A data-driven approach to promote diversity of recruitment in a cancer screening trial

Benjamin Ritchie, Victoria Vickerstaff, Adam R. Brentnall, Francesca Rawlins, Nikhil Mayor, Varsha Sonigra-Patel, Emma Cullen, Ros Eeles, Mark Emberton, Rakesh Heer, Caroline Moore, Hashim Ahmed and Rhian Gabe on behalf of the TRANSFORM trial group

Background

Cancer screening is becoming more personalised to increase test accuracy and reduce overdiagnosis.[1] Large-scale clinical trials continue to play a crucial role in evaluating population screening strategies. The data and biological samples collected in trials are powerful resources for further personalised medicine research.[2] However, the applicability of results to minoritised ethnic groups is limited by their under-representation in trial cohorts.[3, 4]

TRANSFORM is a large-scale randomised trial comparing multiple screening options for prostate cancer.[5] Eligible men are identified by searching the patient registers of general practices (GPs). Since Black men have a higher risk of prostate cancer and are less likely to participate in medical research,[4] TRANSFORM aims to over-represent individuals identifying as Black among those invited. The study has a similar aim for inviting men living in the most deprived areas, specifically those in the lowest quintile of the Index of Multiple Deprivation (IMD).

Methods

A data-driven approach was taken to identify GPs with patient profiles and geographical locations that meet TRANSFORM's aims. This method combined publicly available GP and Lower Super Output Area (LSOA) datasets to determine numbers of registered patients by sex, age, ethnic group and IMD quintile.

For each GP:

- The practice's proportion of patients self-identifying as Black was estimated as a weighted average of the proportions of Black residents in the LSOAs it serves. The weight assigned to each LSOA was the proportion of the GP's patients living in it.
- A similar method approximated the proportion of the practice's patients living in the most deprived quintile of the IMD.

GPs with more than 600 male patients aged 50 or older were categorised in relation to these factors. Those in the proximity of a secondary care site confirmed for participation in TRANSFORM were plotted on maps using the lattice package of the R programming language. Results were cross-referenced against Census 2021 ethnicity maps.

Results

Geographical clusters of GPs with patient profiles in line with the trial's aims were found near the West London and Bath secondary care study sites (Figure 1).

- In West London, there is a cluster of 34 GPs with an estimated proportion of Black patients of more than 10%. The majority (27) of these GPs have more than 20% of patients living in the most deprived quintile.
- For the Bath site, a cluster of 20 GPs with more than 20% of patients living in the most deprived quintile was identified in nearby Bristol. The minority (3) of these GPs have more than 10% Black patients.

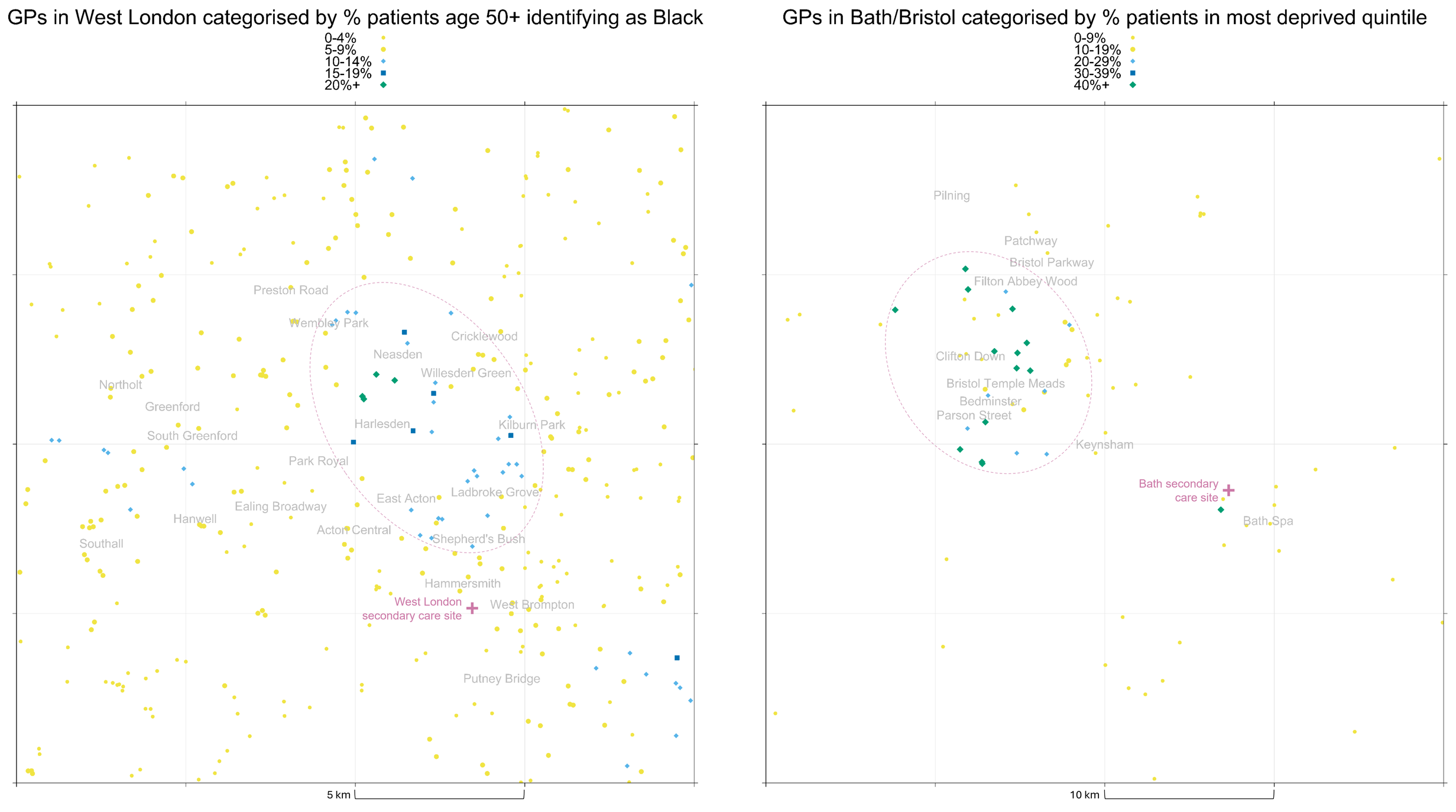

Figure 1. Maps of GPs in West London (left panel) and Bath/Bristol (right panel). Clusters with patient profiles aligned with trial aims are circled with dashed ovals. Nearby train/underground stations are annotated.

Conclusion

In a cancer screening trial recruiting patients from primary care, general practices with patient demographics that would support diversity of recruitment objectives were identified by merging and mapping published datasets. Ongoing work includes engaging with networks of identified GPs to ascertain their capacity to take part in the trial, investigating the feasibility of screening facilities near the GPs and considering travel times for participants.[6] Including a diverse sample of participants in TRANSFORM will increase the applicability of its findings to previously under-represented groups.

Acknowledgements

Funding partners

The Freddie Green and Family Charitable Foundation

Academic collaboration

References

1. K. Chai et al., Front. Oncol. 15 (2025), doi:10.3389/fonc.2025.1511880.
2. A. Esteva et al., npj Digit. Med. 5 (2022), doi:10.1038/s41746-022-00613-w.
3. Z. Iyizoba-Ebozue et al., Clinical Oncology. 34, 674–677 (2022), doi:10.1016/j.clon.2022.07.006.
4. K. H. Kensler et al., JNCI. 116, 34–52 (2023), doi:10.1093/jnci/djad193.
5. T. Burki, The Lancet. 403, 1738 (2024), doi:10.1016/S0140-6736(24)00912-7.
6. M. Berthelsen et al., Cancer Epidemiology. 80, 102244 (2022), doi:10.1016/j.canep.2022.102244.