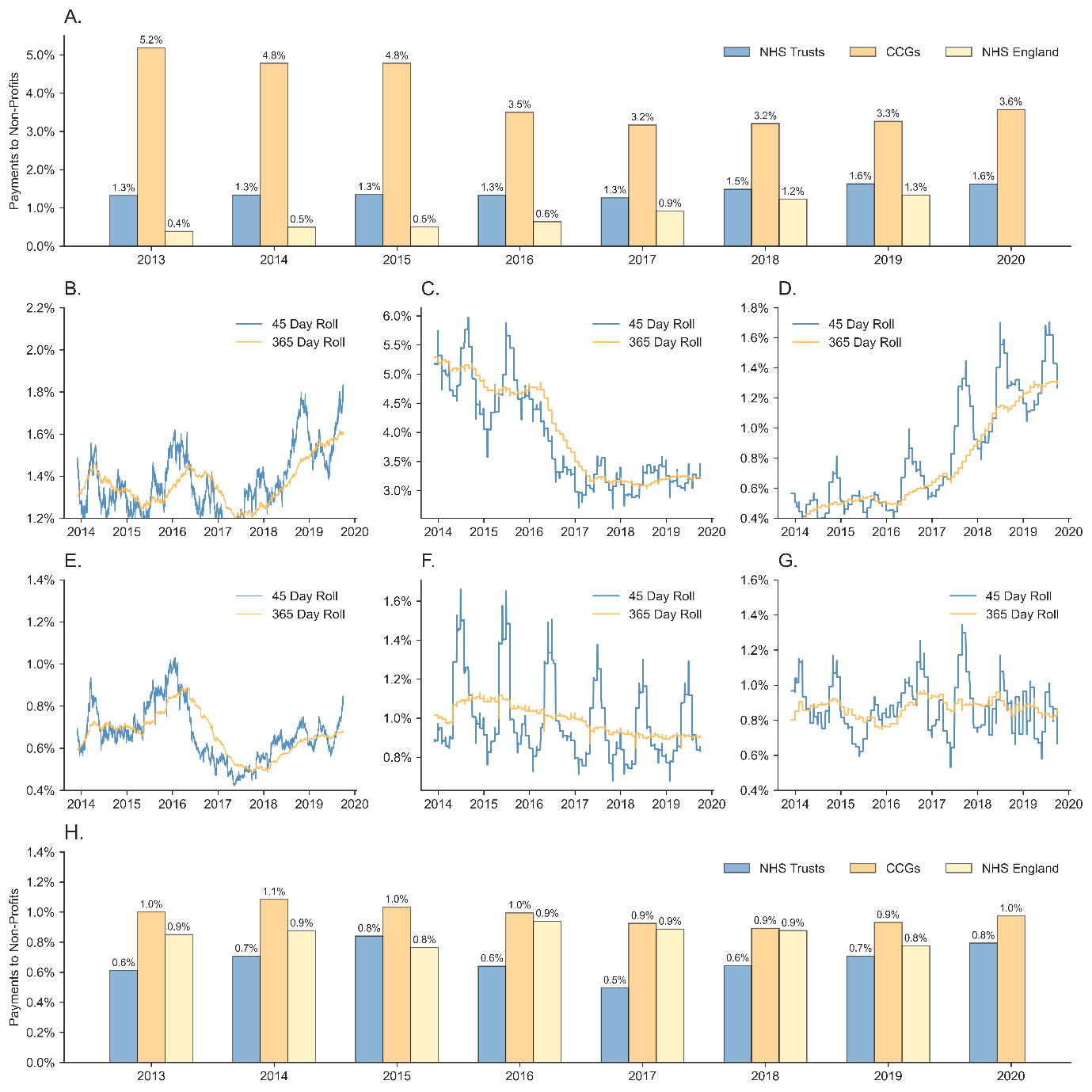
Description automatically generated with low confidence

**Figure 8: Geospatial Distribution of CCG Payments.** Choropleth map of all parsed and reconciled CCG payments as of 31/03/2020. Shapefile provided by NHS Digital. The `No Data' refers to the small number of CCGs which do not make their data available (in a parse-able format). Subfigure A. refers to the number of payments made to institutions registered on the Charity Commission as a percent of all payments made, and B. refers to the cumulative value of such payments.



**Figure 9: Geospatial Distribution of Trust Payments.** Choropleth map of all parsed and reconciled NHS Trust payments as of 31/03/2020. As Trusts do not map into designated polygons, the geocoded location of their headquarters as registered on NHS Digital is provided. These points are layered on top of the CCG Shapefile, provided by NHS Digital. Subfigure A. refers to the number of payments made to institutions registered on the Charity Commission as a percent of all payments made, and B. refers to the cumulative value of such payments.

**Figure 10: Temporal Variation in Procurement from Non-Profits.** Panels A. and H. show the aggregation over years for each of the Trust, CCG and NHS England datasets, showing the cumulative percent of procurement from non-profits for that organisation type. Panels B.-G. plot this on a daily basis using backward-looking 45- and 365-day rolling windows for Trusts (left column), CCGs (middle column) and NHS England (right column). The upper middle row represents number of payments, and the lower represents payment value**.**



**Figure 11: Concentration Ratios Within Non-Profits in 180 Day Rolling Windows.** Each line represents the proportion of payments within all of those which go to institutions registered on the Charity Commission for England and Wales which go to either the five, ten, or twenty highest value recipients within a 180 day rolling window.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Clinical Commissioning Groups | | | | |  | NHS Trusts | | | |
|  | PDF | Excel | CSV | | Total |  | PDF | Excel | CSV | Total |
| Files | 687 | 3756 | 3307 | | 7750 |  | 270 | 3407 | 4912 | 8589 |
| Rows | 50591 | 327255 | 269323 | | 646938 |  | 21461 | 295667 | 344209 | 661226 |
| Value (£m) | 16364 | 100256 | 875796 | | 204111 |  | 32873 | 38493 | 45159 | 86932 |
| Mean (£k) | 323 | 306 | 325 | | 315 |  | 153 | 130 | 131 | 131 |
| Suppliers | 2165 | 4933 | 3982 | | 7973 |  | 1380 | 6883 | 7631 | 11203 |
|  | | | |  | | | | | | |
|  | NHS England | | | | |  | Combined Database | | | |
| Files | 0 | 67 | 1 | | 68 |  | 955 | 7213 | 8148 | 16316 |
| Rows | 0 | 565206 | 36555 | | 601761 |  | 72052 | 1188128 | 650087 | 1909925 |
| Value (£m) | 0 | 147991 | 6830 | | 154820 |  | 19651 | 286740 | 139567 | 445863 |
| Mean (£k) | 0 | 269 | 187 | | 257 |  | 273 | 241 | 215 | 233 |
| Suppliers | 0 | 5732 | 3718 | | 5734 |  | 3300 | 13885 | 12681 | 19122 |
| **Table 1: Distribution of data across file-types and institution.** Files relates to the number of files in each category from which we can extract data which passes the cleaning process described in 2.1. Rows refers to the total rows of data (total number of payments). Value (£m) refers to the total value of these payments. Mean (£k) refers to the average payment value. Suppliers refers to the number of unique suppliers. | | | | | | | | | | |

**Table 2: Highest Value VCS Suppliers.** Top 10 institutions by procurement value mapped to the Charity Commission, ordered by the cumulative value of all contracts they receive (£m). Count refers to the number of payments made in our datasets. ‘Income’ (£m) and ‘Rank’ refer to their total income for the last financial year at the time of writing (2018/2019), as supplied by the Charity Commission as of 31/03/2020.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | £m | Count | Reg No. | ICNPO | Income (£m) | CC Rank |
| St Andrew’s Healthcare | 730.93 | 834 | 1104951 | 3300 | 1370 | 26 |
| Nuffield Health | 326.50 | 3021 | 205533 | 3100 | 4840 | 3 |
| Healthcare QIP | 110.95 | 224 | 1127049 | 3400 | 150 | 427 |
| Hospice UK | 78.28 | 17 | 1014851 | 3200 | 105 | 652 |
| Healthcare Management Trust | 613.26 | 348 | 292880 | 4100 | 177 | 350 |
| Royal Hospital for Neuro Disability | 58.24 | 140 | 205907 | 3100 | 2.90 | 256 |
| Marie Curie | 55.69 | 600 | 207994 | 3400 | 1090 | 33 |
| The Forward Trust | 51.50 | 461 | 1001701 | 7200 | 137 | 480 |
| Horder Healthcare | 50.12 | 448 | 1046624 | 3400 | 192 | 304 |
| MSI Reproductive Choices | 43.33 | 1117 | 265543 | 3400 | 1600 | 16 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Financial Year: 2011/2012 | Minor | Small | Medium | Large | Major | Super-Major |
| Entire CCEW (%) | 42.39 | 36.70 | 16.31 | 3.86 | 0.69 | 0.05 |
| Organisation Count (%) | 0.64 | 3.78 | 37.65 | 41.59 | 14.48 | 1.85 |
| Percent of Payments (%) | 0.71 | 0.51 | 16.78 | 29.85 | 34.33 | 17.82 |
| Percent of Amount (%) | 0.54 | 0.20 | 21.70 | 26.67 | 24.05 | 26.84 |
| Financial Year: 2018/2019 |  |  |  |  |  |  |
| Entire CCEW (%) | 37.47 | 37.56 | 19.84 | 4.32 | 0.75 | 0.06 |
| Organisation Count (%) | 1.18 | 2.45 | 33.46 | 44.59 | 16.42 | 1.89 |
| Percent of Payments (%) | 0.54 | 4.66 | 10.45 | 26.54 | 35.93 | 21.87 |
| Percent of Amount (%) | 0.21 | 8.57 | 12.10 | 22.73 | 27.14 | 29.26 |

**Table 3: Charity Classification by Income.** The upper half of the table details splits across charity type determined by their income as of 2011/2012, and the latter half for 2018/2019. 'Entire CCEW' refers to the whole of the Charity Commission of England and Wales, 'Organisation Count' refers to the number of organisations which meet the thresholds without merging with NHS procurement data. The latter two rows show the distribution of payments by volume and amount.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | CCGs | | NHS Trusts | | NHS England | |
| ICNPO | | Number (%) | Amount (%) | Number (%) | Amount (%) | Number (%) | Amount (%) |
| 1100 | Culture and sport | 1.25 | 0.05 | 1.54 | 0.21 | 1.59 | 0.01 |
| 1200 | Culture and sport | 0.96 | 0.08 | 0.58 | 0.23 | 0.79 | 0.11 |
| 2100 | Education | 0.58 | 0.02 | 1.16 | 0.67 | 1.19 | 0.01 |
| 2120 | Education | 0.1 | 0.02 | 0.19 | 0.02 | 0 | 0 |
| 2200 | Education | 0.29 | 0.11 | 0.58 | 0.21 | 0.4 | 0 |
| 2300 | Education | 0.96 | 0.05 | 2.12 | 1 | 1.19 | 0.03 |
| 2400 | Education | 0.77 | 0.04 | 2.12 | 0.86 | 1.19 | 0.05 |
| 2410 | Education | 0.87 | 0.07 | 2.9 | 5.55 | 2.78 | 1.04 |
| 3100 | Health | 4.33 | 17.78 | 9.07 | 13.44 | 8.33 | 7.23 |
| 3200 | Health | 14.13 | 36.35 | 8.49 | 3.28 | 12.3 | 9.9 |
| 3300 | Health | 9.9 | 6.37 | 10.23 | 9.7 | 7.54 | 57.73 |
| 3400 | Health | 2.88 | 11.19 | 5.41 | 19.97 | 3.97 | 8.51 |
| 4100 | Social Services | 32.88 | 16.43 | 22.78 | 26.62 | 22.22 | 5.41 |
| 4110 | Social Services | 0.19 | 0 | 0 | 0 | 0 | 0 |
| 4150 | Social Services | 0.29 | 0.04 | 0.77 | 0.32 | 0.4 | 0.02 |
| 4160 | Social Services | 0.67 | 0.03 | 0 | 0 | 0 | 0 |
| 4170 | Social Services | 0.29 | 0.21 | 0 | 0 | 0.79 | 0.01 |
| 4180 | Social Services | 0.1 | 0.01 | 0.19 | 0.01 | 0 | 0 |
| 4200 | Social Services | 0.38 | 0.02 | 0.58 | 0.03 | 0 | 0 |
| 4300 | Social Services | 0.58 | 0.15 | 0.19 | 0.02 | 0.79 | 0.06 |
| 5100 | Environment and animals | 0.96 | 0.16 | 0.97 | 0.04 | 0.4 | 0 |
| 5200 | Environment and animals | 0.1 | 0 | 0.39 | 0.03 | 0 | 0 |
| 6100 | Community, employment, housing | 3.46 | 0.44 | 2.51 | 2.74 | 2.78 | 0.93 |
| 6200 | Community, employment, housing | 5.96 | 4.53 | 5.21 | 8.51 | 5.56 | 1.87 |
| 6300 | Community, employment, housing | 1.15 | 0.07 | 2.12 | 0.53 | 1.59 | 0.24 |
| 7100 | Law, advocacy and politics | 2.4 | 0.65 | 2.9 | 1.28 | 4.37 | 1.64 |
| 7200 | Law, advocacy and politics | 2.5 | 0.66 | 3.09 | 1.65 | 1.98 | 3.64 |
| 8100 | Philanthropy and volunteering | 1.54 | 0.33 | 1.93 | 0.17 | 2.38 | 0.3 |
| 8200 | Philanthropy and volunteering | 5.67 | 1.75 | 4.63 | 0.68 | 6.75 | 0.23 |
| 10100 | Religion | 1.44 | 0.93 | 1.35 | 0.26 | 1.19 | 0.07 |
| 11100 | Business and professional | 0.1 | 0 | 1.35 | 0.43 | 0.79 | 0.01 |
| 11200 | Business and professional | 0.87 | 0.2 | 2.51 | 0.68 | 5.16 | 0.92 |
| 1100 | Culture and sport | 1.25 | 0.05 | 1.54 | 0.21 | 1.59 | 0.01 |
| 1200 | Culture and sport | 0.96 | 0.08 | 0.58 | 0.23 | 0.79 | 0.11 |
| 2100 | Education | 0.58 | 0.02 | 1.16 | 0.67 | 1.19 | 0.01 |
| **Table 4: Distribution of payments Across Charity Classes**. Classes determined and provided by the Charity Commission. Note: Multiple classes per organisation possible (in contrast to ICNPO numbers). Amount (%) refers to the cumulative value of all payments made to all institutions of a specific class, while count is the number of payments going to institutions reporting that same specific class. | | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | CCGs | | NHS Trusts | | NHS England | |
| Class | Amount  (%) | Count  (%) | Amount  (%) | Count  (%) | Amount  (%) | Count  (%) |
| Accommodation/housing | 1.75 | 2.26 | 2.15 | 2.12 | 0.8 | 1.74 |
| Acts As An Umbrella Or Resource Body | 0.98 | 2.13 | 1.52 | 2.12 | 1.93 | 2.9 |
| Amateur Sport | 0.19 | 0.68 | 0.19 | 0.6 | 0.01 | 0.22 |
| Animals | 0 | 0.03 | 0.04 | 0.02 |  |  |
| Armed Forces/emergency Service | 0.01 | 0.05 | 0.01 | 0.06 | 0.01 | 0.13 |
| Arts/culture/heritage/science | 0.21 | 0.8 | 0.38 | 0.93 | 0.03 | 0.63 |
| Children/young People | 6.05 | 5.51 | 6.4 | 5.4 | 9.54 | 6.34 |
| Disability | 4.49 | 5.47 | 3.29 | 3.97 | 1.85 | 4.38 |
| Economic/community dev/employment | 1.29 | 2.69 | 2.32 | 2.68 | 0.41 | 2.46 |
| Education/training | 4.64 | 5.89 | 6.59 | 7.06 | 10.43 | 7.01 |
| Elderly/old People | 8.2 | 5.81 | 4.76 | 4.6 | 8.22 | 4.06 |
| Environment/conservation/heritage | 0.4 | 0.42 | 1.47 | 0.54 | 0.18 | 0.63 |
| General Charitable Purposes | 1.99 | 3.75 | 2.84 | 3.24 | 1.63 | 3.17 |
| Rights/religious/racial equality/Diversity | 0.46 | 0.53 | 0.22 | 0.58 | 0.12 | 0.49 |
| Makes Grants To Individuals | 0.25 | 0.8 | 1.2 | 1.23 | 0.19 | 1.25 |
| Makes Grants To Organisations | 1.08 | 1.02 | 1.95 | 1.84 | 1.01 | 2.01 |
| Other Charitable Activities | 0.61 | 1.55 | 0.97 | 1.62 | 1.39 | 1.65 |
| Other Charitable Purposes | 1.39 | 1.28 | 1.21 | 1.6 | 0.71 | 1.52 |
| Other Charities Or Voluntary Bodies | 1.24 | 2.25 | 2.71 | 2.63 | 1.02 | 2.77 |
| Other Defined Groups | 3.46 | 3.61 | 4.16 | 3.95 | 3.28 | 4.33 |
| Overseas Aid/famine Relief | 0.1 | 0.1 | 0.43 | 0.09 | 0.01 | 0.22 |
| Particular Ethnic Or Racial Origin | 0.6 | 1.4 | 1.04 | 1.47 | 0.55 | 1.21 |
| People With Disabilities | 8.25 | 7.14 | 6.06 | 5.83 | 9.53 | 6.39 |
| Provides Advocacy/advice/information | 6.63 | 7.98 | 6.93 | 7.45 | 4.44 | 7.99 |
| Provides Buildings/facilities/open Space | 4.57 | 4.5 | 3.57 | 4.23 | 2.12 | 3.44 |
| Provides Human Resources | 3.63 | 4.16 | 3.56 | 2.98 | 2.15 | 2.99 |
| Provides Other Finance | 0.33 | 0.21 | 0.68 | 0.45 | 1.06 | 0.58 |
| Provides Services | 12.7 | 10.09 | 9.6 | 9.37 | 11.97 | 8.62 |
| Recreation | 0.12 | 0.59 | 0.06 | 0.32 | 0.03 | 0.31 |
| Religious Activities | 0.16 | 0.3 | 0.87 | 0.26 | 0.01 | 0.18 |
| Sponsors Or Undertakes Research | 3.74 | 2.54 | 4.4 | 3.69 | 10.63 | 4.91 |
| Advancement Of Health Or Saving Of Lives | 11.18 | 6.97 | 9.14 | 7.8 | 11.13 | 7.77 |
| General Public/mankind | 8 | 4.75 | 7.22 | 6.48 | 3.29 | 5.49 |
| Prevention Or Relief Of Poverty | 1.29 | 2.73 | 2.04 | 2.79 | 0.33 | 2.19 |
| **Table 5: Distribution of payments Across ICNPO Codes**. Based on the International Classification of Non-Profit Organisations. Note: Unlike with Charity Commission classes, ICNPO categories are unique across organisations. Amount (%) refers to the cumulative value of all payments made to all institutions of a specificc ICNPO, while count is the number of payments going to institutions reporting that same specificc ICNPO. | | | | | | |

1. For correspondence: Charles Rahal, Leverhulme Centre for Demographic Science, Department of Sociology, University of Oxford, OX1 1JD, United Kingdom. Tel: 01865 286170. Email: charles.rahal@sociology,ox.ac.uk. The extensive code library and copies of the curated data resources which accompany this work can be found at https://www.github.com/crahal/NHSSpend and https://zenodo/record/abcxyz123456. The authors are grateful to comments on earlier versions of the work from Mark Exworthy, David Stuckler, Martin Mckee, Lucy Reynolds, James Rees and Aaron Reeves. Technical research assistance provided by Ian Knowles, and Figure 1 implemented with the help of James May. Data audits were provided by Steve Barnard, Ben Goodair, Max Hattersly and Yu Pei. Funding gratefully acknowledged from the British Academy, with ESRC ES/M010392/1 laying the foundation as a ‘Proof of Concept’. Insightful comments acknowledged from participants at the International Conference for Administrative Data Research, 2018. [↑](#footnote-ref-1)
2. We predominantly use the term ‘non-profit’ rather than voluntary, nongovernmental, third sector, or otherwise, following United Nations terminology in their System of National Accounts. More technically, the institutions which we examine might be termed ‘non-profits listed on the Charity Commission for England and Wales’, as not all non-profits are: some are exempt by characterisation (such as charitable service funds of armed forces with an income under £100,000), or those subject to other regulators (such as museums, galleries and other institutions of national importance, which are regulated by the Department for Digital, Culture, Media and Sport). [↑](#footnote-ref-2)
3. This is in addition to the mandated provision of data at other administrative levels (and at various thresholds), such as local authorities (Department for Communities and Local Government, 2015) and smaller local councils (Department for Communities and Local Government, 2014). [↑](#footnote-ref-3)
4. This work directly advances the pipeline discussed in that process in terms of database size, parsing fecundity (in the form of .pdfs), and substantial innovations in terms of reconciliation algorithms. [↑](#footnote-ref-4)
5. A cursory example of this would be ‘Nuff Health’ and ‘Nuffeld Health’ both relating to the register entry: ‘Nuffield Health’. [↑](#footnote-ref-5)
6. Here we focus on curating data on; Special Health Authorities in England, Commissioning Support Units and Data Management and Integration Centres, Executive Agencies of the Department of Health, Local Service Providers (LSP), LSP Sites, Cancer Networks, Strategic Health Authority Sites (closed), Special Health Authority sites, Other Statutory Authorities, OSA Sites, Executive Agencies of the Department of Health, Executive Agency Programmes, Executive Agency Programme Departments, Executive Agency Sites, Government Departments, Government Department Sites, Public Health Observatories, Cancer Registries, Channel Island Health Organisations, Military Hospitals, Clinical Networks, Application Service Providers, National Application Service Providers, NHS England Area Team Sites, NHS Support Agencies in England, GP Practices, NHS Trusts, Care Trusts, Welsh Local Health Boards, Prisons, Schools and Local Authorities [↑](#footnote-ref-6)
7. However, the charitable status of large, highly commercial, fee charging charities such as Nuffield Health is frequently contested, despite protestations by the charities themselves that they do not operate to make profit for shareholders nor investors, reinvesting for their cause and public benefit. [↑](#footnote-ref-7)
8. These categories are as follows: 1100 - Culture and Arts, 1200 - Sports, 1300 - Other Recreation and Social Clubs, 2100 - Primary and Secondary Education, 2110 - Parent Teacher Associations, 2120 - Educational Foundations, 2130 - Playgroups and nurseries, 2200 - Higher Education, 2210 - Student Unions, 2300 - Other Education, 2400 - Research, 2410 - Medical Research, 3100 - Hospitals and Rehabilitation, 3200 - Nursing Homes, 3210 - Hospices, 3300 - Mental Health and Crisis Intervention, 3400 - Other Health Services, 4100 - Social Services, 4110 - Scouts, guides and other groups, 4150 - Social services for children, young people and families, 4160 - Social services for older people, 4170 - Social services for adults with learning disabilities, 4180 - Social services for people with disabilities, 4200 - Emergency and Relief, 4300 - Income Support and Maintenance, 5100 - Environment, 5200 - Animal Protection, 6100 - Economic, Social and Community Development, 6110 - Village Hall, 6200 - Housing, 6300 - Employment and Training, 7100 - Civic and Advocacy Organizations, 7200 - Law and Legal Services, 8100 - Grant-making foundations, 8200 - Other philanthropic intermediaries and voluntarism promotion, 9100 - International activities, 10100 - Religious congregations and associations, 11100 - Business associations, 11200 - Professional associations, 12100 - Not Elsewhere classified. [↑](#footnote-ref-8)
9. The identification of hospital catchment areas is an significant challenge in health services research, as discussed in Garnick et al. 1987. Hospitals operated by Trusts receive patients from varied geographical dispersions: some of which represent a large proportion of the hospital's total activity, and some of which are physically far removed from the hospital. [↑](#footnote-ref-9)
10. With multiple works such as Wells (1982), Jack (1987), and Dorling (2007) reach a vague consensus that the dividing boundary line might run somewhere roughly from Bristol to somewhere between just south of Hull and north of King’s Lynn, we follow a rough approximation (as per Buchan *et al*., 2017) from the recent social epidemiology literature which groups the ‘East’, ‘London’, ‘South East’ and ‘South West’ NUTS regions as the ‘South’. [↑](#footnote-ref-10)
11. This variation occurs at a level as granular as to the use of tests in UK primary care (O'Sullivan et al, 2018). [↑](#footnote-ref-11)
12. Mohan (2012a, 2012b) talks about the likely continuation of this 'grey area'. [↑](#footnote-ref-12)