RESEARCH ARTICLE

Enrolment in clinical research at UCLH and geographically distributed indices of deprivation [version 1; peer review: awaiting peer review]

Holger Engleitner[1], Ashwani Jha[1], Daniel Herron[2], Amy Nelson[1], Geraint Rees[1,3], Nick McNally[2], Bryan Williams[2,4], Parashkev Nachev[1]

[1]UCL Queen Square Institute of Neurology, University College London, London, WC1N 3GB, UK
[2]NIHR University College London Hospitals Biomedical Research Centre, University College London Hospitals NHS Trust, London, NW1 2BU, UK
[3]Faculty of Life Sciences, University College London, London, WC1E 6BT, UK
[4]UCL Institute of Cardiovascular Sciences, University College London, London, WC1E 6BT, UK

v1 **First published:** 14 Dec 2021, **6**:342
https://doi.org/10.12688/wellcomeopenres.17300.1
Latest published: 14 Dec 2021, **6**:342
https://doi.org/10.12688/wellcomeopenres.17300.1

Abstract

Healthcare should be judged by its equity as well as its quality. Both aspects depend not only on the characteristics of service delivery but also on the research and innovation that ultimately shape them. Conducting a fully-inclusive evaluation of the relationship between enrolment in primary research studies at University College London Hospitals NHS Trust and indices of deprivation, here we demonstrate a quantitative approach to evaluating equity in healthcare research and innovation.

We surveyed the geographical locations, aggregated into Lower Layer Super Output Areas (LSOAs), of all England-resident UCLH patients registered as enrolled in primary clinical research studies. We compared the distributions of ten established indices of deprivation across enrolled and non-enrolled areas within Greater London and within a distance-matched subset across England. Bayesian Poisson regression models were used to examine the relation between deprivation and the volume of enrolment standardized by population density and local disease prevalence.

A total of 54593 enrolments covered 4401 LSOAs in Greater London and 10150 in England, revealing wide geographical reach. The distributions of deprivation indices were similar between enrolled and non-enrolled areas, exhibiting median differences from 0.26% to 8.73%. Across Greater London, enrolled areas were significantly more deprived on most indices, including the Index of Multiple Deprivation; across England, a more balanced relationship to deprivation emerged. Regression analyses of enrolment volumes yielded weak biases, in

Open Peer Review

Reviewer Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

favour of greater deprivation for most indices, with little modulation by local disease prevalence.

Primary clinical research at UCLH has wide geographical reach. Areas with enrolled patients show similar distributions of established indices of deprivation to those without, both within Greater London, and across distance-matched areas of England. We illustrate a robust approach to quantifying an important aspect of equity in clinical research and provide a flexible set of tools for replicating it across other institutions.

Keywords
Equity, fairness, clinical research, geostatistics, deprivation.

Corresponding author: Parashkev Nachev (p.nachev@ucl.ac.uk)

Author roles: Engleitner H: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Jha A**: Conceptualization, Formal Analysis, Investigation, Methodology, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Herron D**: Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Validation, Writing – Review & Editing; **Nelson A**: Conceptualization, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Review & Editing; **Rees G**: Conceptualization, Formal Analysis, Funding Acquisition, Resources, Writing – Review & Editing; **McNally N**: Conceptualization, Investigation, Methodology, Project Administration, Resources, Validation, Visualization, Writing – Review & Editing; **Williams B**: Conceptualization, Funding Acquisition, Resources, Writing – Review & Editing; **Nachev P**: Conceptualization, Formal Analysis, Funding Acquisition, Methodology, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: A.J. has received fees from Sanofi/Genzyme, Britannia and Novartis. The other authors have no competing interests to declare.

Grant information: This work was funded by Wellcome (213038; to PN, AJ and GR); and the NIHR UCLH Biomedical Research Centre.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Engleitner H, Jha A, Herron D *et al.* **Enrolment in clinical research at UCLH and geographically distributed indices of deprivation [version 1; peer review: awaiting peer review]** Wellcome Open Research 2021, **6**:342 https://doi.org/10.12688/wellcomeopenres.17300.1

First published: 14 Dec 2021, **6**:342 https://doi.org/10.12688/wellcomeopenres.17300.1

Introduction

Equity of care is a central tenet of medicine. Its pursuit has conventionally focused on the structures of healthcare delivery[1] rather than the research and innovation that precede them[2]. But a failure of equity may arise at any stage during the translation of biological insights into front-line care. Unrecognized or mischaracterised heterogeneities in the underlying biology, in environmental factors—social, cultural, and physical—in access to specific therapeutic approaches, in trust of and engagement with healthcare services, may all result in unwarranted variations in the outcomes following any treatment[3,4]. As machine learning, in our pursuit of closer individuation, widens the field of factors brought to bear on clinical decision making, the effects of such heterogeneities may be magnified, potentially increasing the disparities between those from whom the guiding intelligence is drawn and those to whom it is merely applied[5].

How should equity of care be promoted in innovation? We must first be clear about what the notion, rooted in Aristotle's *epieikeia*[6], implies: equality of opportunity to benefit, across the full diversity of patient subpopulations, defined by as many characteristics as differentiate them. A subpopulation need not be defined by demographics alone: its distinctive characteristics may span a wide array of interacting factors only complex generative statistical models, given sufficient data, could adequately summarise. Nor need the heterogeneities distinguishing subpopulations necessarily be static: they could vary across both time and place. The problem is analogous to that of estimating heterogeneous treatment effects[7], and is just as difficult to solve.

One approach is to promote equity in innovation through reason-based interventions, such as distributing research activity closer to the populations it serves[8]. But any intervention here is in as much need of empirical evidence as the novel treatments research aims to discover. Evidential grounding presupposes an understanding and quantification, on the one hand, of the heterogeneities innovation needs to address, and on the other, of the disparities in research participation that could maintain them. Methods quantifying heterogeneities and disparities are widely used in clinical medicine, epidemiology and sociology, but have not yet been formulated as a principled framework that could guide the promotion of equity in innovation. In its absence, interventions are not only speculative, their impact, whether positive or negative, remains speculative too. And however plausible the wisdom of a rationally guided intervention, the extreme complexity of innovation makes unevidenced, purely reason-based inferences here as unsafe as they are in the realm of biology itself.

Note the problem extends much further than the co-localisation of research recruitment with disease prevalence[9]. Equity is dependent on the inclusivity of diverse patterns of disease and associated factors in the sampled population, not on the density of any one pattern. Studying a relatively homogeneous population with a high prevalence of a given disorder will yield a less equitable (and also less generalisable) model than a heterogeneous population with a lower prevalence. If *equity* is the objective, it is *diversity* we need to capture, described comprehensively yet succinctly enough to render variations in care readily actionable.

Theoretically the ideal approach is to reveal the deep structure of our target population with highly expressive generative models of all meaningful patient characteristics, projecting heterogeneity onto a compact *latent space* where its patterns can be discerned and comparative outcomes determined[10]. This immensely challenging task requires sustained engagement with target populations. It also requires innovation, in data acquisition as well as modelling, that will take time to deliver. But people are naturally clustered in geographical, *real-space* communities distributed at multiple spatial scales, whose varying healthcare needs and outcomes geography can thereby render intelligible.

Such geographically informed analysis is readily implementable, and can be used to evaluate and encourage equity along the entire chain of clinical innovation. We can further stratify geographically-organised subpopulations by indices, such as deprivation[11], reasonably presumed to exert a tangible impact on equity. An important aspect of patient heterogeneity can thereby be examined and related to innovation.

Here we report the relationship between participation in primary research studies at University College London Hospital NHS Trust and a key set of geographically organised heterogeneous indices of deprivation, normalised by population and local disease prevalence. The objective is to quantify the association between geographically distributed deprivation and enrolment in research studies at our organisation, a large London NHS Trust, illustrating in the process a quantitative approach to a key aspect of the ethics of innovation.

Methods

Data

For each primary research study registered on UCLH's electronic healthcare records system, we extracted the postcodes of all participants residing in England enrolled in studies initiated between 10/05/2001 and 24/08/2021. The date of individual enrolment was not available. Each postcode was associated with a set of deprivation indices derived from the English Indices of Deprivation 2019 document published by the UK government[11]. The set included Barriers to Housing and Services; Crime; Education, Skills and Training; Employment Deprivation; Health Deprivation and Disability; Income Deprivation; Income Deprivation Affecting Children Index (IDACI); Income Deprivation Affecting Older People Index (IDAOPI); Living Environment Deprivation; and a global summary Index of Multiple Deprivation (IMD). The NHS Quality Outcomes Framework[12] provided locations of GP practices in England with their prevalence rates for common diseases clustered into seven groups: cardiovascular, respiratory, lifestyle, high dependency, mental health and neurology, musculoskeletal, and fertility, obstetrics and gynaecology.

Pre-processing

Participant postcodes in England were aggregated into Lower Layer Super Output Areas (LSOA) defined by the 2011 Census in order to yield partitions of roughly comparable size: ~1000–3000 residents per area. The 5301 LSOAs inside the M25 were identified as Greater London; those 32844 inside its national boundaries as England. In total, 43160 enrolments were located in 4401 LSOAs in Greater London and 54593 participants in 10150 LSOAs in England. Deprivation data was identically aggregated to enable matched comparison. Participant counts for each LSOA were associated with the deprivation data published for each respective area.

The disease prevalence rate for each LSOA was derived by selecting the GP practice participating in the Quality Outcomes Framework closest to the area's geographical centre and taking the geometric mean of all non-zero counts within the 7 disease groups in the dataset. This resulted in 5562 LSOAs with disease prevalence information in the first instance, as not every LSOA contains a GP practice. In a second step, we assigned to each of the remaining LSOAs the disease prevalence rate of its nearest geographical neighbour, as defined by the Euclidean distance between their centroids.

Participant enrolment counts within each LSOA were normalised in two ways: population-adjusted and need-adjusted. Population-adjusted counts per 10000 residents were derived by dividing the participant count of an area by its population and multiplying by 10000. Need-adjusted counts per 10000 residents were produced by dividing the participant count by the population-adjusted number of cases (disease prevalence multiplied by population) and multiplying by 10000.

To account for the natural decrement in participation beyond Greater London, a distance-matched subsample of non-enrolled LSOAs was created as follows. For each of the 10150 LSOAs with at least one participant we matched geometrically the closest LSOA with no participants, yielding a distance-matched sample of 10150 non-enrolled LSOAs. All England-level distribution comparisons used this sample. Where tied distances were encountered, one of the ties was uniformly randomly chosen.

To enable comparison of differences across deprivation indices varying in their scale, each index was normalised within its 0.005^{th} to 0.995^{th} cumulative distribution centiles across all sampled areas (both enrolled and non-enrolled) to the range 0 to 100.

All pre-processing was done with custom Python scripts, using the pandas and geopandas libraries. Postcodes were matched with LSOAs based on the ONS Postcode Directory.

Statistical analysis

Deprivation distributions for areas with and without enrolment were visualised with non-parametric kernel density estimates computed by akde1d. For each index of deprivation, areas with and without enrolments were formally compared with a two-sample, two-sided Mann–Whitney U test. The effect size was calculated as a difference in medians of the data, as long as the visualised histograms were of similar shape. The p-values for an alpha=0.05 were computed asymptotically, correcting for ties, and subsequently Bonferroni corrected for multiple comparisons. The correction factor was 10, treating the 10 deprivation distributions for each of the two regions separately. Non-parametric tests were favoured owing to the evident non-normality of the data. The tests were run using the scipy.stats.mannwhitneyu method implemented in SciPy.

To evaluate the linear relation between volume of enrolment and deprivation indices, population-adjusted or needs-adjusted enrolment counts for each LSOA were entered as a dependent variable into a Bayesian Poisson regression model, with a univariate deprivation index as predictor: separate models were constructed for each deprivation index. All models employed a ridge prior. Using BayesReg version 1.9.1 and its Markov chain Monte Carlo (MCMC) Gibbs sampling procedure, we estimated posterior parameters from every 5th sample (thinning) from one chain holding 3000000 samples, after discarding the first 1000000 samples (burn-in). Posterior distributions were summarised by a mean and 95% credibility intervals. Separate sets of models were run for raw and need-adjusted enrolment data, and Greater London and England. Effective sample sizes of posterior parameter estimates ranged from 24.6 to 98.8% for the population-adjusted models and from 18.6 to 99.2% for the need-adjusted models. All diagrams were rendered using custom Python scripts employing the Altair visualisation library.

Geographical maps

All maps were produced using geometry published by the ONS and assembled using QGIS. The labels for towns and cities were extracted from Open Street Map data (© OpenStreetMap contributors).

Ethics

This local service evaluation does not require ethical approval.

Results

Geographical distribution

A geographical map of LSOAs with at least one enrolled patient revealed wide coverage across England, spanning 31% of its LSOA-parcellated territory (Figure 1). Across Greater London, 83% of areas were enrolled, with density highest in the North-eastern sector that UCLH primarily serves (Figure 2). Maps of the top three disciplines with the highest enrolled number of participants, haematology and oncology (35% of total), neurology (21%), and infectious diseases (15%), show domain-specific variations in sampling (Figure 3).

Distributional comparisons

We compared the distributions of deprivation indices of LSOAs that included at least one enrolled patient vs none, quantifying the magnitude and significance of any difference for each index, and reporting its direction based on the difference in medians. Given the wide but unequally sampled geographical catchment area of UCLH, comparisons were made separately for regions within London and within England.

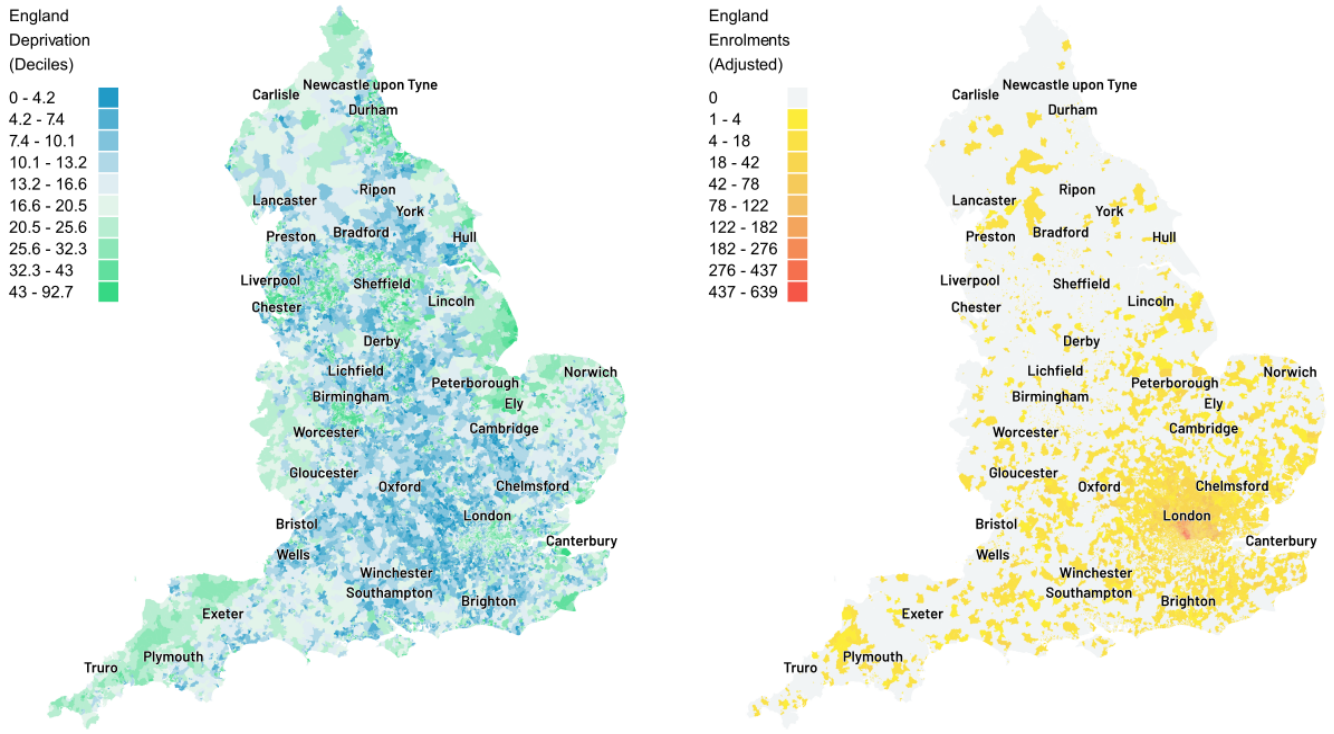

Figure 1. Enrolment and deprivation across England. Map of the Index of Multiple Deprivation (left, deciles), and population-adjusted research study enrolments at UCLH per 10000 residents of England, partitioned by LSOAs (right).

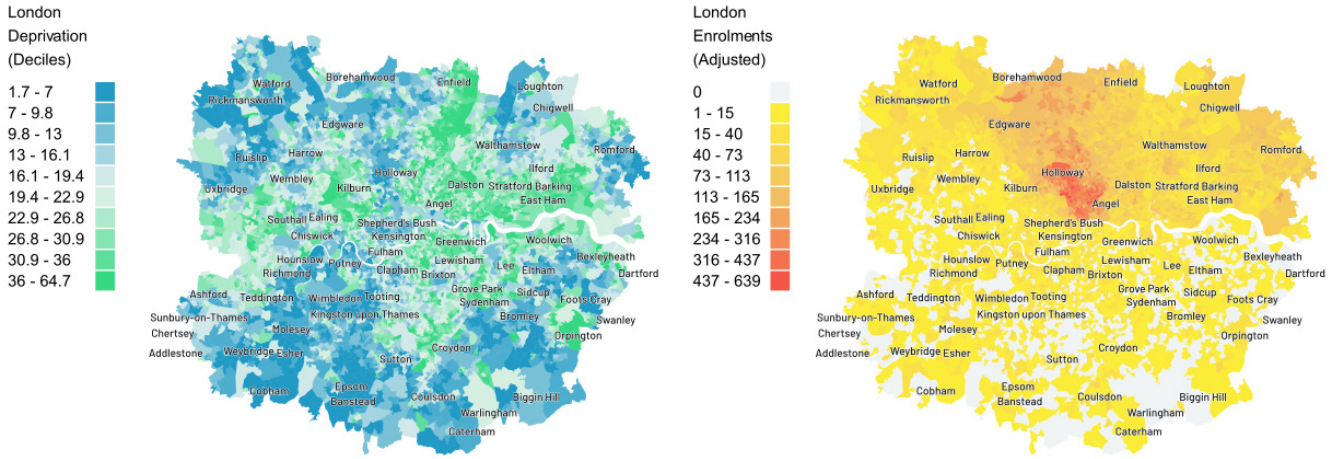

Figure 2. Enrolment and deprivation across Greater London. Map of the Index of Multiple Deprivation (left, deciles), and population-adjusted research study enrolments at UCLH per 10000 residents of Greater London, partitioned by LSOAs (right).

For all LSOAs within Greater London (Table 1 and Figure 4), the distributions were very similar on visual inspection; statistically indistinguishable in Employment Deprivation, Health Deprivation and Disability, and Income Deprivation Affecting Children Index; and significantly different in all others. For Education, Skills and Training, enrolled areas exhibited 4.62% lower median deprivation. For all others, including the aggregate Index of Multiple Deprivation, enrolled areas exhibited greater median deprivation, ranging from 0.92% to 7.94%.

For LSOAs within England (including London) (Table 2 and Figure 5), the distributions were very similar on visual inspection;

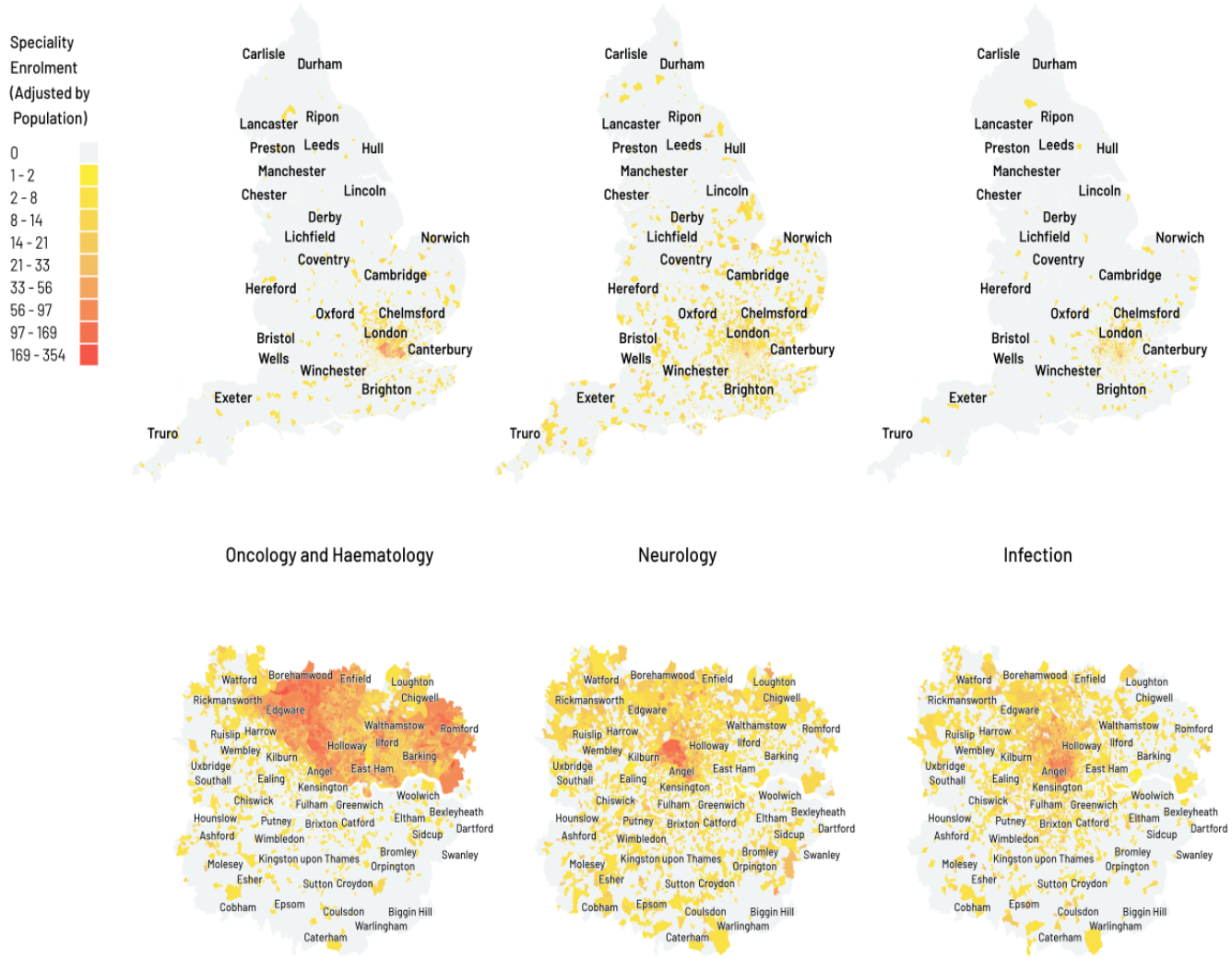

Figure 3. Enrolment by top research domains. UCLH population-adjusted enrolments per 10000 residents of England (top) and Greater London (bottom), for numerically the top three research domains: oncology and haematology (left), neurology (middle), and infectious disease (right).

statistically indistinguishable for Income Deprivation Affecting Children Index; and significantly different for all others. Index of Multiple Deprivation (IMD); Employment Deprivation; Education, Skills and Training; and Health Deprivation and Disability median scores were 0.64% to 6.85% lower in enrolled areas, and all others were 0.26% to 8.73% higher.

Regression analyses

The preceding analyses compare the distributions of areas without enrolment *versus* those with at least one. To measure the association between the volume of enrolment within areas with at least one participant, we conducted a series of linear regressions. Bayesian Poisson regression was used to predict the number of enrolments from individual indices of deprivation, taken separately. To adjust for differences in population density and estimated healthcare need, we normalised enrolment counts, across separate sets of models, by local population density (population adjusted) and disease prevalence (need adjusted). In keeping with the distributional analyses, the regression coefficients for each index revealed a weak association between increased participation and greater deprivation for all indices, including the Multiple Deprivation Index, except for Barriers to Housing Services (within Greater London), and Education, Skills and Training (within Greater London and England), and Health Deprivation and Disability (within England). Population-adjusted and need-adjusted models yielded a broadly similar picture (Figure 6).

Table 1. Results of Mann–Whitney U test comparing the distribution of indices of deprivation or enrolled (n=900) *versus* non-enrolled (n=4401) LSOAs across Greater London.

Deprivation Index	Enrolled Median	Non-enrolled Median	Difference	Normalised difference	U Statistic	p-value
Index of Multiple Deprivation (IMD)	19.665	17.979	1.69	3.64%	2124955.5	0.001
Income Deprivation	0.121	0.111	0.01	3.11%	2105102.5	0.003
Employment Deprivation	0.076	0.074	0.00	0.92%	2018638.0	0.361
Education, Skills and Training	10.605	12.710	-2.11	-4.62%	1775215.5	<0.001
Health Deprivation and Disability	-0.344	-0.398	0.05	1.47%	1985245.0	0.909
Crime	0.267	0.181	0.09	2.94%	2149180.5	<0.001
Barriers to Housing and Services	30.205	29.092	1.11	2.32%	2189675.5	<0.001
Living Environment Deprivation	28.042	23.630	4.41	7.46%	2358108.5	<0.001
Income Deprivation Affecting Children Index (IDACI)	0.156	0.142	0.01	3.35%	2088444.0	0.010
Income Deprivation Affecting Older People Index (IDAOPI)	0.205	0.159	0.05	7.94%	2331494.5	<0.001

Discussion

Examining the relation between enrolment in research studies at UCLH and a comprehensive set of geographically distributed indices of deprivation reveals minor differences generally favouring areas of greater deprivation. Similar patterns are observed within Greater London, and areas of England with matched proximity to UCLH, across distributional comparisons of binary participation, and Poisson regression models of participation count, with and without adjustment for local disease prevalence. For some indices, most consistently Education, Skills and Training, lower enrolment is associated with higher deprivation.

Our analysis demonstrates that research recruitment at UCLH is not markedly biased against those living in deprived areas: in general, it favours them, though some indices are consistently lower in enrolled areas. It emphasises the need for detailed, quantitative evidence in determining the equity of clinical research and innovation. A hospital cannot be presumed on the grounds of its profile, history, or location to be more or less equitable: it must be shown through analyses of the kind conducted here, resolved to finely granular indices that enable targeted, specific, measurable action.

Deprivation is only one, even if very important, geographically distributed index of patient heterogeneity plausibly material to the equitable delivery of care. The same approach can—and should—be applied to other indices. If the confluence of a set of characteristics is such as to define a distinctive subpopulation, its relation to healthcare outcomes ought to be measured. This is true regardless of whether the characteristics are individually recognized to be of ethical concern, such as age, sex, and ethnic background: any systematic cause of unequal opportunity must be identified and addressed.

It should be noted that some scores, especially Health Deprivation and Disability, are bound to interact with the clinical service the host hospital provides within a closely overlapping geography. For example, quality of care, both primary and secondary, may be expected to modulate the index of acute morbidity, which carries a weight of 0.256 in the overall score[13]. A well-performing hospital may, indeed should, thus reduce health deprivation within the clinical basin from which a large proportion of its research participants is likely to be drawn. This does not mean that adverse associations with this or any other potentially interactive factor can be neglected, only that they should be carefully interpreted.

Equally, the overall bias towards deprived areas reflects the natural demographics of UCLH's clinical population. In common with other large metropolitan centres, London exhibits substantial variations in deprivation at fine spatial scales that are invisible from a crude, macroscopic standpoint. We should resist the temptation to treat London populations as uniform noise centred on the mean of the entire region: there is striking local heterogeneity here that must be carefully identified and respected.

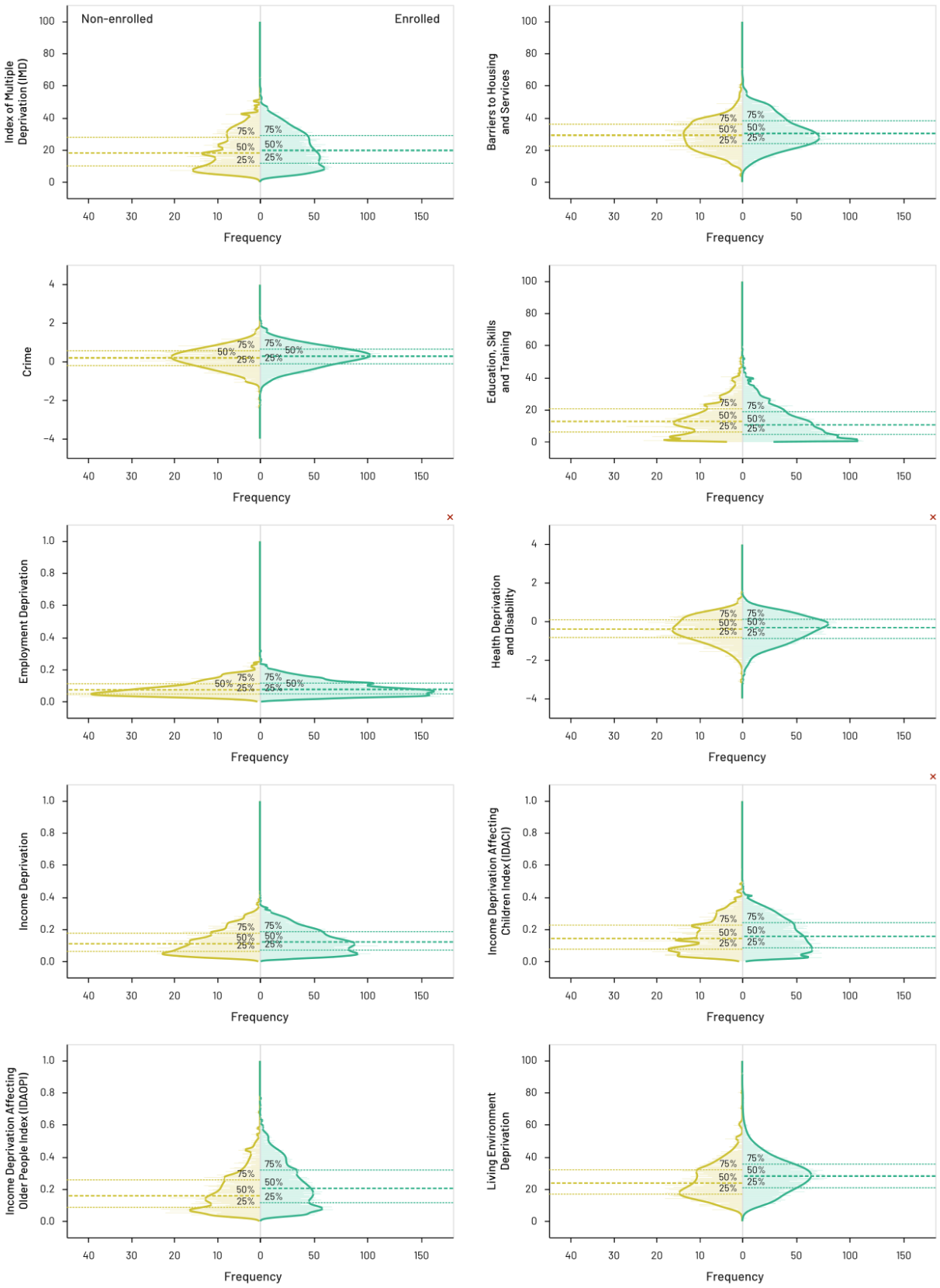

Figure 4. Distributional comparisons across Greater London. Distributions of enrolled (green) and non-enrolled (yellow) LSOAs in Greater London across ten standard indices of deprivation. Each distribution is visualised as a histogram overlaid by an adaptive kernel density estimate with quartiles shown as dashed lines. See Table 1 for formal statistical comparisons.

Table 2. Results of Mann–Whitney U test comparing enrolled *versus* unenrolled distributions in England for deprivation indices.

Deprivation Index	Enrolled Median	Non-enrolled Median	Difference	Normalised difference	U Statistic	p-value
Index of Multiple Deprivation (IMD)	15.596	16.004	-0.41	-0.64%	49605278.5	<0.001
Income Deprivation	0.091	0.090	0.00	0.26%	50267435.5	0.003
Employment Deprivation	0.067	0.073	-0.01	-2.11%	46920821.5	<0.001
Education, Skills and Training	11.755	17.149	-5.39	-6.85%	39620797.0	<0.001
Health Deprivation and Disability	-0.379	-0.155	-0.22	-5.53%	42155368.5	<0.001
Crime	0.017	-0.104	0.12	3.19%	55190264.5	<0.001
Barriers to Housing and Services	25.561	21.292	4.27	8.73%	62126033.5	<0.001
Living Environment Deprivation	21.121	16.313	4.81	7.54%	58213450.0	<0.001
Income Deprivation Affecting Children Index (IDACI)	0.115	0.117	0.00	-0.42%	50375503.5	0.007
Income Deprivation Affecting Older People Index (IDAOPI)	0.127	0.113	0.01	2.35%	55030513.5	<0.001

Our analysis is limited to primary research studies whose participants are individually identifiable. This excludes large-scale secondary studies employing anonymized data, whose lesser selectivity and lower barriers to access are likely to bring the inferred patterns of participation closer to those of the clinical population as a whole. Their exclusion therefore makes our test of equity more, not less, stringent. Nonetheless, the use of complex, machine learning-dependent models acutely sensitive to the range and inclusivity of large-scale data will increasingly make secondary studies important to the comprehensive evaluation of equity.

We rely on several approximations and assumptions. We do not assess the deprivation indices of recruited patients themselves, rather assume that they can be approximated by the mean deprivation of the region they live in. Since participant data were collected over several years and deprivation scores based on 2019 data, our comparisons may be confounded by longitudinal change. Total disease burden was calculated using an averaged prevalence of seven disease categories which may not represent the true disease burden of the research participants involved. Our Poisson regression analyses only modelled linear relations (with a log link function) and did not allow for over-dispersion, rendering our coefficient estimates potentially over-confident or biased. The scale and inclusivity of the modelled data should nonetheless minimize the impact of such effects.

Metrics of equity are themselves diverse, each throwing unwarranted variation into slightly different relief. Rather than focusing, as we have done here, on the relation between an outcome and a factor reasoned negatively to modulate it, a more general approach would be to quantify the variability in an outcome across a set of subpopulations of interest. The critical move remains fractionating the population into locally homogeneous partitions over which comparable recruitment, amongst other measures of engagement, should be observed.

Finally, though our concern here is with *post hoc* evaluation, there is no reason not to perform analyses of this kind either in advance of enrolment, extending the technique of stratified sampling, or adaptively during the process of enrolment itself, again in extension of familiar research practices. The provision of a robust framework for implementing such pre-emptive action is critical to its wider use given the already inhibitory complexities of clinical research studies, especially those of the interventional kind.

Conclusion

Enrolment to UCLH primary research studies is broadly equitably distributed across deprivation, weakly favouring areas with higher multiple and most domain-specific indices, both with and without adjustment for estimated disease prevalence. We propose a framework for continuous evaluation and optimisation of research recruitment responsive to the diversity of the served population.

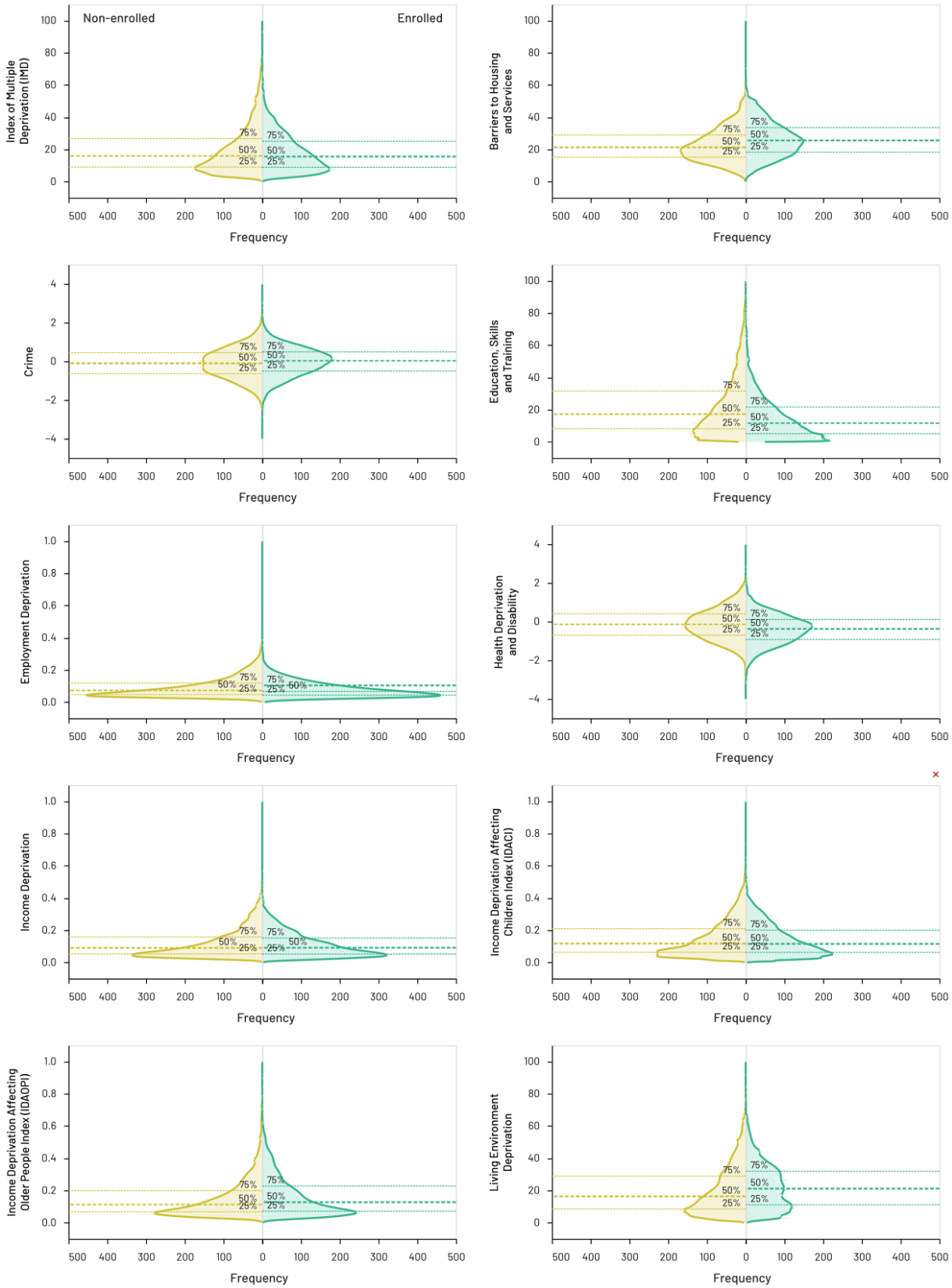

Figure 5. Distributional comparisons across England. Distributions of enrolled (green) and distance-matched non-enrolled (yellow) LSOAs in England across ten standard indices of deprivation. Each distribution is visualised as a histogram overlaid by an adaptive kernel density estimate with quartiles shown as dashed lines. See Table 2 for formal statistical comparisons.

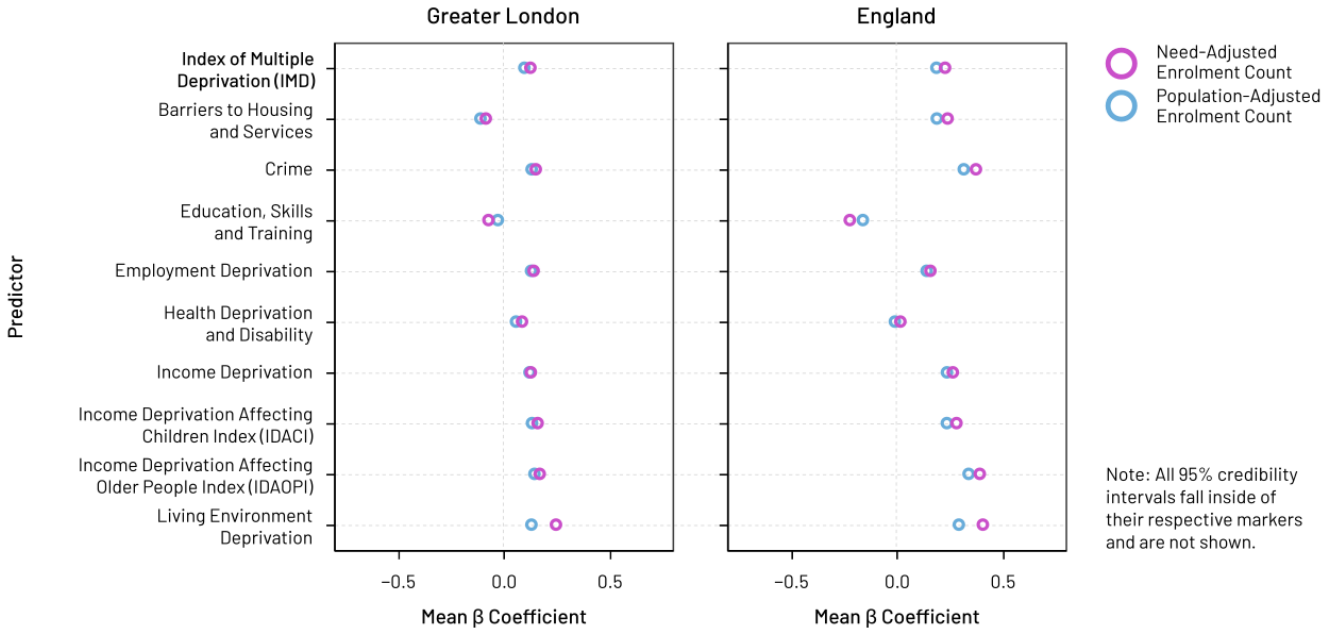

Figure 6. Regression analyses. Plots of the β coefficients derived from Bayesian Poisson regression models for Greater London (left) and England (right), for need-adjusted (magenta) and population-adjusted (blue) enrolment counts. Note all 95% credibility intervals fall inside each marker and are therefore not shown. Positive beta coefficients indicate a positive association with deprivation, negative the converse.

Data availability

Data are potentially identifying and cannot be made public. The aggregate maps and code in this study are available for reviewers and readers from the corresponding author on request by email for non-commercial purposes. The raw data were analysed within an internal service evaluation project; the UCLH's Information Governance Committee has not permitted external release owing to the risk of patient identification.

References

1. Ruger JP: **Health and social justice.** *Lancet.* 2004; **364**(9439): 1075–1080.
PubMed Abstract | Publisher Full Text | Free Full Text
2. Mehta P: **Promoting equality and diversity in UK biomedical and clinical research.** *Nat Rev Genet.* 2006; **7**(9): 668–668.
PubMed Abstract | Publisher Full Text
3. Fry A, Littlejohns TJ, Sudlow C, *et al.*: **Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population.** *Am J Epidemiol.* 2017; **186**(9): 1026–1034.
PubMed Abstract | Publisher Full Text | Free Full Text
4. McDonald AM, Knight RC, Campbell MK, *et al.*: **What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies.** *Trials.* 2006; **7**: 9.
PubMed Abstract | Publisher Full Text | Free Full Text
5. Mehrabi N, Morstatter F, Saxena N, *et al.*: **A survey on bias and fairness in machine learning.** *ACM Comput Surv CSUR.* 2021; **54**: 1–35.
Publisher Full Text
6. Crisp R: **Aristotle: Nicomachean Ethics.** *Cambridge Core.*
Publisher Full Text
7. Alaa A, Schaar M: **Limits of Estimating Heterogeneous Treatment Effects: Guidelines for Practical Algorithm Design.** *Int Conf Mach Learn* 2018; 129–138. (PMLR, 2018).
Reference Source
8. Rai T, Dixon S, Ziebland S: **Shifting research culture to address the mismatch between where trials recruit and where populations with the most disease live: a qualitative study.** *BMC Med Res Methodol.* 2021; **21**(1): 80.
PubMed Abstract | Publisher Full Text | Free Full Text
9. Bower P, Grigoroglou C, Anselmi L, *et al.*: **Is health research undertaken where the burden of disease is greatest? Observational study of geographical inequalities in recruitment to research in England 2013-2018.** *BMC Med.* 2020; **18**(1): 133.
PubMed Abstract | Publisher Full Text | Free Full Text
10. Bengio Y, Courville A, Vincent P: **Representation learning: A review and new perspectives.** *IEEE Trans Pattern Anal Mach Intell.* 2013; **35**(8): 1798–1828.
PubMed Abstract | Publisher Full Text
11. **English indices of deprivation 2019.** *GOV.UK.*
Reference Source
12. **Quality Outcomes Framework (QOF).** *NHS Digital.*
Reference Source
13. McLennan D, Noble S, Noble M, *et al.*: **The English indices of deprivation 2019: Technical report.** 2019.
Reference Source