Cortical patterning of abnormal morphometric similarity in psychosis is associated with brain expression of schizophrenia-related genes

Sarah E Morgan\textsuperscript{a,1}, Jakob Seidlitz\textsuperscript{a,1}, Kirstie J Whitaker\textsuperscript{a,c}, Rafael Romero-García\textsuperscript{a}, Nicholas E Clifton\textsuperscript{d,e}, Cristina Scarpazza\textsuperscript{a}, Therese van Amelsvoort\textsuperscript{a}, Machted Marcelis\textsuperscript{a}, Jim van Os\textsuperscript{f}, Gary Donohoe\textsuperscript{g}, David Mothersill\textsuperscript{h}, Aiden Corvin\textsuperscript{i}, Andrew Pocklington\textsuperscript{j}, Armin Raznahan\textsuperscript{k}, Philip McGuire\textsuperscript{l}, Petra E Vértes\textsuperscript{b,c,k,2}, and Edward T Bullmore\textsuperscript{a,h,2}

\textsuperscript{a}Department of Psychiatry, University of Cambridge, Cambridge CB2 0SZ, UK; \textsuperscript{b}Developmental Neurogenomics Unit, National Institute of Mental Health, Bethesda, MD 20892, USA; \textsuperscript{c}The Alan Turing Institute, London, NW1 2DB, UK; \textsuperscript{d}Neuroscience and Mental Health Research Institute, Cardiff University, Cardiff, UK; \textsuperscript{e}MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK; \textsuperscript{f}Department of Psychology, King's College London, London SE5 8AF, UK; \textsuperscript{g}Department of General Psychology, University of Padova, 35131 Padova, Italy; \textsuperscript{h}Maastricht University, Maastricht, The Netherlands, 6166ZM MD; \textsuperscript{i}School of Psychology, NUI Galway, Galway, Ireland, H91 TK33; \textsuperscript{j}Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, UK; \textsuperscript{k}Institute of Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK; \textsuperscript{l}School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom; \textsuperscript{m}GlaxoSmithKline R&D, Stevenage SG1 2NY, UK.

This manuscript was compiled on March 25, 2019

Schizophrenia has been conceived as a disorder of brain connectivity but it is unclear how this network phenotype is related to the underlying genetics. We used morphometric similarity analysis of magnetic resonance imaging (MRI) data as a marker of inter-areal cortical connectivity in three prior case-control studies of psychosis: in total, N=185 cases and N=227 controls. Psychosis was associated with globally reduced morphometric similarity (MS) in all 3 studies. There was also a replicable pattern of case-control differences in MS which was significantly reduced in patients in frontal and temporal cortical areas, but increased in parietal cortex. Using prior brain-wide gene expression data, we found that the cortical map of case-control differences in MS was spatially correlated with cortical expression of a weighted combination of genes enriched for neurobiologically relevant ontology terms and pathways. In addition, genes that were normally over-expressed in cortical areas with reduced MS were significantly up-regulated in three prior postmortem studies of schizophrenia. We propose that this combined analysis of neuroimaging and transcriptional data provides new insight into how previously implicated genes and proteins, as well as a number of unreported genes in their topological vicinity on the protein interaction network, may drive structural brain network changes mediating the genetic risk of schizophrenia.

Significance Statement

Despite significant research, the biological mechanisms underlying schizophrenia are still unclear. We shed fresh light on structural brain differences in psychosis using a new approach called morphometric similarity mapping, which quantifies the structural similarity between brain regions. Morphometric similarity was globally reduced in psychosis patients in three independent datasets, implying that patients’ brain regions were more differentiated from each other, less interconnected. Similarity was especially decreased in frontal and temporal regions. This anatomical pattern was correlated with expression of genes enriched for nervous system development and synaptic signaling, and genes previously associated with schizophrenia and anti-psychotic treatments. So we begin to see how combining genomics and imaging can give a more integrative understanding of schizophrenia, which might inform future treatments.

E.T.B. is employed half-time by the University of Cambridge and half-time by GlaxoSmithKline; he holds stock in GSK.

P.E.V. and E.T.B. contributed equally to this work.

1To whom correspondence should be addressed. E-mail: sem91@cam.ac.uk

www.pnas.org/cgi/doi/10.1073/pnas.XXXXXXXXXX

PNAS | March 25, 2019 | vol. XXX | no. XX | 1–6
networks and structural covariance networks, MS networks included a greater proportion of connections between human cortical areas of the same cytoarchitectonic class. Individual differences in regional mean MS, or “hubness” of cortical nodes in MS networks, accounted for about 40% of the individual differences in IQ in a sample of 300 healthy young people. These results suggest that MS mapping could provide a useful new tool to analyse psychologically relevant biological differences in brain structure.

Here we used MS mapping to test the disconnection hypothesis of psychosis in three independent case-control MRI datasets: the Maastricht GROUP study (83 cases, 68 controls) and the Dublin study (33 cases and 82 controls), both made available as legacy datasets for the PSYSCAN project, and the publicly available Cobre dataset (69 cases and 77 controls). See Methods. We mapped case-control MS differences at global and nodal levels of resolution individually in each dataset to assess replicability and we tested for significant differences in network organization that were consistent across studies. We used partial least squares (PLS) regression to test the hypothesis that this MRI network phenotype of psychosis was correlated with anatomically patterned gene expression using data from the Allen Human Brain Atlas (AHBA). This analytical approach to combine imaging and genomics data has been methodologically established (5, 6) and applied in the context of neuropsychiatric disorders (7, 8). We used it to test the pathogenic hypothesis that the genes most strongly associated with case-control differences in MS were enriched: (i) for genes that have been ontologically linked to relevant neurobiological processes; and (ii) for genes that are abnormally expressed in post-mortem studies of schizophrenia.

Results

Samples. Socio-demographic and clinical data available on the sample are in Table S1. There was considerable heterogeneity in clinical measures between studies, e.g., the Maastricht patients had relatively low mean scores on psychotic symptoms scales.

Case-control differences in global morphometric similarity. Globally, MS was reduced in cases compared to controls in all 3 datasets (Fig. S2). Regional MS had an approximately Normal distribution over all 308 regions (after regressing age, sex and age × sex) and for a subset of MS differences in all 3 datasets, we found a significant case-control difference in this distribution ($P < 0.001$, Kolmogorov-Smirnov test). Modal values of regional MS were more frequent, and extreme values less frequent, in cases compared to controls (Figs. 1a and S2).

Case-control differences in regional morphometric similarity. The cortical map of regional MS in Fig. 1c summarises the anatomical distribution of areas of positive and negative similarity on average over controls from all 3 datasets. The results are similar to those reported in an independent sample (3), with high positive MS in frontal and temporal cortical areas and high negative MS in occipital, somatosensory and motor cortex. This confirms the replicability of this pattern of regional MS in healthy individuals and is consistent with prior knowledge that primary cortex is more histologically differentiated than association cortex.

We mapped the t-statistics and corresponding Hedge’s $g$ effect sizes for the case-control differences in regional MS at each cortical area (Fig. 1d). A positive t-statistic means MS increased in patients whereas a negative t-statistic means MS decreased. We found somewhat similar patterns of case-control difference across all 3 datasets, with increased regional MS in occipital and parietal areas in patients, and decreased regional MS in frontal and temporal cortex. The case-control t-map for the Dublin study was significantly correlated with both the Maastricht and the Cobre t-maps ($r = 0.42, P < 0.001$ and $r = 0.47, P < 0.001$, respectively), although the Maastricht and Cobre t-maps were not significantly correlated ($r = 0.058, P = 0.31$), see Fig. S4. However, a large number of patients in the Maastricht dataset had very low symptom scores (below the threshold for “borderline mentally ill” (9)). If those non-psychotic patients were excluded from the analysis, the Maastricht case-control t-map was correlated significantly with the Cobre map ($r = 0.22, P < 0.001$, see section S6.2).

Combining the P-values for case-control differences across all 3 datasets, we identified 18 cortical regions where MS was robustly and significantly different between groups (Fig. 1e). MS decreased in patients in 15 regions located in the superior frontal, caudal middle frontal, pre-central, par triangularis and superior temporal areas and increased in 3 regions located in superior parietal and post-central areas (Table S2).

To contextualise the regional MS case-control differences, we referred them to two prior classifications of cortical areas: the von Economo atlas of cortex classified by cytoarchitectonic criteria (5); and the Yeo atlas of cortex classified according to resting state networks derived from functional MRI (10, 11). MS was significantly reduced in von Economo class 2 (association cortex) and in the ventral attention, frontoparietal and default mode Ye networks (all $P_{FDR} < 0.05$; Tables S12 and S13).

There was a strong negative correlation between regional MS in the control subjects and the case-control differences in regional MS (both averaged over all 3 datasets; $P_{perm} = 0.002$) (Fig. 1f). Hence areas with highest positive MS in controls tended to show the greatest decrease of MS in patients; and, conversely, areas with highest negative MS in healthy controls had the greatest increase of MS in psychosis. This result is analogous to the observation that highly connected ‘hub’ regions are the most likely to show reduced connectivity in disease in fMRI and DTI brain networks (12). We tested for correlations between mean MS and a range of clinical measures, including symptom scores, anti-psychotic medication use and cannabis use, see section S6.3. The only significant associations after FDR correction were with cannabis use, which was positively correlated with mean global MS in the Maastricht study ($P_{FDR} = 5 \times 10^{-4}$), as well as with mean MS averaged across the 15 regions with significantly decreased MS in Fig. 1e ($P_{FDR}=0.0017$).

Gene expression related to morphometric similarity. We used PLS regression to identify patterns of gene expression that were correlated with the anatomical distribution of case-control MS differences. The first PLS component explained 13% of the variance in the case-control MS differences, combining data from all 3 studies, significantly more than expected by chance (permutation test, $P < 0.001$). PLS1 gene expression weights were positively correlated with case-control MS differences in the Dublin study ($r = 0.49, P < 0.001$) and the Cobre study ($r = 0.37, P < 0.001$ (Fig. 2a); but not in the Maastricht study ($r = 0.006, P = 0.94$). These positive correlations mean...
Fig. 1. Case-control differences in regional MS. a) Distributions of regional MS strength, i.e. the average similarity of each region to all other regions, for cases and controls from all datasets. b) Distributions of regional MS strength for a region with significantly reduced MS in cases, namely left hemisphere caudal middle frontal part 1. c) Regional MS averaged over controls from all 3 datasets. d) t-statistics and Hedge’s g effect sizes for the case-control differences in regional MS in each dataset. e) t-statistics for regional case-control differences averaged across datasets in all regions and in the 18 cortical areas where the difference was statistically significant across datasets (FDR = 0.05). f) Scatterplot of mean control regional MS (x-axis) versus case-control t-statistic (y-axis): Control MS (from panel c) is strongly negatively correlated with case-control MS differences (from panel d) (Pearson’s r = −0.76, P < 0.001). Most cortical regions have positive MS in controls which decreases in patients (47% of regions) or negative MS in controls which increases in patients (36% of regions). Statistically significant regions are circled in red/blue according to whether their mean t-statistic increases/decreases in patients.

Fig. 2. Gene expression profiles related to case-control differences in morphometric similarity a) Scatterplot of regional PLS1 scores (weighted sum of 20,647 gene expression scores) versus case-control differences in regional MS (Cobre dataset). b) Cortical map of regional PLS1 scores. c) Cortical map of mean case-control MS differences, averaged across all datasets. Here we include intra-hemispheric left hemisphere edges only (see Methods). d) Genes that are strongly positively weighted on PLS1 (e.g., LYSMD4) correlate positively with case-control differences in regional MS (r = 0.44, P < 0.001); whereas genes that are strongly negatively weighted on PLS1 (e.g., C1orf95) correlate negatively with case-control differences in MS (r = −0.37, P < 0.001).
that genes positively weighted on PLS1 are over-expressed in regions where MS was increased in patients, whilst negatively weighted genes are over-expressed in regions where MS was decreased in patients (Fig. 2d). Hence genes which are positively (or negatively) weighted on PLS1 were related to increased (or decreased) MS in cases compared to controls.

**Enrichment analysis of genes transcriptionally related to morphometric similarity.** We found 1110 genes with normalised PLS1 weights $Z < -3$, which we denote the PLS- gene set, and 1979 genes with $Z > 3$, which we denote the PLS+ gene set. We first consider PLS- genes (the equivalent results for PLS+ genes are also given below).

We mapped the network of known interactions between proteins coded by the PLS- gene set (13) (Fig. 3). The resulting protein-protein interaction (PPI) network had 341 connected proteins and 1022 edges, significantly more than the 802 edges expected by chance (permutation test, $P < 1^{-13}$). We also tested the PLS- gene set for significant GO enrichment of biological processes and enrichment of KEGG pathways. Enriched biological processes included “nervous system development”, “synaptic signaling” and “adenylate cyclase-modulating G-protein coupled receptor signaling pathway.” (see Dataset S1). There were two significantly enriched KEGG pathways: “neuroactive ligand-receptor interaction” and “retrograde endocannabinoid signaling” (Fig. S13). The proteins coded by genes enriched for “adenylate cyclase-modulating G-protein coupled receptor signaling pathway” and the two KEGG pathways formed the most strongly inter-connected cluster of nodes in the PPI network, see Fig. 3, compatible with them sharing a specialised functional role for GPCR signaling.

Genes recently reported as up-regulated in post mortem brain tissue from patients with schizophrenia (14) were highly enriched among genes that were negatively weighted on PLS1 (permutation test, $P < 0.001$, after FDR correction). The relationship between the sign of PLS1 weights of gene expression related to the MRI case-control phenotype and the sign of case-control differences in the histological measures of brain gene expression was highly non-random (Wilcoxon rank sum test, $P < 10^{-26}$).

In other words, genes that were up-regulated in post mortem brain tissue from patients with schizophrenia are normally over-expressed in association cortical areas that have reduced MS in psychosis. This association between gene expression in regions with reduced MS and genes up-regulated in schizophrenia was replicated by analysis of two alternative datasets provided by the PsychENCODE consortium (15) and by (16), see section S8.5. We also observed enrichment by genes up-regulated in other psychiatric disorders, e.g., autistic spectrum disorders, which is compatible with the substantial overlap between genes which are up (or down) regulated in common between schizophrenia and other neurodevelopmental disorders (15).

The PLS+ genes coded proteins that formed a PPI network with significantly more edges than expected by chance ($P < 10^{-5}$), which was enriched for the biological process “nucleic acid metabolic process” but no KEGG pathways, see Fig. S14. Genes which are down-regulated post mortem in schizophrenia (14) were highly enriched among genes that were positively weighted on PLS1 (permutation test, $P < 0.001$ after FDR correction). This result was reproduced with genes reported as down-regulated in schizophrenia by (16), although not by the PsychENCODE consortium (15), see section S8.5.

There was no significant enrichment of PLS- or PLS+ genes for common sequence variants associated with schizophrenia, derived from a recent genome-wide association study (GWAS) of PGC and CLOZUK samples (17) ($P > 0.05$).

**Discussion.**

**Morphometric similarity network phenotypes.** Morphometric similarity mapping disclosed a robust and replicable cortical pattern of differences in psychosis patients. MS was significantly reduced in frontal and temporal cortical areas, and significantly increased in parietal cortical areas. This pattern was consistent across 3 independent datasets, with different samples, locations, scanners and scanning parameters.

What does this novel MRI phenotype of psychosis represent? Morphometric similarity mapping quantifies the correspondence or kinship of two cortical areas in terms of multiple macro-structural features, e.g., cortical thickness, and micro-structural features, e.g., fractional anisotropy (FA), that are measurable by MRI. We assume that high MS between a pair of cortical regions indicates that there is a high degree of correspondence between them in terms of cytoarchitectonic and myeloarchitectonic features that we cannot directly observe, given the limited spatial resolution and cellular specificity of MRI. Prior work also showed that morphometrically similar cortical regions are
more likely to be axonally connected to each other, i.e., MS is a proxy marker for anatomical connectivity (3). We therefore interpret the reduced MS we observe in frontal and temporal brain regions in psychosis as indicating that there is reduced architectonic similarity, or greater architectonic differentiation, between these areas and the rest of the cortex, which is probably indicative of reduced anatomical connectivity to and from the less similar, more differentiated cortical areas.

There is a well-evidenced and articulated prior theory of schizophrenia as a dysconnectivity syndrome, specifically functional dysconnectivity of frontal and temporal cortical areas has been recognised as a marker of brain network disorganization in schizophrenia (18). Our results of reduced MS in frontal and temporal cortex - implying increased architectonic differentiation and decreased axonal connectivity - are descriptively consistent with this theory. Our complementary finding of abnormally increased MS in parietal cortex - implying increased architectonic similarity and axonal connectivity - is plausible but not so clearly precedent, given the relatively limited prior data on the parietal cortex in studies of schizophrenia as a dysconnectivity syndrome (19, 20).

Encouragingly, this novel MRI network marker of psychosis was highly reliable across three independent and methodologically various case-control studies. This implies that the measurement is robust enough to be plausible as a candidate imaging biomarker of cortical network organization in large-scale, multi-centre studies of psychosis.

Transcriptional profiling of MS network phenotypes. In an effort to connect these novel MRI phenotypes to the emerging genetics and functional genomics of schizophrenia, we first used partial least squares to identify the weighted combination of genes in the whole genome that has a cortical expression map most similar to the cortical map of case-control MS differences. Then we tested the mechanistic hypothesis that the genes with greatest (positive or negative) weight on PLS1 were enriched for genes previously implicated in the pathogenesis of schizophrenia.

We found that the genes that are normally over-expressed in frontal and temporal areas of reduced MS in psychosis, were significantly enriched for genes that are up-regulated in post mortem brain tissue from patients with schizophrenia (14). Conversely, the genes that are normally over-expressed in parietal and other areas of increased MS in psychosis were significantly enriched for genes that are down-regulated in post-mortem data (14). This tight coupling between MRI-derived transcriptional weights and gene transcription measured histologically was highly significant and replicated across three prior post-mortem datasets.

Further investigation showed that the proteins coded by the PLS- genes formed a dense, topologically clustered interaction network that was significantly enriched for a number of relevant GO biological processes and KEGG pathways. The cluster of interactive proteins related to GPCR signaling included multiple proteins coded by genes previously linked to anti-psychotic mechanisms of action, including DRD4 (21), HTR1 (22), NTSR1 (25) and ADRA2C (24); reported in transcriptional studies of post-mortem brain tissue, e.g., PTGER3, S1PR1, ITPR2 and EDNRB (14, 25); or associated with risk SNPs for schizophrenia, e.g., DRD5, OPRM1 and CNR1 (26–28). The remarkable density of therapeutically relevant genes in the GPCR-related cluster suggests that other, topologically neighboring genes may deserve further attention as novel targets for anti-psychotic interventions.

Risk genes identified by the largest extant GWAS studies of schizophrenia were not significantly enriched among PLS- or PLS+ genes. Nevertheless, the involvement of PLS- genes further down the causal pathway is still mechanistically revealing and potentially useful.

Methodological considerations. Some limitations of this study should be highlighted. The whole brain data on “normal” brain tissue expression of the genome were measured post mortem in 6 adult brains (mean age = 43 years) and not in age-matched subjects or patients with schizophrenia (such data are not currently available to our knowledge). Also, the transcriptional experiments we use to label genes as up- or down-regulated in schizophrenia were performed in regions of the parietal or prefrontal cortex (14), whereas the neuroimaging results are for the whole brain. We have used MRI data from 3 independent studies to measure MS networks but the studies used different scanning protocols, leading to estimation of morphometric similarity between regions on the basis of 7 MRI parameters that were measurable in all studies. Future work could usefully explore the opportunity to further improve sensitivity and reliability of the MS network biomarker of schizophrenia by optimising and standardising the MRI procedures to measure the most informative set of morphometric features. Finally, the datasets have varied, limited clinical information available, making it difficult to assess the clinical significance of the MS phenotype.

Samples. We used MRI data from 3 prior case-control studies: the Maastricht GROUP study (29) from the Netherlands; the Dublin database which was acquired and scanned at the Trinity College Institute of Neuroscience as part of a Science Foundation Ireland-funded neuroimaging genetics study (“A structural and functional MRI investigation of genetics, cognition and emotion in schizophrenia”); and the publicly available Cobre dataset (30). The Maastricht and Dublin datasets were PSYSCAN legacy datasets. All patients satisfied DSM-IV diagnostic criteria for schizophrenia or other non-affective psychotic disorders. MRI data were quality controlled for motion artifacts (section S1). The Euler number, which quantifies image quality (31), was not significantly different between groups in any of the studies but it was different between studies, indicating that the studies were ranked Dublin > Cobre > Maastricht in terms of image quality (Table S1).

Morphometric similarity mapping. The T1-weighted MRI data (MPRAGE sequence) and the diffusion-weighted imaging (DWI) data from all participants were pre-processed using a previously defined computational pipeline (5). Briefly, we used the recon-all (32) and trac-all (33) commands from FreeSurfer (version 6.0). Following (3), the surfaces were then parcellated using an atlas with the publicly available Cobre dataset (30). The Maastricht and Dublin datasets were PSYSCAN legacy datasets. All patients satisfied DSM-IV diagnostic criteria for schizophrenia or other non-affective psychotic disorders. MRI data were quality controlled for motion artifacts (section S1). The Euler number, which quantifies image quality (31), was not significantly different between groups in any of the studies but it was different between studies, indicating that the studies were ranked Dublin > Cobre > Maastricht in terms of image quality (Table S1).

Case-control analysis of MS networks. The global mean MS for each participant is the average of $M_i$. The regional mean $M_{ij}$, for the jth participant at each region, $j = 1, \ldots, 308$, is the average of the
Transcriptomic analysis. We used the AHBA transcriptomic dataset, with gene expression measurements in 6 post-mortem adult brains (35) (http://human.brain-map.org), aged 24-57. Each tissue sample was assigned to an anatomical structure using the AHBA MRI data for each donor (36). Samples were pooled between bilaterally homologous cortical areas. Regional expression levels for each gene were compiled to form a 308 × 20,647 regional transcription matrix (36). Since the AHBA only includes data for the right hemisphere for two subjects, in our analyses relating gene expression to MRI data we only consider intra-hemispheric left hemisphere edges (37).

We used PLS to relate the regional MS case-control differences (t-statistics from the 152 cortical regions in the left hemisphere, calculated from intra-hemispheric edges only) to the post mortem gene expression measurements for all 20,647 genes. PLS uses the gene expression measurements (the predictor variables) to predict the intra-left hemisphere regional MS patient/controls t-statistics (t-control variables). The first PLS component (PLS1) is the linear combination of the weighted gene expression scores that has a cortical expression map that is most strongly correlated with the map of case-control MS differences. The statistical significance of the variance explained by PLS1 was tested by permuting the response variables 1,000 times. The error in estimating each gene’s PLS1 weight was assessed by bootstrapping (resampling with replacement of the 308 cortical regions), and the ratio of the weight of each gene to its bootstrap standard error was used to calculate the Z-scores and hence rank the genes according to their contribution to PLS1 (5).

We constructed FPI networks from the genes with PLS1 weights Z > 3 and Z < −3 (all P < 0.05) using STRING version 10.5 (13). Our key results were robust to changing these thresholds to Z > 4 and Z < −4 (all P < 0.01), see section S8.3. We used DAVID (38, 39) to calculate enrichments of KEGG pathways and GO enrichments of biological processes for genes with Z > 3 or Z < −3, using a background gene list of 15,745 brain-expressed genes, again see section S8.3 (37).

We adopted a similar procedure to test for enrichment of PLS-derived gene sets by genes previously associated with schizophrenia by transcriptional data (14). The median rank of each risk gene set in the PLS gene list was compared to the median rank of 10,000 randomly selected brain-expressed gene sets (3).

Data and code availability. All code and processed data used for the analyses will be made available on GitHub on publication.

ACKNOWLEDGMENTS. This study was supported by grants from the European Commission (PSYSCAN – Translating neuropsychiatric findings from research into clinical practice; ID: 603196) and the NIHR Cambridge Biomedical Research Centre (Mental Health). The Cobre data was downloaded from the COLlarative Informatics and Neuroimaging Suite Data Exchange tool (COINS; http://coins.mrn.org/dx/) and data collection was performed at the Mind Research Network, and funded by the Center of Biomedical Research Excellence (COBRE) grant 5P20RR021936/2P20GM103472 from the NIH to Dr. Vince Calhoun. SEM holds a Henslow Fellowship at Lucy Cavendish College, University of Cambridge, funded by the Cambridge Philosophical Society. PEV was supported by the Medical Research Council (MR/K020706/1) and an MQ fellowship (MQF17_24) and is a Fellow of the Alan Turing Institute funded under the EPSRC grant EP/N501012/1. KJIW was funded by an Alan Turing Institute Research Fellowship under EPSRC Research grant TUA/A/000017. ETB is supported by a NIHR Senior Investigator Award. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.