Journal of the Royal Statistical Society Series A: Statistics in Society, 2025, **188**, 1085–1106
https://doi.org/10.1093/jrsssa/qnae092
Advance access publication 27 September 2024
Original Article

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series A
Statistics in Society

The role of the third sector in public health service provision: evidence from 25,338 heterogeneous procurement datasets

Charles Rahal[1,2] and John Mohan[3]

[1]Leverhulme Centre for Demographic Science, Demographic Science Unit, University of Oxford, Oxford OX1 1JD, UK
[2]Nuffield College, University of Oxford, Oxford, UK
[3]Third Sector Research Centre, University of Birmingham, Birmingham B15 2TT, UK

Address for correspondence: Charles Rahal, Leverhulme Centre for Demographic Science, Demographic Science Unit, University of Oxford, Oxford OX1 1JD, UK. Email: charles.rahal@demography.ox.ac.uk

Abstract

We examine the role of nonprofits within publicly funded healthcare which runs parallel to private provision in a 'two-tier' system through a unique Big Data pipeline. Scraping tens of thousands of heterogeneous transaction datasets across a commissioning hierarchy, the processed dataset contains over £445Bn worth of transactions in 1.9m+ rows of clean data, spanning 2012–2020. We utilize this dataset to test a range of hypotheses related to the introduction of the Health and Social Care Act of 2012. The proportion of contracts placed with nonprofit organizations is relatively low, with limited evidence as to whether the Act increased the involvement of third sector organizations in line with NHS 'marketization'. We analyse the pattern of procurement by corporate category, the field of service delivery (International Classification of Nonprofit Organizations and Standard Industrial Classification codes), and 'expense area' to show the unique array of services which voluntary organizations supply. We also analyse the pattern of commissioning across entity class and size distributions of registered charities. We conclude with a consideration of high-value Community Interest Companies, and discuss potential further areas of research within a healthcare context which such government transaction data makes possible.

Keywords: big data, data visualization, healthcare supply, National Health Service

JEL codes: H10, H41, H51, I11, L10

1 Introduction

The UK's National Health Service is often seen as an exception by international standards because of the extent of public financing, control and delivery of health services. It is true that elements of private provision of core NHS functions have been in place since the service was created in 1948 (Salter, 1995). However, governments since at least the 1980s have sought to expand the scope for provision of NHS services by commercial and voluntary sector entities. While critics charge that such initiatives amount to privatization and the introduction of market forces into a nonmarket sphere, governments have consistently argued that these policies were intended to introduce a plurality of providers into healthcare, especially by offering opportunities for provision by non-profit entities such as charities or social enterprises. However, establishing the extent to which such opportunities have been taken—for instance, by quantifying the scale of contractual arrangements with such entities, and understanding the characteristics of those organizations in receipt of them—have been problematic. Our focus is on establishing a firmer evidence base on the scale, distribution and growth of NHS provision by nonprofit and voluntary organizations. We focus

Received: March 14, 2022. **Revised:** August 6, 2024. **Accepted:** August 18, 2024

specifically on the extent of changes over the decade since 2012, when the Coalition government (2010–2015) introduced the Health and Social Care Act of 2012 (hereafter, 'HSCA 2012'), which marked an especially noteworthy change in the commissioning of NHS England services. We are aware of no similar attempt at the analyses we undertook, and break further new ground by exploiting novel sources of data in a sophisticated way. Our primary sources are the very considerable volumes of data on public sector procurement, made available because of the same Coalition government's transparency agenda, which (from May 2010) gave significant impetus to efforts to make data on government transactions available. This data made it possible to track payments from central government through to individual providers of NHS services. Public sector organizations must publish information on transactions of over £25,000, with rows of data including information required on the supplier, the date of the transaction, the transaction value, and many other auxiliary fields.[1] Our dataset captures this between the period of 2012–2020, allowing a substantial analysis of the post-HSCA 2012 landscape.

The resulting datasets are large-scale and were usually updated monthly, requiring complex computational approaches to capture, manage, and utilize. This allows a research design that matches data on NHS procurement to institutional registers (that is, lists of potential providers of services) to draw conclusions on our substantive question of interest: what role do voluntary sector providers play in the post-HSCA 2012 public healthcare regime (and have opportunities generally increased)? What can be said about the characteristics of those voluntary sector entities which have attracted contracts? To what extent are different organizational components of the voluntary sector (e.g. charities or social enterprises) successful in attracting contracts? Our approach offers significant and substantive insights into how NHS procurement data can best be harnessed and analysed, and insights into the dynamics of relationships between the state, the nonprofit sector, and the market in the commissioning of public services.

We begin by discussing the rationale for the involvement of nonprofit and commercial entities in the provision of publicly funded healthcare, and the background to successive reforms of the NHS which have sought to expand the scope for the provision by private entities (Section 2.1). We also summarize the existing evidence base on how much is spent on these entities (Section 2.2), concluding with a concrete statement of the research questions which we have sought to answer (Section 2.3). We then describe the sources of data upon which we draw, the immense challenges encountered in using them, and our computational approach to solving such challenges (Section 3.1 and online supplementary material, Section S1–S2). This is followed by a discussion of the institutional registers which we link to (Section 3.2). We then reconcile this data in Section 3.3 (with technical methodological developments described in online supplementary material, Section S3) to create our final, linked dataset. This is followed by our findings regarding the involvement of nonprofit and commercial providers (Section 4). We conclude with a generalized discussion of our findings of and their implications for future research and policy (Section 5). The accompanying online code repository—indexed via Zenodo as Rahal et al. (2023)—contains all datasets, computer code, and analytical outputs discussed throughout.

2 Nonstate providers of healthcare in the British NHS: rationale, extent of involvement, and difficulties in quantification

2.1 Arguments for private provision of healthcare in a public health service

The provision of NHS services has never been an absolute state monopoly. Health authorities have always been able to agree contracts for treatment with nonstate providers of services, but the extent and character of provision of NHS healthcare by commercial and voluntary agencies has received relatively little academic attention, and a key aim of this paper is to shed light on this issue. In general, voluntary organizations share several characteristics. They are private, or independent of government; their activities entail elements of voluntarism in either or both of governance and service delivery; they are self-governing, which means they can close themselves down without reference to external authority; and they are nonprofit distributing, where any surpluses are retained for investment in the mission of the organization (Salamon & Anheier, 1992). One rationale for

[1] 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).

involving such entities in public service provision is that in complex fields of activity such as health and social care—where there are problems of information asymmetries between providers and recipients—the nonprofit form offers reassurance to service users that professional judgements are not compromised by commercial motivations.

A second rationale concerns failures of both markets and states to respond to diverse and specific categories of need, either because of insufficient effective market demand or because of the absence of democratic endorsement, both of which result in under provision. A third rationale is that the state and the voluntary sector can form effective partnerships, recognizing their respective strengths (Billis & Glennerster, 1998; Hansmann, 1979; Weisbrod, 2009). Commercial, for-profit providers are likely to focus on services where pricing is more straightforward, such as routine surgical interventions (joint replacements, for example). Large-scale provision enables—for commercial entities—a degree of certainty in the potential to generate excess profit. The ability of commercial providers to raise capital through financial markets may also secure desirable public objectives, such as the rapid development of new facilities in contrast to the delays associated with the public funding of capital investment.

2.2 Policy towards contractual arrangements for the provision of NHS care

Contractual arrangements between the NHS and nonprofit providers were not unknown in the first four decades of the NHS, but largely resulted from ad hoc local arrangements: pragmatic responses to local issues in NHS capacity. The Conservative governments of the 1980s and early 1990s encouraged health authorities to place contracts with non-NHS service providers, partly as an effect of important shifts towards community-based models of service provision in fields such as mental healthcare, and care for people with disabilities. However, the amounts of money and the numbers of organizations involved were small (Higgins, 1988). The 1991 reforms of the NHS established what was known as the 'purchaser-provider split': NHS budgets were allocated to local 'purchasers' of services, known variously as Primary Care Trusts (PCTs, from 2001 to 2013) and subsequently as Clinical Commissioning Groups (CCGs) for the period covered by this paper (some budgets were also devolved direct to general practitioners). These local entities receive allocations of funds which vary according to assessments of the relative needs of their resident populations. They deploy these funds to procure services to meet those needs from whichever organizations they deem appropriate, such as large NHS Trusts (which provide hospitals and community services), GPs, and external organizations such as charities and private companies. Despite the promarket rhetoric accompanying the 1991 reforms, the scope for purchasers of services to switch contracts between providers was constrained; governments were cautious, on the grounds that the service needed stability rather than the destabilization of markets. However, the Labour governments subsequently moved in various directions to enhance the scope for commercial provision of services. Most notably, Labour entered into a 'concordat' with the private sector, granting commercial providers of clinical services confidence in the availability of long-term, large-scale contracts with the NHS (we do not discuss these here as their establishment and development has been considered by previous authors; see Leys, 2001; Leys & Player, 2011). Other policies favoured contestability in service provision between types of providers. Thus, greater encouragement was given to social enterprises, with a new nonprofit structure—the Community Interest Company (CIC)—becoming available from 2004 and providing greater flexibility than traditional charitable models of nonprofit provision. NHS employees were supported to manage 'spinouts' of elements of NHS provision into freestanding enterprises or charities, often characterized as employee-run mutuals, on the basis that greater employee control would improve performance and outcomes (Miller & Lyon, 2016) and facilitate greater entrepreneurial freedoms.

These policies signalled a direction of travel away from state dominance in service provision, to one in which provider pluralism was the way forward. It is against this backdrop that proposals were introduced for the reform of the National Health Service through the HSCA 2012. The intention was to promote a strongly competitive dynamic for reshaping health services by eliminating various geographically based structures for oversight of the delivery of healthcare, and the decentralization of budgets to local service commissioning consortia (CCGs). Diversification of the provider market for healthcare was an explicit goal: services could henceforth be commissioned

from 'Any Qualified Provider'. Providers would compete on a 'level playing field' for NHS contracts, reducing barriers to entry while supposedly offering considerable opportunities to the voluntary and community sector. The post-2010 government had presented its policies as examples of welfare pluralism and as providing opportunities for community-based provision of services led by social enterprizes, but critics charged that the intention was to expand the scope for commercial service provision to a much greater extent than in any previous reforms (Delamothe & Godlee, 2011; Leys & Player, 2011).

2.3 To what extent have these reforms led to expansion in the role of commercial and nonprofit provision of NHS care?

How much of the NHS's budget is spent on voluntary and nonprofit organizations? Prior to the implementation of the HSCA 2012, the voluntary sector provided a diverse range of services. Healthcare supply by the voluntary sector ranges from specialist clinical provision complementary to the NHS where it often dominates the market—such as the hospice movement (Field & Johnson, 1992)—to disease-specific advice services (such as Diabetes UK) and general wellness support and advocacy. National government accounts (such as from, at the time of our exercise, the Department of Health) quote a figure of 6.1%–7.3% as going to 'Expenditure on non-NHS bodies', but that estimate is much contested for various reasons (Rowland, 2019), and does not provide any differentiation between charitable or profit-driven entities. The most prominent quantitative work on the funding of the UK's voluntary sector by the National Council for Voluntary Organisations (NCVO) has focussed on estimating the total resources of voluntary organizations and major components of their income, such as donations, legacies, fees for services, and government grants and contracts. They capture data from the accounts of registered charities, with an emphasis on estimating economic aggregates; sampling fractions were heavily weighted towards larger organizations. That data, combined with information from large representative surveys of voluntary organizations, can be used to generate a baseline estimate of the scale of voluntary sector involvement in delivering NHS care prior to the 2012 reforms.[2] This resource seeks to capture detailed data on the distinctive funding sources received by English and Welsh charities. These are described in the notes to charities' accounts, and they provide greater detail than the very broad categories which must be reported by registered charities to the Charity Commission for England and Wales (CCEW, the main registry spine used herein). In a typical year—with between 9,000 and 10,000 accounts being captured—information from the notes to the accounts generates around 50 lines of text per charity, or up to half a million items of text data each year. This is then classified by NCVO so that key funding sources (central or local government; NHS; European funds) can be identified. Typically, NCVO use the data to estimate aggregate flows of public funding to the voluntary sector; calculations using their data show that around £1.15Bn was received by charities from the NHS in the years immediately prior to 2010. Of those charities (n = 9,497) with incomes greater than £500K, 1,071 (11.3%) received NHS funding. While the threshold of £500K excludes the majority of charities (the median income of which is c. £20,000), it includes organizations most likely to be of a sufficient scale to deliver public service contracts. Organizations of this size must include greater detail in their accounts as well as prepare them in a standardized fashion.

Two authoritative surveys of the voluntary sector in England can also be used to estimate how many organizations received funding from the NHS. These were the National Survey of Third Sector Organisations (NSTSO) and its successor, renamed the National Survey of Charities and Social Enterprises (NSCSE). With over 40,000 responses each, these were the largest surveys ever undertaken of voluntary organizations in one country. The surveys did not ask for estimates of the amount of funding received from particular sources but organizations were presented with a lengthy menu of types of public sector organizations, and asked to indicate those from which they received funding. The NSTSO was designed to provide a representative sample of third sector organizations, including not only charities, but also CICs, Companies Limited by Guarantee (CLGs), and Industrial and Provident Societies (IPSs). In the 2008 survey, estimates suggest that 2,233

[2] The sampled accounts data have been deposited at the UK Data Service as part of a wider suite of resources generated by the Third Sector Research centre: see ESRC Grant Reference RES-595-28-0001.

charities with incomes greater than £500K received at least some funding from the NHS. Narrowing down further and considering only charitable companies (on the basis that organizations should not be entering into contracts if they do not have the protections of company status), around 29% of such entities with incomes of over £500K reported receipt of NHS funding. These figures suggest a baseline estimate that around 2,000 charities were in receipt of receiving contracts for NHS provision in the years immediately prior to 2010. To these, we could add a small number of CLGs and CICs. The latter had only been introduced in 2004—so we would not expect numbers to be high only 4–6 years later—but these are of substantive interest considering the claims made by the Conservative government that their 2012 reforms would have significantly expanded the scope for social enterprises within the NHS. In addition to deep descriptive characterizations that we have built (Section 4.1), we were formally interested in three specific research questions related to nonprofit provision within the NHS that we believe that transactional data is best positioned to illuminate:

1. What is it that charitable organizations were supplying to NHS organizations, and within what fields of service provision did/do they operate (Section 4.2)?
2. Have opportunities for the voluntary sector expanded over the period covered by this research (Section 4.3)?
3. What types of charitable and social entities are successful in entertaining these opportunities (Section 4.4)?

3 Data

Public sector agencies are required to publish monthly information by working day 15 after each month end. It 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 are provided under the Open Government Licence v3.0 (The National Archives, 2001). Given that there are several hundred discrete entities publishing data monthly, and many hundreds of thousands of organizations that might, in principle, be the recipients of government contracts, the matching of procurement data to organizational databases is a formidable task. 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 back to Newcombe et al. (1959). The Company, ORganization and Firm name Unifier (CORFU) approach of J. M. Alvarez-Rodriguez et al. (2015) reconciles the suppliers of public procurement with supplier names in Australia between 2004 and 2012 (containing 77,526 unique names across 430,188 payments). Svátek et al. (2014) 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 J. Alvarez-Rodríguez et al., 2012)—addressed the matching task using techniques such as spreading activation and resource description framework (RDF) classification. The most similar works in terms of method and application are Rahal (2018, 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 which analyses central government procurement data and provides three examples 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. This work directly advances and substantially expands that pipeline in terms of database size, parsing fecundity (specifically with regards to files in PDF format, but also otherwise), substantial innovations in terms of reconciliation algorithms, and in research design more broadly. In terms of pure application, the Institute for Government (Davis & Hornung, 2018) uses data from the Spend Network to provide a broad, macro-level overview as to what the UK government procures. We further elucidate upon the dimensions of our data and methods in the following subsections.

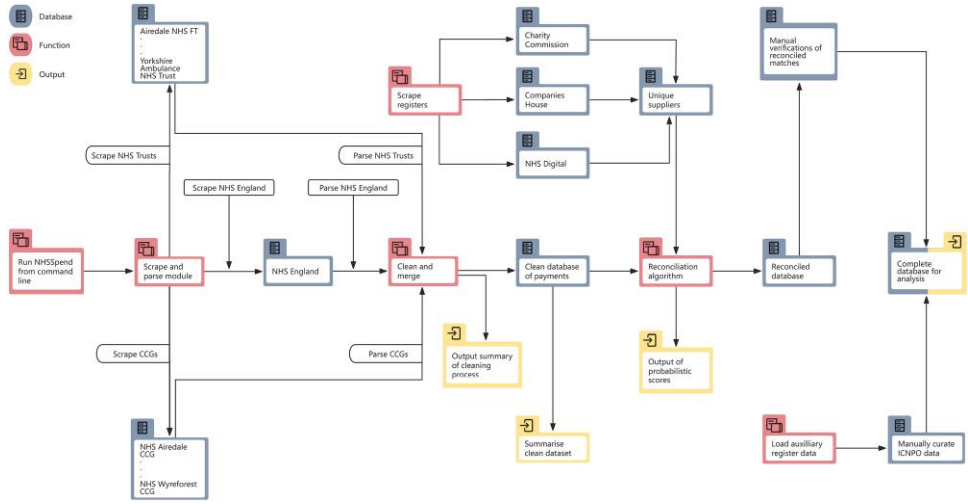

Figure 1. Schematic detailing the operation of NHSSpend. Red nodes identify functions and processes, blue nodes identify the creation of a database, and yellow nodes denotes the creation of an output of analysis on a database.

3.1 Raw procurement data

In addition to NHS England, every CCG and NHS Trust (of which we consider 191 and 225, respectively, although there is substantial longitudinal fluidity in completeness) publishes their expenditure data in a different place, using a different format, and at different points in time. Our complex pipeline for curating a final dataset for analysis is visualized in Figure 1 (and is described in detail in online supplementary material, Section S1–S2). We manually curated a look-up table for all 191 CCGs and all 225 NHS Trusts which were active as of 1 April 2020. Data are typically provided from around the time of the introduction of the transparency requirements, or the introduction of CCGs (typically as of 2012, but some combined output files result from regional consolidation). Longitudinal data (grouped into individual months as discussed above) were programmatically obtained from the public domain (in some format)[3] for 189 CCG and 205 Trust websites, and for 68 months for NHS England (additional information is available in online supplementary material, Section S1). The data predominantly originated from each individual CCG and Trust website, and failing this, was seldom found on data.gov.uk. It also occasionally incorporated bulk downloads hosted on nongovernmental organization websites hosting the outcome of previous Freedom of Information requests made by other interested parties.

A small number of Trusts did not publish data, and some published 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 and automated wrangling). Some Trusts regularly overlooked the need to maintain links either on their homepage, or on data.gov.uk, leading to several '404 Not Found' style errors. In total, our data curation process curates a total of 25,338 raw procurement datasets (12,709CCG files, 12,561NHS Trust files, and 68 monthly files from NHS England) which get passed as inputs into our data parsing module (online supplementary material, Section S1). The official guidance from HM Treasury (2018) 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, and in addition to issues of coverage described above, huge variation exists, making parsing the data a substantial challenge. Each row of each file must have had a minimum of data in

[3] The files are provided almost exclusively as .pdf, .csv, and .xls or .xlsx format. A small fraction was provided in.xlsb or.xlsm format.

Table 1. Distribution of data across file-types and institution

	Clinical Commissioning Groups				Trusts			
	PDF	Excel	CSV	Total	PDF	XLS	CSV	Total
Files	687	3,756	3,307	7,750	270	3,407	4,912	8,589
Rows	50,591	327,255	269,323	646,938	21,461	295,667	344,209	661,226
Value (£Mn)	16,364	100,256	87,579	204,111	3,287	38,493	45,159	86,931
Mean (£K)	323	306	325	316	153	130	131	131
Suppliers	2,165	4,933	3,982	7,973	1,380	6,883	7,631	11,203
	NHS England				Aggregated			
Files	0	67	1	68	955	7,213	8,148	16,316
Rows	0	565,206	36,555	601,761	72,052	1,188,128	650,087	1,909,925
Value (£Mn)	0	147,991	6,829	1,548,208	19,651	286,740	139,567	445,863
Mean (£K)	0	2,624	187	257	273	241	215	233
Suppliers	0	5,732	3,718	5,734	3,300	13,885	12,681	19,122

Note. Files relates to the number of files in each category. Rows refers to the total rows of data (total number of payments). Value (£Mn) refers to the total value of these payments. Mean (£K) refers to the average payment value. Suppliers refers to the number of unique suppliers.

three key fields to be included in our database: a date of payment, the value of the payment (which can be negative in the form of rebates), and the name of the supplier (which is latterly mapped to one of several institutional registers). Table 1 shows the distribution of the file-types across NHS Trusts, CCGs, and NHS England. Several files were seemingly (and somewhat unnecessarily) converted at source out of Comma Separated Values (.csv extensions) or Microsoft Excel (.xls extensions) style spreadsheets and into the more difficult to parse Portable Document Format (.pdf). There is a general lack of awareness of the Berners-Lee (2016) criteria,[4] where the barriers to saving a file in an open, tabular '3*' format in comparison to a '2*' format are almost nonexistent; 6% of files are in a .pdf format, and 45% are in the proprietary '2*' Excel file format (i.e. .xls or .xlsx extension).

Our custom parser crawled through the cached files sequentially, and then cleaned the merged database for certain attributes (online supplementary material, Section S1–S2). First, we drop information on payments under £25,000 (in absolute value) from the small number of institutions which (unnecessarily) provide it, to standardize reporting criteria 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 standardized 'dd-mm-yyyy' format. Supplier names must be strings greater than length three (to avoid ambiguity). We check for the presence of eight substrings that indicate a redacted supplier name (indicated as variations of '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 £14Mn. We find one payment which has the 'supplier name' entry of 'Various' (worth £54,996). We also drop rows where suppliers are purely numerical (but not alphanumerical). 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 £445,863Mn (£204,111Mn, £86,932Mn, and £154,821Mn for CCGs, Trusts, and NHS England, respectively). Note, importantly, that while all three of our substantive databases contain approximately the same number of rows (with shares of 33.387%, 34.62%, and 31.51%), the

[4] 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*; 'nonproprietary format (e.g. CSV instead of Excel)' ratings at best (Berners-Lee, 2016).

distribution of payment values (and unique suppliers) across the three datasets is less evenly distributed (45.78%, 19.5%, and 34.72%). This is commensurate with the hierarchical nature of the commissioning process where subcontracts from NHS Trusts to other providers are for relatively smaller amounts. Table 1 provides indicative evidence of this, in that the average payment made from Trusts is substantially lower in value (£131K) in comparison to payments made by CCGs (£315K) or NHS England (£257K). It must also be appreciated, for example, that some net positive value of payments from NHS England may go to CCGs or NHS Trusts, and some may flow from CCGs to NHS Trusts (and vice versa).

3.2 Registers for record linkage

Administrative registers for record linkage are essential for generating an understanding of the direction of payments, and to eliminate false positive matches from 'suppliers' in the transaction data to unique 'named entities'. We used 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 to provide consistency with supplier names in the procurement data. This comprised 4,490,685 unique strings, and if a normalized version of the supplier string is duplicated, both are dropped to avoid any ambiguity (4,489,992 unique normalized supplier strings remain; see Section 3.3). A similar procedure is followed for the CCEW (the nonministerial government department that regulates registered charities in England and Wales) which maintains the Central Register of Charities, and formed the core of our analysis. The Register contained 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.[5] There are 45,739 unique raw entries, and 45,560 normalized ones on this specific register, although we note that they may not all be potential 'providers' of goods and services. As indicated in Figure 2, 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 normalized unique names on both Companies House and the Charity Commission, 1,408 on both Companies House and NHS Digital, and 1,280 on both the Charity Commission and NHS Digital registers.

3.3 Reconciled dataset

Online supplementary material, S3 describes our automated matching procedure, otherwise known as the 'reconciliation' stage which links named suppliers to institutional registers. The institutions with the highest value of received payments from our matching 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 first of these two specific Trusts is not surprising: it is one of the UK's largest 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

[5] 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.

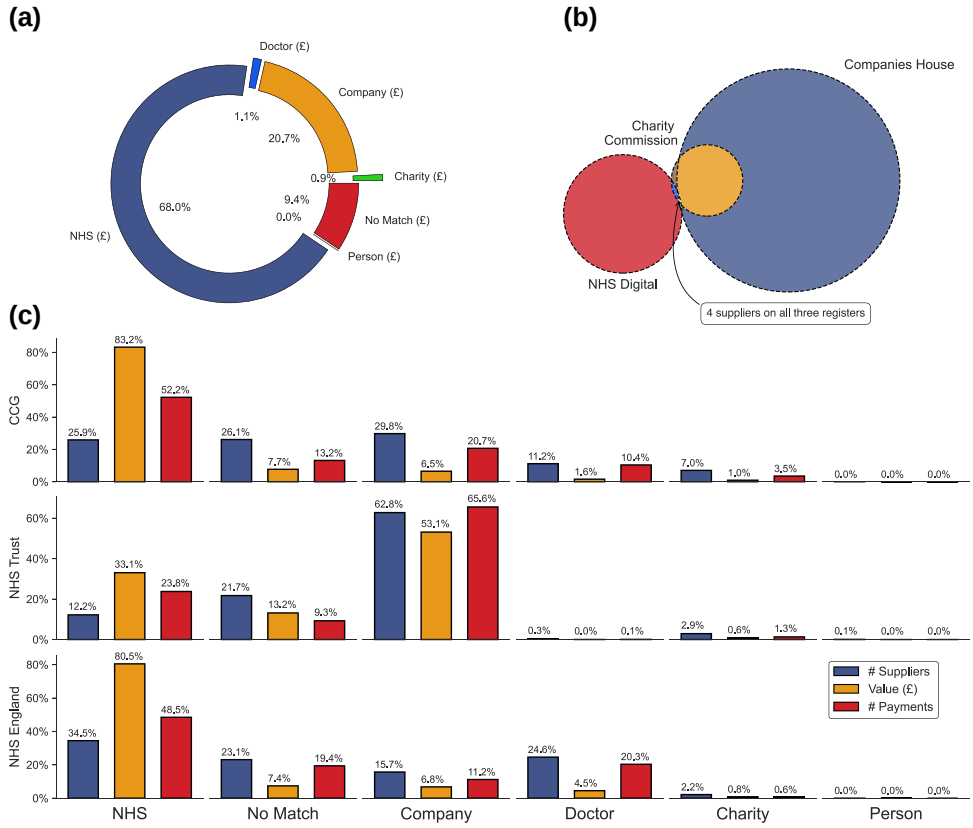

Figure 2. Match distribution across institution type: The doughnut chart (Panel 'a') shows the cumulative value of all payments made to one specific sector (including where an institution features on more than one register). The Venn diagram (Panel 'b') outlines the issue of institutional overlap between the named suppliers in our merged dataset. The three bar charts (Panel 'c') document the unique mapping of institutions exactly one of the six categories.

fact: that entities registered on NHS Digital (and the entities subsequently mapped to them) are comprised of two types of organizations: core components of the NHS itself, as well as health authorities and support (executive) agencies. A visualization of the reconstructed dataset with supplier reconciliation can be seen in Figure 2. For all entities matched to the CCEW (and a large subset of the highest value Company suppliers), we not only undertook a significant manual verification exercise pertaining to the validity of the multiple approximate string matches, but also a much deeper manual examination and an extensive database audit. Additional details pertaining to this data audit—in additional to all the raw and processed datasets—are available in the online supplementary material, indexed as Rahal et al. (2023).

One of the main generative overviews of our approach can be seen in Figure 2a, which shows the distribution of matches across all six registers combined across both the CCG and NHS Trust datasets. It shows that our approach can map over 90.5% of the total value of payments in our database. Cursory analysis indicates that approximately 2.57% of payments within the NHSSpend database go to 'public' entities outside of the NHS (4.0% by value)[6]; 45,876 supplier payment strings contain 'council', for example. In terms of unmapped payments (allocated to the 'No Match' category), the three most seen supplier strings relate to either ministerial departments or

[6] This number is approximated by searching for a range of terms related to public organizations within supplier string names. This list includes: 'council', 'municipality', 'local auth*', 'library', 'government', 'police', 'department of', 'her majest', and 'job centre'. We note that this list is subjective in the absence of a recognized lookup set. Most identified payments originate from CCGs (5.07% of the total number of payments), followed by Trusts (1.89%), with the fewest payments found originating from NHS England (0.62%). These payments are—on average—larger than those which are not flagged in such a way, with an average value of £401,919.

executive agencies which are not listed on NHS Digital; 'Inland Revenue CIS' (a Construction Industry Scheme for contractors), the National Offender Management Service (NOMS), and the Ministry of Justice (MoJ). Including these three main suppliers, an extensive manual review of all unmapped payments indicates that we can explain up to approximately 10.508% of unmapped payments by value to 'public sector' entities.

One immediate point of interest from Figure 2a is the large aggregate value of payments (20.7% and 0.9%) going to entities registered on the Companies House and CCEW registers; these numbers combined are substantially larger than the 6.1%–7.3% quoted in national government accounts as going to 'Expenditure on non-NHS bodies' (see Section 2). The figure in those accounts is however not a reasonable 'framing' of payment distributions (Powell & Miller, 2014) to use as a reference for two reasons. The first is that these numbers in Panel 'a' do not represent the final destinations of expenditures: money from NHS England in our database flows to CCGs, and then through to NHS Trusts, for example. 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). This itself is also substantively interesting, pointing as it does to the ways in which the NHS—and indeed public services more generally—are becoming subcontracted to specialist delivery vehicles, constructed as trading entities.

4 Procurement from third sector organizations

Our databases contain 36,639 (23,842 for CCGs, 9,008 for Trusts, and 3,789 for NHS England) payments (each with a minimum value of £25K) to entities matched onto the CCEW, worth a total of £3.860Bn (£1.980Bn, £0.556Bn, and £1.324Bn, respectively). These numbers—while slightly lower than NVCO estimates—are, like the NCVO analysis, based on a sub-sample of all NHS payments (i.e. a large majority), given the limited availability of a small amount of procurement data. We estimate that there are 1,845 unique CCEW-registered entities within our dataset, which is directly comparable to the NSTSO estimate of 2,000. While the proportion of suppliers from the CCEW 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).

4.1 Simple descriptive statistics

The highest value 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 frequency distributions

Table 2. Highest value Charity Commission for England and Wales suppliers

	Amount	Count	Reg #	ICNPO	Income	Rank
St Andrew's Healthcare	730.93	834	1,104,951	3,300	1,370	26
Nuffield Health	326.5	3,021	205,533	3,100	4,840	3
Healthcare QIP	110.95	224	1,127,049	3,400	150	427
Hospice UK	78.28	17	1,014,851	3,200	105	652
Healthcare Management Trust	613.26	348	292,880	4,100	177	350
Royal Hospital for Neuro Disability	58.24	140	205,907	3,100	2.9	256
Marie Curie	55.69	600	207,994	3,400	1,090	33
The Forward Trust	51.5	461	1,001,701	7,200	137	480
Horder Healthcare	50.12	448	1,046,624	3,400	192	304
MSI Reproductive Choices	43.33	1,117	265,543	3,400	1,600	16

Note. Top 10 institutions by procurement value mapped to the Charity Commission, ordered by the cumulative value of all contracts they receive (£Mn). Count refers to the number of payments made in our datasets. 'Income' (£Mn) and 'Rank' refer to their total income for the last full financial year of our NHSSpend dataset, as supplied by the Charity Commission (i.e. 2018/2019).

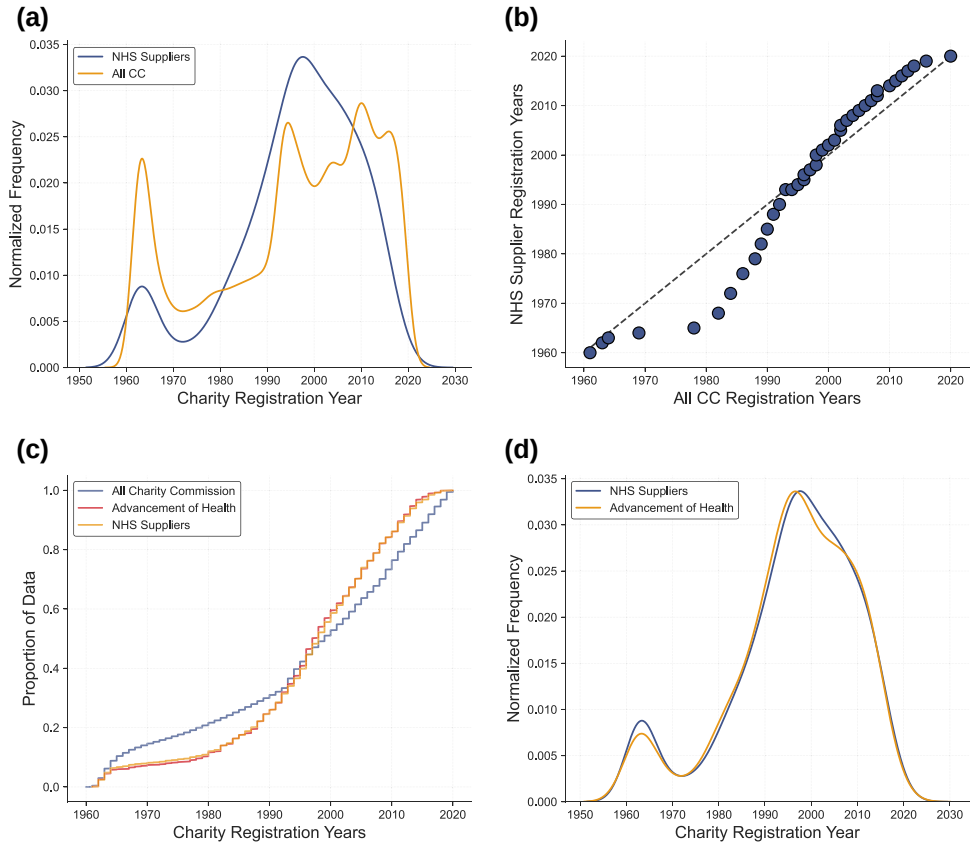

Figure 3. Analysing charity registration dates. Panel (a) plots the distribution of years of registration of all entries on the Charity Commission for England and Wales, and the subset of those that supply the NHS. Panel (b) is a Q-Q plot of the same two arrays. Panel (c) is an empirical cumulative distribution, which also considers all charities which report as the 'Advancement of Health' class, which is also visualized in contrast to all NHS suppliers in (d).

across each of the three databases is unsurprising. It is not just the largest healthcare charity within the UK, it also typically ranks among the top five charities by annual income.[7] Similarly, St. Andrews Healthcare is a large independent charity which provides highly specialized 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. The Trust which procured the largest proportion of its budget from a single CCEW-registered entity was Derbyshire Healthcare NHS Trust.

Figure 3 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 organizations have been registered relatively recently compared to the broader population of charities; they have been on the register for a mean 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 have been registered for approximately the same length of time. The modern register of charities was compiled from 1960 onwards, and large numbers of entities were in existence prior to that date, but there is no reliable data on precisely when they were founded.

[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.

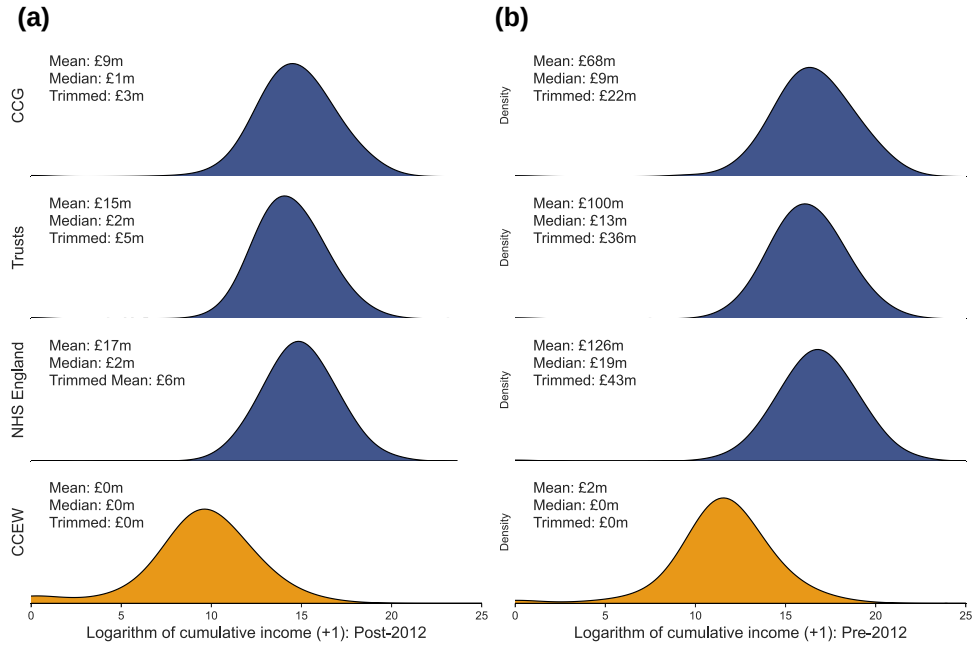

Figure 4. Income distributions across NHS suppliers. The figures show the logarithm of cumulative income as detailed on the Charity Commission for England and Wales (CCEW), both post-2012 (Panel 'a') and pre-2012 (Panel 'b'). The logarithm of cumulative income +1 is taken to deal with zero values on the CCEW. The trimmed mean is calculated by slicing 10% off either side of the distribution.

Figure 4 shows another characteristic of NHS supplying organizations: their cumulative income both before and after the implementation of the HSCA. Across the three tiers, those charities which supply the NHS are substantially larger, even when trimming the mean to consider the skewing effect of very large charities such as Nuffield Health. NHS England appears to procure from charities that are generally larger than the wider charity population, which might be expected as NHS England commissions more from national, as opposed to potentially smaller, regional suppliers. A widely used division of charities into income bands (Davis & Hornung, 2018) is used here; we decompose nonprofits into 'Micro' (<£10K), 'Small' (£10K to £100K), 'Medium' (£100K to £1Mn), 'Large' (£1Mn to £10Mn), 'Major' (£10Mn to £100Mn), and 'Super major' (>£100Mn). This is shown in Table 3, and—consistent with Figure 4—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 feature in the topmost income band (with incomes of over £100Mn 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 the value of procurement to entities registered on the CCEW. This is consistent with the wider population of charities, where financial resources are highly concentrated within a relatively small number of entities (Backus & Clifford, 2013).

4.2 Activities of charity suppliers, and services supplied

Having outlined the stylistic characteristics of CCEW suppliers in comparison to others which supply the NHS, we now explore exactly which roles they undertake; their automated and manually curated International Classification of Nonprofit Organizations (ICNPO) number (Section 4.2.1), their 'class' as provided by the CCEW (Section 4.2.2), an analysis of the free-text description of the charity (Section 4.2.3), an analysis of the Standard Industry Classifiers (Section 4.2.4), and an analysis of 'expense area' found within the raw transaction data (Section 4.2.5). This sheds light on

Table 3. Charity classification by income

	Minor	Small	Medium	Large	Major	Super-major
Financial year: 2011/2012						
Entire CCEW (%)	42.39	36.7	16.31	3.86	0.69	0.05
Organization 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	0.2	21.7	26.67	24.05	26.84
Financial year: 2018/2019						
Entire CCEW (%)	37.47	37.56	19.84	4.32	0.75	0.06
Organization 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	12.1	22.73	27.14	29.26

Note. 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, 'Organization Count' refers to the number of organizations that meet the thresholds without merging with NHS procurement data. The latter two rows show the distribution of payments by volume (count) and amount.

the distinctive contributions made by third sector organizations to the mix of services provided by the NHS, which is to be expected from arguments underpinning the rationale for third sector provision.

4.2.1 ICNPO analysis

The ICNPO—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 nonprofit sector overlooked by the existing Standard Industry Classifier system (Salamon & Anheier, 1992, 1996). We utilize the expansion designed by the NCVO in the UK, which adds a small number of appropriate new categories,[8] and we manually curate additional ICNPO numbers for charities elsewhere uncharacterized. With respect to Section 2, the NCVO (Hornung et al., 2020) estimate that the ICNPO grouping of 'Health' (10%) has the largest proportion of large, major, and super-major organizations compared to 4% of those in other sectors. Analysis by ICNPO number can be seen in online supplementary material, Table S1. As might have been expected given the nature of the commissioning—and consistent with the other modes of analysis in this section—it is those organizations within the 'Health' classifications (3100, 3200, 3300, and 3400) and the umbrella 'Social Sciences' categorization (4100), which receives the largest number of payments and the cumulative highest values. There are differences across the commissioning hierarchy, which reflect differences in the responsibilities of each of these commissioning entities. For example, a high proportion of services are commissioned by NHS England from mental health charities (ICNPO class 3300) because these are highly specialized, and only undertaken by a few organizations which possess expertise. Charities across almost every ICNPO receive payments, showing the vast diversity and the range of contributions made by the sector in general, as well as perhaps hinting at the challenges of

[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.

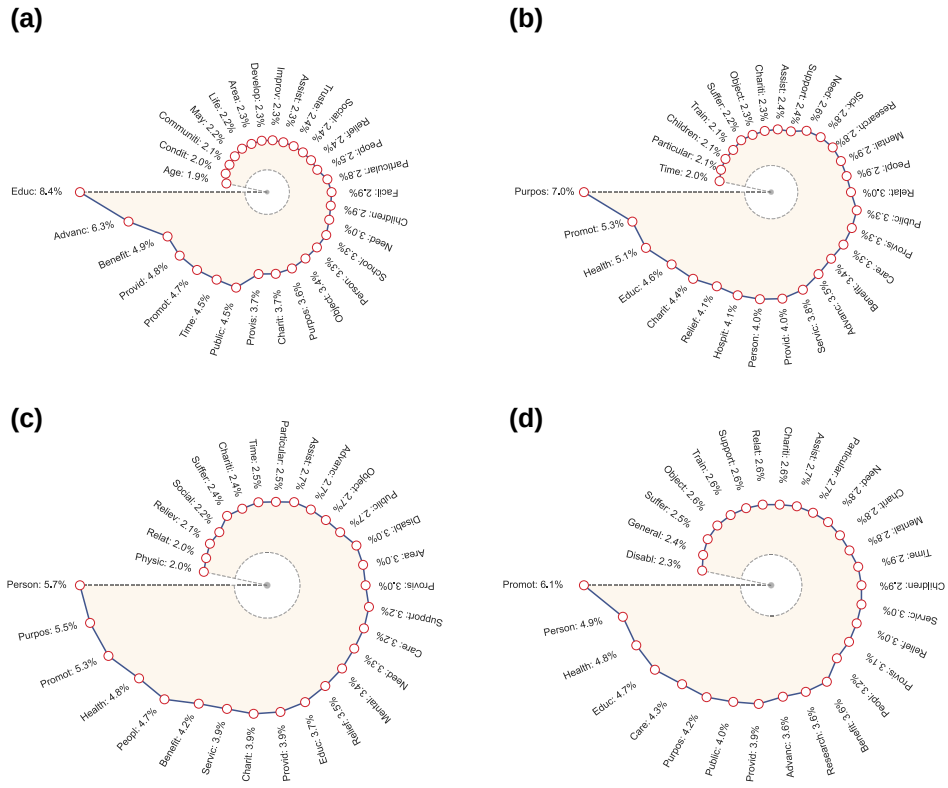

Figure 5. Analysing supplier 'Objects'. The first panel (a) visualizes the frequency distribution of cleaned, stemmed words found in the descriptive ambitions of all charities registered and active on the Charity Commission for England and Wales. Panels (b)–(d) visualize this for just the subset of charities which supply each of NHS Trusts, Clinical Commissioning Groups, and NHS England, respectively.

classification in the field, largely based on analyses of textual information provided in the governing documents of charities, rather than the self-assigned choice of Standard Industrial Classification (SIC) code at registration. We turn to this next in Section 4.2.4.

4.2.2 Charity classes

Appreciating the potential limitations of the ICNPO classification typology, we now extend our analysis in multiple directions. There are two main differences across the analysis concerning ICNPO categories and that which considers 'classes' provided by 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 ICNPOs are unique, charities can register as multiple classes. This distribution is seen in online supplementary material, Table S2. In addition to the 'Advancement Of Health Or Saving Of Lives' category, most payments are made to service providers and those that provide services to children, young people, and the differently abled.

4.2.3 Charity descriptions

Figure 5 provides a visualization of word frequency counts (after the cleaning of 'stop words' and the 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, terms such as 'health', 'care', 'sick', and 'suffer' appear frequently in the classes of CCEW organizations which supply the NHS, but not in the distribution of CCEW classes 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'.

Table 4. Distribution of payments across Standard Industrial Classification (SIC) codes for non-Charity Commission for England and Wales entities (CCEW)

		Trusts		CCGs		NHS England	
Code	Name	Count	£	Count	£	Count	£
86900	Other human health activities	10.4	7.2	27.4	37.6	15.5	22.8
46460	Wholesale of pharmaceutical goods	10.3	5.3	0.2	0.1	0.1	0.1
86210	General medical practice activities	2.8	2.9	14.4	19.8	10.2	7.3
86101	Hospital activities	2.7	6	17.5	22.7	7.2	24.1
84110	General public administration activities	7.4	9.3	2.5	3.5	1.8	3.9
78200	Temporary employment agency activities	6.8	7.8	0.7	0.3	0.4	0.5
82990	Other business support service activities n.e.c.	4.5	4.8	4.5	3.1	3.6	1.7
62090	Other information technology service activities	5	4.3	1.7	1.5	4.3	2.9
62020	Information technology consultancy activities	3.8	2.7	1.6	1.0	7.8	6.3
21100	Manufacture of basic pharmaceutical products	4.8	2.5	0	0.0	0	0

Note. Payments made to entities not registered on the CCEW. Ordered by total amount of payments. Top ten highest value SIC codes shown. All values are percentages; 'Count' represents percent of payments by volume, and '£' represents percent of payments by value.

Further work along this trajectory would consider dynamic topic models to see how the involvement of nonprofits in the NHS changes over time, which is something we consider from a slightly different perspective in Section 4.3.

4.2.4 Charities' standard industry classifiers

We also examine the distribution of Standard Industry Classifiers as provided by Companies House. We do so by comparing entities registered on Companies House and the CCEW, and those exclusively on Companies House.[9] Unsurprisingly, for each of the two types of suppliers, the SIC code '86900—Other human health activities' dominates (14.487% of the value of all non-CCEW registered entities, and 42532% for CCEW-registered entities). However, following this, a divergence occurs. For non-CCEW-registered entities, the most seen SIC codes (following 86900) are; '46460—Wholesale of pharmaceutical goods', '86210—General medical practice activities', '86101—Hospital activities', '84110—General public administration activities', and '78200—Temporary employment agency activities' (see Table 4). We expect the latter to be more prominent in the case of NHS Trusts, which are service providers, in comparison to agencies which commission NHS services, such as CCGs, or NHS England. Table 5 shows the distribution over SIC codes for CCEW registered entities, ordered by total procurement amount, for each of the Trusts, CCGs, and NHS England datasets. Specifically, specialized charity-specific services become immediately apparent, primarily in the form of social work and care (residential and nonresidential) activities. The most noticeable disaggregated charity-specific supply of services comes at the NHS England level in the form of goods and services procured from '87200—Residential care activities for learning difficulties, mental health, and substance abuse' (57.772% of all NHS England procurement from CCEW-registered suppliers by value, and 19.336% by payment count). Such patterns are broadly what we might expect for the third sector. The aggregate pattern of NHS commissioning would show a very substantial concentration of spending on acute hospital services, whereas spending by NHS Trusts and by NHS England is often for more specialized activities which plays to third sector strengths in areas that require greater personalization and long-term support.

4.2.5 Specific services supplied by CCEW-registered entities

Our approach also allows us to gain insight into specific services being provided by organizations registered on the CCEW in contrast to other services offered by other providers. While there is

[9] Entities can submit up to four SIC codes per registration. We expand this list for each entity, and map payments to each registered SIC code to treat each code equally, as opposed to focussing only on the first mentioned code.

Table 5. Distribution of payments across Standard Industrial Classification (SIC) codes for Charity Commission for England and Wales entities (CCEW) entities

		Trusts		CCGs		NHS England	
Code	Name	Count	£	Count	£	Count	£
86900	Other human health activities	65.2	63.8	48.5	55.2	32.7	16.6
88990	Other social work w/o accommodation n.e.c.	22.6	27.1	15.3	10.5	25.5	6.7
86101	Hospital activities	2.5	3.2	14.2	19.5	4.0	1.5
87200	Residential care activities for learning difficulties	8.8	8.1	8.3	5.2	19.3	57.8
86220	Specialists medical practice activities	4.8	3.5	8.3	9.2	8.9	3.9
88100	Social work w/o accommodation (elderly and disabled)	3.3	2.7	8.9	6.2	5.4	1.0
87100	Residential nursing care facilities	2.7	2.0	9.5	9.24	1.9	0.6
87900	Other residential care activities n.e.c.	2.3	2.5	5.3	2.8	16.2	6.1
87300	Residential care activities for the elderly and disabled	1.73	1.4	6.4	3.4	4.1	2.3
94990	Activities of other membership organizations n.e.c.	1.3	1.1	4.0	2.2	1.6	6.3

Note. Payments made to CCEW-registered entities only. Ordered by total amount of payments. Top 10 highest value SIC codes shown. All values are percentages; 'Count' represents percent of payments by volume, and '£' represents percent of payments by value.

guidance that individual lines of procurement data should never reveal explicit 'invoice detail', it does mandate that information on 'expense areas'—descriptions of broad categories of NHS service provision—is provided. Our data extraction pipeline was unable to obtain information on 'expense areas' if it was not present in raw files, nor was it able to be extracted when present in a small number of low-quality files (predominantly in the case of 1* .pdf files). However, it was available for 93.51% of all transactions (100% for NHS England, 99.25% in CCGs, and 81.97% of NHS Trusts; see Table 1 and Section 3.1 for a discussion regarding data quality, to which this pertains, especially regarding the higher prevalence of .pdf files for Trusts). Analysis of this can be seen in online supplementary material, Figure S2. While it is expected and is shown that most of the procurement from entities which were not registered on the CCEW were involved in the provision of services related to 'Acute Commissioning', 'Primary Care' (and Secondary Dental), and—in the case of NHS Trusts—'Ambulance Services', entities listed on the CCEW provided a different kind of service. For example, CCGs primarily expend in 'Hospices' when procuring services from CCEW-listed entities, as well as in the areas of planned and palliative care. Expenditure by NHS England is typically dominated by 'Specialized Commissioning'. This features entities that often operate contracts for relatively small numbers of individuals across many different sites, such as specialist psychiatric services supporting people experiencing specific, severe and sometimes long-term challenges.

4.3 Temporal variation in the role of nonprofits

The intention of the HSCA 2012 was to foster competitive markets by opening them to 'external' organizations, including, but not limited to the nonprofit sector. Payments across each ICNPO number are relatively consistent across procuring organizations, and over time. The exception to this is procurement from nonprofits 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 11 months of the year. This likely indicates the annual contractual nature of this specific type of service supplied at this level of the procurement chain.

In terms of absolute change over time, Figure 6 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 nonprofits (ranging from 4.7% to 2.9% per annum, as shown in

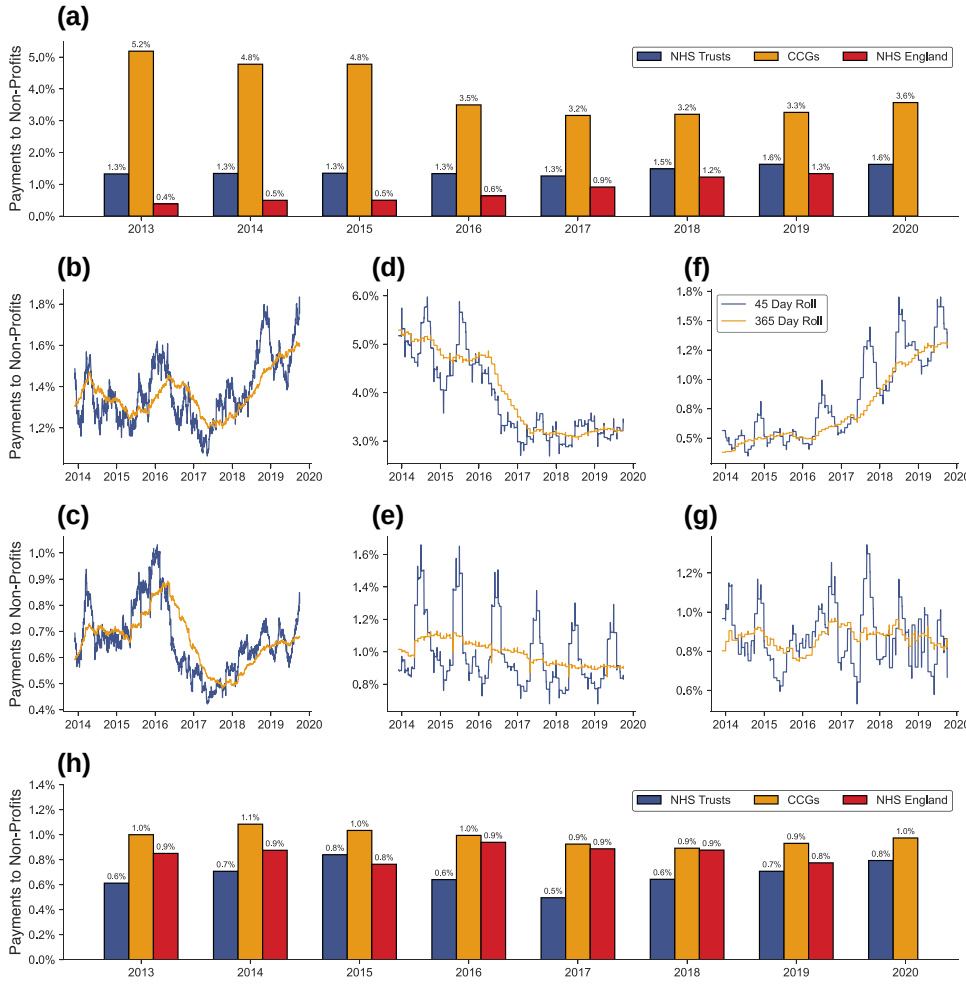

Figure 6. Temporal variation in procurement from nonprofits. Panels (a) and (h) show the aggregation over years for each of the Trust, Clinical Commissioning Group (CCG), and NHS England datasets, showing the cumulative percent of procurement from nonprofits for that organization type. Panels (b)–(g) plot this daily 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 the number of payments, and the lower represents payment value. *Note*: NHS England data curation finishes near to the start of 2020.

Figure 6a), there is a resounding level of consistency in the 'amount' of payments made across the hierarchy of procurement. CCGs, for example, procure at least 0.9% of their goods and services form the nonprofit sector, and at most 1.1% (as shown in Figure 6h). In addition to the annual contractual pattern mentioned above (which can also be seen in Figure 6f), two other patterns emerge. The first is the slight upward trend in the number of payments made to nonprofits by NHS England, 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 nonprofit sector. Online supplementary material, Figure S3 shows a 'concentration ratio': the ratio of the value of payments going to the top five, 10, and 20 nonprofit organizations 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 10 highest value, and 0.609–0.446 for the 20 highest value recipients between the years of 2013 and 2019. We were also able to consider two additional dimensions which analyse 'market entry' (the first time a unique organization is seen in our dataset), and the median value of payments to CCEW-registered suppliers. Figure 7a directly allows us to

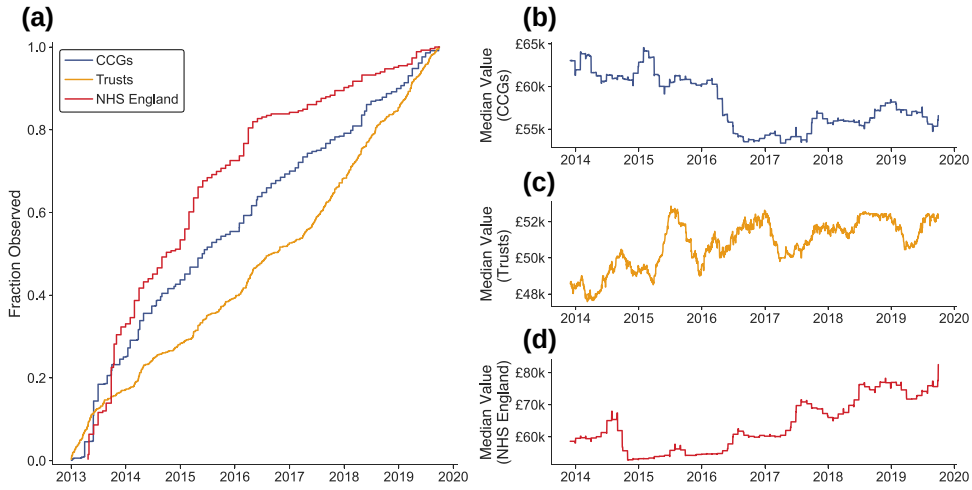

Figure 7. Market entry and median payment values. Panel (a) depicts the rate at which organizations are first seen in the NHSSpend dataset. Panels (b)–(d) depict the median payment value over time.

evaluate the speed at which new entities began to supply the NHS. Suppliers of NHS England were most quick to provide services and be seen within the NHSSpend dataset, with relatively fewer entities continuing to enter. NHS Trusts were slower to procure from a wider variety of CCEW-registered entities, likely due to their smaller, more diffuse nature in comparison to NHS England. The median value of NHS transactions to CCEW-registered entities increased substantially for entities supplying NHS England, decreased slightly for CCGs, and was relatively stable for NHS Trusts (Figure 7b).

We additionally undertake a series of regressions to more formally test whether third sector involvement in NHS procurement is increasing. First, we run regressions of the form $y_t = \alpha + \beta T + \lambda D_t + \varepsilon_t$ on (calendar) monthly counts and amounts as percentages of total spend which go to entities registered on the CCEW (upper third of Table 6), delineating spend by tier in the hierarchy and as aggregated together. For different estimations, the time trend is both significant and insignificant. We next consider rolling windows of the same percentages—counts and amounts (lower two-thirds of Table 6)—for both 90- and 365-day rolling windows (as visualized in the middle two rows of Figure 6). Here, all coefficients are 'statistically significant' at the 0.01 level, although the time trend coefficients are small in value and may be considered 'economically insignificant' (and for volume of payments made by CCGs, even negative).

4.4 Types of charitable entities and CICs

Finally, we can decompose the various types of corporate entities which supply NHS organizations, enabling the identification of entities other than registered charities. Data held by Companies House allows the identification of CICs: prosocial enterprises operating for the good of the community, rather than to make a profit for the owners. While most corporate suppliers of the NHS organizations are in the form of Private Limited Company (82.758% by unique organizations, 83.1778% by volume of payments, and 78.798% by payment amount to companies), there is also evidence of substantial involvement by CICs (202 organizations, 2.025% by payment volume, and 3.857% by payment amount to companies). Charitable Incorporated Organizations (55 organizations, 0.0562% by volume, 0.0355% by amount to companies) feature commonly, and in two rare instances, so do Scottish Charitable Incorporated Organizations (0.002% by volume, 0.001% by amount to companies). As expected (by virtue of regulations about the governance and management of different types of entity), no CICs are also simultaneously matched to the CCEW in our database, and a manual review of the 202 CIC entities indicates that they are all denominated as CICs via their registered 'Company Name' (by including 'CIC', 'C.I.C', or 'Community Interest Company'), even if it does not appear

Table 6. Time-series regression on monthly and rolling counts and amounts

Monthly values		All	Trusts	CCGs	NHS England
	Constant	0.848***	1.253***	5.1	0.204***
Count	Trend	0.000	0.004***	−0.0272	0.015***
	April	0.385***	−0.099	0.25233	0.06
	Constant	1.383***	0.699***	0.944***	0.853***
Amount	Trend	0.013***	−0.001	−0.001	0.000
	April	0.047	−0.054	0.849***	−0.01
90-day roll		**All**	**Trusts**	**CCGs**	**NHS England**
Count	Constant	1.574***	1.254***	5.110***	0.275***
	Trend	0.000***	0.000***	−0.001***	0.000***
Amount	Constant	0.955***	0.743***	1.084***	0.884***
	Trend	0.000***	0.000***	0.000***	0.000***
365-day roll		**All**	**Trusts**	**CCGs**	**NHS England**
Count	Constant	1.5***	1.294***	5.322***	0.252
	Trend	0.000***	0.000***	−0.001***	0.000***
Amount	Constant	0.946***	0.74***	1.097***	0.854***
	Trend	0.000***	0.000***	0.000***	0.000***

Note. *** indicates statistical significance at the 0.01 level.

in the 'raw' supplier name in the procurement data.[10] CICs supply £2.733Bn in total, which represents 0.61% of the entire NHSSpend database, with this percent rising from 0.508% in 2013 to 0.78% in 2019. The CIC that supplies the highest value of healthcare services is City HealthCare Partnership CIC (£368Mn). Unsurprisingly, the highest value 'expense area' within which these services are procured is 'Community Services' (£1.361Bn), far greater than the second (Public Health, £195Mn) and third highest expense areas ('Health and Justice', £187Mn). These figures are very large by the standards of many charities. The reason is that a small number of CICs have been established through the process of spinouts of what were formerly components of the NHS to provide community-based services. Thus, they were very large from the point of establishment, which is reflected in the size of the organizations and of the payments made to them.

5 Discussion and conclusion

The procurement data on which we have drawn has presented an opportunity for new insights into the pattern of contracting for NHS services from the voluntary sector. These insights concern the distinctive nature of the niches occupied by nonprofit organizations, including their position in the NHS commissioning hierarchy and the fields of service provision in which they operate; whether opportunities for the voluntary sector have expanded or not, and whether the result has been to reinforce the concentration of resources in large entities. We also document the types of nonprofit entity—charities or social enterprises—that are successful in taking up the opportunities through recent changes in the NHS. We have argued that there are good reasons why the NHS might wish to involve nonprofit organizations in the delivery of health care, because of the distinctive 'goods' offered by such entities. That is their ability to respond to diverse and specific categories of need of service users, and the ability to provide care in circumstances where conventional market provision is impossible because of the challenges of incomplete information about the severity and likely progression of a health condition.

[10] An example of this would be 'Hope Citadel Healthcare', legally registered as 'Hope Citadel Healthcare Community Interest Company'.

In terms of estimating the scale of provision, we show that the numbers of organizations involved are broadly comparable with estimates from previous national social surveys. Estimates in the region of 2,200 charities emerge from surveys and, though these are now over a decade old, the slightly lower figures from our exercise can be judged comparable not least because, with a minimum threshold of a payment in any one month of £25K, the procurement data we have used will exclude many small charities which may play a subsidiary role in delivering NHS services (perhaps receiving subcontracts from a larger charitable provider). The data we have used allows more to be said than has been possible with sources used in previous research in relation to the position of third sector organizations in the delivery of services. The accounts-based data captured by NCVO contains some information on the source of funding; there is enough information to judge that money has emanated from the NHS, but not always from what sort of entity. Our tabulations of data from CCGs, NHS Trusts, and NHS England, augmented with information about the broad 'expense area' (which characterizes the services being purchased) builds on this previous research. It points to the roles played by third sector organizations in meeting areas of specialist need. The prominence of third sector organizations working in mental health receiving contracts from NHS England is a good illustration. These are specialist services, commissioned on a national basis to ensure that needs are met, and they are fields in which provision by third sector organizations is in line with expectations, given the rationale for the existence of third sector organizations.

Part of the motivation for this research was to chart changes as a result of the legislation of the post-2010 UK Coalition government to promote greater competition by permitting contracts to be placed with 'Any Qualified Provider' via the HSCA 2012. Pre-2012 data are not available from our source, so we cannot definitively draw conclusions about the situation prior to the passage of that Act. For the time period for which data were available, we were unable to document consistent support for a substantially increasing involvement of CCEW-registered entities following the Act's implementation. However, we did find some evidence that opportunities had opened up to a wider population of organizations, based on evidence from data on the concentration of NHS contracts across organizations of varying sizes. The financial resources of registered charities are known to be highly concentrated in small numbers of very large entities. In addition to showing how the distribution of institutions supplying the NHS varies across well-defined size boundaries, we show how these concentrations have varied over time, which is critically important given the commentaries and concerns surrounding the HSCA 2012. The reductions in these concentrations provide support in favour of the 'opening up' of the healthcare market and alleviate concerns that it would allow a small number of institutions to consistently consolidate their power. There have been long-running debates about whether big charities are becoming more dominant in financial terms—that is, securing greater shares of resources (Backus & Clifford, 2013)—but there is no evidence that the biggest charities account for a growing share of total voluntary sector income received from the NHS over the period of analysis.

Finally, in terms of the changing character of the third sector, we find a growth in the value of payments to social enterprises registered as CICs. This is in line with expectations; this legal form was introduced in 2004 in part to provide an alternative to registration as a charity for those wishing to form enterprises with a social purpose. A distinctive feature is that many of the largest such providers, receiving over £100Mn each, are former sections of NHS Trusts which have been 'spun out' of the service into private nonprofit ownership. This has been a deliberate aim of government policy from before 2010, but the procurement data seems to imply an acceleration since that date. There remains considerable potential for further work on these datasets, but we have not developed additional lines of inquiry for reasons of space. One possibility would be to map the spatial distribution of contracts in more detail; online supplementary material, Figures S4–S5 plot this by Trust and CCG. However, further work is required to examine whether is there greater reliance on third sector organizations in some locations compared to others, potentially as a function of population age structure, the uneven distribution of third sector organizations, or unequal demand due to health inequalities. There is evidence that contracts are placed with local organizations, or do we find that some large entities deliver services across a wide range of locations and are able to respond to a diverse range of geographically varied opportunities? To what extent do we find that individual voluntary organizations are heavily reliant on contracts from specific commissioners of services to deliver the bulk of their income? As successive waves of commissioning data are accumulated and further policies are implemented, we will be better able to causally analyse the dynamics of contract

placement, such as whether individual charities or social enterprises develop long-term relationships with commissioners, and better analyse policy changes ex-ante.

While the UK generally leads the world in terms of transparency in public processes, more work is needed to evaluate the evolution and impact of procurement policies in response to policy changes such as the HSCA 2012. We find some evidence of change over the period since that Act, but arguably it is too early to judge whether the aspirations of the Act to create 'the largest and most vibrant social enterprise sector in the world' have been fulfilled, despite the fact that procurement data are uniquely positioned to evaluate such claims. Recent regulation regarding the compulsory integration of Contracts Finder for some procuring agents (Cabinet Office, 2021) updates legal and policy requirements. This goes some way towards making such data more accessible for academic and policy-relevant research, such as, for example, by mandating the use of unique institution identifiers to facilitate a better understanding of subcontracting relationships between hierarchical parts of the supply chain. The more complete reporting (and indexing) of procurement data would allow a deeper decomposition of how significant the provision of health (and social) care is for the financing of individual organizations. It also allows a deeper, discontinuity-based analysis given that for future policy implementations, transparently available procurement data will be available for pretreatment and posttreatment analysis and comparison. This goes far beyond the HSCA 2012 in other ways (and the recommendations previously made by Carter, 2016); transparently available procurement data will facilitate a better understanding of the contributions that public, private, and charitable organizations make to the functioning of universal healthcare systems across the world.

Acknowledgments

The extensive code library which accompanies this work can be found at github.com/crahal/NHSSpend. The authors are grateful to comments on earlier versions of the work from Mark Exworthy, David Stuckler, Martin Mckee, Lucy Reynolds, and James Rees. Technical research assistance provided by Ian Knowles. Insightful comments were gratefully received from participants at the International Conference for Administrative Data Research, the Economic Insights team at the Office for National Statistics, the Spatial Unit at the Department for Levelling Up, the Government Data Science Community Meetup, two editors, and two anonymous referees.

Conflicts of interest: The authors declare no conflicts of interest.

Funding

The work originates from a scoping and prototyping exercise funded by the ESRC (grant numbers ES/M010392/1 and latterly ES/X000524/1), with majority funding latterly and gratefully acknowledged from the British Academy and the Leverhulme Trust (Grant RC-2018-003), the Leverhulme Centre for Demographic Science (LCDS), and Nuffield College. Insightful comments were gratefully received from participants at the International Conference for Administrative Data Research, the Economic Insights team at the Office for National Statistics, the Spatial Unit at the Department for Levelling Up, the Government Data Science Community Meetup, two editors, and two anonymous referees.

Data availability

The entirety of this project is open and available online at github.com/crahal/NHSSpend, and this GitHub repository is indexed into Zenodo as Rahal et al. (2023).

Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series A*.

References

Alvarez-Rodríguez J., Gayo J. L., Silva F., Alor-Hernández G., Sanchez-Ramirez C., & Luna J. (2012, May). Towards a pan-European e-procurement platform to aggregate, publish and search public procurement notices powered by linked open data: The moldeas approach. *International Journal of Software Engineering and Knowledge Engineering (IJSEKE)*, *22*(3), 365–384. https://doi.org/10.1142/S0218194012400086

Alvarez-Rodriguez J. M., Vafopoulos M., & Llorensm J. (2015). Enabling policy making processes by unifying and reconciling corporate names in public procurement data. The CORFU technique. *Computer Standards and Interfaces 41*(1), 28–38. https://doi.org/10.1016/j.csi.2015.02.009

Backus P., & Clifford D. (2013). Are big charities becoming more dominant? Coss-sectional and longitudinal perspectives. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(3), 761–776. https://doi.org/10.1111/j.1467-985X.2012.01057.x

Berners-Lee T. (2016). Linked data. https://www.w3.org/DesignIssues/LinkedData.html.

Billis D., & Glennerster H. (1998). Human services and the voluntary sector: Towards a theory of comparative advantage. *Journal of Social Policy*, *27*(1), 79–98. https://doi.org/10.1017/S0047279497005175

Cabinet Office (2021). Procurement policy note—update to legal and policy requirements to publish procurement information on contracts finder. *Action Note PPN 07/21*.

Carter P. R. (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*. Department of Health.

Davis J., & Hornung L. (2018, November). Britains biggest charities: Key features. *NCVO Research Briefing*.

Delamothe T., & Godlee F. (2011). Dr Lansley's Monster. *The BMJ*, *342*. https://doi.org/10.1136/bmj.d408. PMID: 21257659.

Department for Communities and Local Government (2014). Local Government Transparency Code 2015. (February).

Department for Communities and Local Government (2015). Local Government Transparency Code 2015. (February).

Field D., & Johnson I. (1992). Volunteers in the British hospice movement. *The Sociological Review*, *40*(1_suppl), 198–217. https://doi.org/10.1111/j.1467-954X.1992.tb03393.x

Hansmann H. B. (1979). The role of nonprofit enterprise. *Yale LJ*, *89*, 835. https://doi.org/10.2307/796089

Higgins J. (1988). *The business of medicine: Private health care in Britain*. Springer.

HM Treasury (2018). Public expenditure statistical analyses. https://assets.publishing.service.gov.uk (July).

Hornung L., Chan O., Jochum V., Lawson M., McGarvey A., & Rooney K. (2020). *UK civil society almanac*. National Council for Voluntary Organisations.

Leys C. (2001). *Market-driven politics: Neoliberal democracy and the public interest*. Verso.

Leys C., & Player S. (2011). *The plot against the NHS*. Merlin Press.

Miller R., & Lyon F. (2016). Spinning with substance? The creation of new third sector organisations from public services. In *The third sector delivering public services* (pp. 87–106). Policy Press.

Newcombe H. B., Kennedy J. M., Axford S. J., & James A. P. (1959). Automatic linkage of vital records. *Science*, *130*(3381), 954–959. https://doi.org/10.1126/science.130.3381.954

Powell M., & Miller R. (2014). Framing privatisation in the English national health service. *Journal of Social Policy*, *43*(3), 575–594. https://doi.org/10.1017/S0047279414000269

Rahal C. (2018). The keys to unlocking public payments data. *Kyklos*, *71*(2), 310–337. https://doi.org/10.1111/kykl.v71.2

Rahal C. (2019). Tools for transparency in central government spending. *International Journal of Population Data Science*, *4*(1), 1–10. https://doi.org/10.23889/ijpds.v4i1.1092

Rahal C., Knowles I., Barnard S., & Mohan J. (2023, December). Introducing NHSSpend: Data and code to parse, harmonize and reconcile NHS procurement data.

Rowland D. (2019). Flawed data? Why NHS spending on the independent sector may actually be much more than 7%. *Blog Entry, British Politics and Policy, London School of Economics*, *1*.

Salamon L. M., & Anheier H. K. (1992). In search of the non-profit sector II: The problem of classification. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, *3*(3), 267–309. https://doi.org/10.1007/BF01397460

Salamon L. M., & Anheier H. K. (1996). *The international classification of nonprofit organizations: ICNPO-Revision 1, 1996*. Johns Hopkins University Institute for Policy Studies Baltimore Mar.

Salter B. (1995). The private sector and the NHS: Redefining the welfare state. *Policy & Politics*, *23*(1), 17–30. https://doi.org/10.1332/030557395782227366

Svátek V., Mynarz J., Wecel K., Klímek J., Knap T., & Nečaský M. (2014). *Linked open data for public procurement* (pp. 196–213). Springer International Publishing.

The National Archives (2001). Open government license. http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/.

Weisbrod B. (2009). *The nonprofit economy*. Harvard University Press.