Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Clusters of people with multiple long-term conditions across socioeconomic and ethnic groups and their association with unplanned hospitalisations: A population-based cohort study from England

Maria T Sanchez-Santos
University of Oxford
Xihang Chen
University of Oxford
Leena Elhussein
University of Oxford
Marta Pineda-Moncusi
University of Oxford
Gianluca Fabiano
University of Oxford
Antonella Delmestri
University of Oxford
Rafael Pinedo-Villanueva
University of Oxford
Lindsay Cherry
University of Southampton
Daniel Prieto Alhambra
daniel.prietoalhambra@ndorms.ox.ac.uk
University of Oxford
Mari Carmen Portillo
University of Southampton
Dorit Kunkel
University of Southampton

Research Article

Keywords: (Three to ten keywords representing the main content of the article)

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11469894/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. RPV has received consultancy fees from Kyowa Kirin Services and UCB Biopharma unrelated to this study. DPA is Principal Investigator for Real World Evidence studies sponsored by UCB Biopharma and Theramex, and a co-investigator for a study sponsored by Eli Lilly, and previous co-investigator for a study sponsored by Novartis. He has been a co-investigator in IMI-funded projects (EHDEN, EU-PEARL) and is currently involved in OPTIMA ONCOLOGY, while leading the IHI-funded GREG (ihi-greg.eu) programme. His research group has received public funding from NIHR, and the European Medicines Agency. He is also a member of the ENCePP Steering Group. LE is a full-time employee at Novartis Pharmaceuticals UK Ltd, however, her contribution to this study was made solely during her prior employment at the University of Oxford and is fully independent of her current role. MTSS, XC, MPM, GF, AD, LC, MCP and DK have no conflict of interest.

Clusters of people with multiple long-term conditions across socioeconomic and ethnic groups and their association with unplanned hospitalisations: A population-based cohort study from England

Maria T Sanchez-Santos[1], Xihang Chen[1], Leena Elhussein[1], Marta Pineda-Moncusí[1], Gianluca Fabiano[1], Antonella Delmestri[1], Rafael Pinedo-Villanueva[1], Lindsey Cherry[2,3], Daniel Prieto-Alhambra[1†], Mari Camen Portillo[2,3**], Dorit Kunkel[2,3**]

Authors' affiliations:

[1]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, United Kingdom

[2]School of Health Sciences, University of Southampton, Southampton, United Kingdom

[3]NIHR Applied Research Collaboration Wessex

[†]Corresponding author: Prof Daniel Prieto-Alhambra, MD MSc PhD

daniel.prietoalhambra@ndorms.ox.ac.uk

Botnar Research Centre, OX3 7LD, Oxford, United Kingdom

**Joint-senior authors

Main text: 4,304 words

Tables: 4

Figures: 3

References: 47

Supplementary files: 1

Abstract

Background:

Multiple long-term conditions (MLTC) are highly prevalent and lead to poor health and increased healthcare utilisation. Evidence is needed to identify clusters of people with MLTC across socioeconomic and ethnic groups and examine their association with unplanned hospitalisation, to better understand differences in MLTC patterns and healthcare needs.

Methods:

We used data from Clinical Practice Research Datalink (CPRD) Gold linked to Hospital Episode Statistics (HES). We included patients aged 18+ years on 01/01/2018, with ≥2 long-term conditions (LTC), and with ≥1 year of data visibility. Conditions of interest were identified based on clinical knowledge and prevalence ≥5%. We applied latent class analysis (LCA) to identify clusters with similar MLTC patterns overall and stratified by area deprivation and ethnicity. We assessed the association between cluster membership and the risk of one-year unplanned hospital admission through Cox proportional hazards regression.

Results:

237,156 (57.7% women) patients were included. LCA identified a total of five distinct MLTC clusters. Across all stratified analyses, we identified five distinct MLTC clusters. Four clusters were identified in the least deprived areas, while three were identified in the most deprived areas and three within each ethnic group. “High-MLTC” (characterised by patients with a high prevalence of most conditions) and “mental health” clusters were observed in all sub-groups. An “inflammatory” cluster (defined by high prevalence of inflammatory conditions) was observed only in the least deprived areas and White ethnicity, whereas a “cardiometabolic” cluster (with high prevalence of cardio-metabolic conditions) was observed solely in the most deprived areas and among Black and Asian ethnic groups. An “obesity” cluster, characterised

by obesity with fewer other conditions, was observed only among those residing in the least deprived areas.

Across all sub-groups, the “High-MLTC” cluster showed the highest unplanned hospitalisation rate: hazard ratio from 2.06 [95%CI 1.94-2.18] to 2.15 [95%CI 1.83-2.54] compared with the lowest risk cluster across deprivation and ethnicity strata.

Conclusions:

This research highlights distinct patient clusters characterised by different patterns of MLTC that have major implications for healthcare use. MLTC clusters across socioeconomic and ethnic sub-groups could inform targeted workforce, resources and personalised interventions directed towards people with MLTC.

Keywords (Three to ten keywords representing the main content of the article):

Clusters, Multiple Long-Term Conditions, Socioeconomic factors, ethnicity, Electronic health records, Primary care data

Introduction

In 2020, the number of people in the United Kingdom (UK) living with multiple long-term conditions (MLTC), defined as two or more co-occurring chronic conditions, exceeded 7 million and is increasing rapidly [1-5]. A recent systematic review estimated the global prevalence of MLTC among adults to be 37.2%, although substantial heterogeneity was observed across studies [6]. In England, the prevalence of MLTC ranges from 0.9% for 0-19 years olds to 68.2% in those aged 80 years and older, with an average overall prevalence of 14.8% [7]. This is projected to increase by one-third by 2049 [8]. Individuals living with MLTC are more likely to have poor health-related quality of life, reduced physical function, greater psychological needs, and increased healthcare utilisation compared to those without MLTC [9, 10]. This highlights the urgent need to address both the condition burden and the inequalities associated with it.

Previous research has shown that long-term conditions (LTC) commonly cluster into distinct patterns [11], highlighting the need for more integrated, person-centred models of care. In parallel, advances in statistical clustering methods have enabled increasingly sophisticated analyses of large population-based cohorts, resulting in a growing body of literature identifying distinct MLTC patient subgroups based on patterns of co-occurring conditions [12]. Importantly, evidence indicates that MLTC are disproportionately prevalent among socioeconomically disadvantaged and ethnic minoritised populations [8, 13]. Nevertheless, scarce evidence shows how socioeconomic and ethnic factors shape patterns of MLTC clustering among people living with multiple conditions [14, 15].

In addition, the burden of MLTC is further reflected in healthcare utilisation, particularly unplanned hospital admissions, where one in three admissions involves individuals living with 5 or more LTC [16, 17]. However, despite emerging evidence linking MLTC clusters with health

care utilisation [15, 18], how these associations vary across demographic populations remains underexplored. To our knowledge, no studies have investigated the association between MLTC clusters and the rate of unplanned hospitalisation across socioeconomic and ethnic groups. Examining these groups separately is important to avoid treating populations with diverse ethnic and socioeconomic backgrounds as homogeneous in their MLTC profiles and associated healthcare risks. Identifying high-risk MLTC subgroups is clinically and strategically important for informing population-level workforce planning, resource allocation and service provision. At the individual level, this could also inform tailored care pathways aimed at improving disease management and reducing health inequalities, particularly among disadvantaged populations, with limited access to resources.

To address these evidence gaps, we sought to (1) identify distinct clusters of patients with co-occurring MLTC specific to different socioeconomic and ethnic groups, and (2) assess the one-year risk of unplanned hospital admissions associated with each MLTC cluster, overall and across socioeconomic and ethnic strata.

Methods

Ethical approval

Clinical Practice Research Datalink (CPRD) has ethics approval from the Health Research Authority to support research using anonymised patient data. The study has been approved by the Independent Scientific Advisory Committee for CPRD (protocol number 24_004361). https://www.cprd.com/approved-studies/identifying-and-characterising-groups-people-multiple-long-term-conditions.

CO-ACTION Programme

CO-ACTION is a multimethod, multisectoral, biopsychosocial applied research programme that aims to improve health, wellbeing, and care coordination for people living with MLTC, with a focus on those in underserved populations (Grant number: NIHR206260). It combines systematic evidence synthesis; data-driven identification of high-need MLTC groups, and participatory co-production to inform a personalised intervention, that integrates health, social care, and voluntary sector support. The intervention will be evaluated in a multi-centre pragmatic randomised controlled trial with process, economic and implementation evaluation to assess its effectiveness, scalability and sustainability. This paper describes the data-driven identification of high-need MLTC clusters used to inform a meaningful and longstanding intervention targeting.

Study design and participants

This is a retrospective cohort study using data from CPRD Gold (January 2025 release), a large primary care database, containing over 21.7 million anonymised records of patients across the UK (England, Wales, Northern Ireland and Scotland) [19, 20].

CPRD Gold data for patients in England were linked to Hospital Episode Statistics (HES) Admitted Patient Care (APC) and Accident and Emergency (A&E), and Outpatient (OP) databases for hospital admissions, to the Townsend Deprivation Index database for socioeconomic status at a residential area level, and to CPRD Gold Ethnicity Record data [21-25]. Linkage to HES was available for eligible patients registered at participating general practices in England who consented to participate in the linkage scheme, who had not opted out or dissented from the sharing of their records, and met the requirements for linkage [26]. Therefore, analysis was restricted to patients within England. All data was mapped to the

Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) to increase future reproducibility in other databases.

Patients were included if they: (i) were registered with an up-to-standard practice for at least one year before 1st January 2018 (index date); (ii) were eligible for linkage to HES; (iii) were aged ≥18 years; and (iv) had at least two prior LTC diagnoses on index date.

Initial selection of MLTC

The selection of long-term conditions was guided by other CO-ACTION programme intervention development work packages, which aimed to define a set of clinically meaningful and high-burden conditions for a targeted multimorbidity intervention. Condition selection was conducted a priori through a structured, multi-source process overseen by the CO-ACTION multidisciplinary MLTC applicants with input from the steering group, including clinical experts, methodological researchers, and Patient and Public Involvement (PPI) contributors.

Condition inclusion was based on triangulation of (i) evidence from systematic literature reviews on MLTC burden and intervention outcomes (PROSPERO; registration number: CRD420251164577), (ii) PPI insights from people living with MLTC highlighting functional limitation, pain, and wellbeing impacts, (iii) national policy priorities including the NIHR MLTC strategy [27], and (iv) epidemiological evidence on associations with unplanned hospital admissions [28, 29] and higher prevalence in socioeconomically deprived and ethnically minoritised populations [30-32].

This process resulted in the pre-specification of 58 high-impact long-term conditions, which were included as candidate variables for clustering analyses (Additional file 1: Table S1). An

LTC was defined as any diagnosis of the condition recorded in CPRD Gold at any time prior to the index date.

Outcome: unplanned hospital admissions

The outcome was one-year unplanned hospital admissions, defined as the first unpredictable and emergency inpatient admission recorded in HES APC within one year of index date. Emergency admissions were identified using the HES admission method field, restricted to non-elective admission codes (including via A&E, GP referral, and other emergency routes). Where A&E attendance records were linked to an inpatient admission, the episode was classified as unplanned if the subsequent admission was coded as an emergency impatient spell. The outcome date was the admission start date of the first eligible hospital episode. Participants were followed from baseline until first unplanned hospital admission, death, deregistration, or end of one-year follow-up.

Explanatory variables

Patients' age was calculated as at index date, and sex was obtained from patient file as recorded (men/women).

Stratification variables for assessing inequalities in MLTC clusters

Socioeconomic status was assessed using the Townsend Deprivation Index [33], an area-level measure of deprivation. The index is derived from UK Census data and comprises four weighted components: unemployment, non-car ownership, non-home ownership, and household overcrowding. Each component is standardised and combined to produce an overall deprivation score for the corresponding area, with higher values indicating higher

levels of socioeconomic deprivation. Patients were categorised into quintiles based on the score, ranging from the least deprived (quintile 1) to the most deprived (quintile 5) [33]. Ethnicity was documented using CPRD Gold Ethnicity Records [34]. This provides a single derived ethnicity category for each patient. Ethnicity data were available and categorised into three ethnic groups: White, Black (Black African, Black Caribbean, and Black Other), and South Asian (Indian, Bangladeshi, and Pakistani).

Statistical Analysis

All analyses were conducted for the whole population and separately stratified by socioeconomic deprivation (first and fifth quintile, least and most deprived areas, respectively) and ethnic groups ('White', 'Black', and 'Asian').

Descriptive analyses

Among all patients with at least two LTC and stratified by deprivation area and ethnic group, key characteristics including age and sex and the number of LTC per patient were summarised, and the frequency of each patient's LTC calculated. No missingness was present for LTC, age and sex. Absence of a recorded condition was considered absence of that condition.

Cluster analysis

Conditions with a prevalence of less than 5% among patients aged 50 and over were excluded for the clustering analysis (Additional file 1: Table S1). A threshold of 5% was selected because the substantial diagnostic heterogeneity and variability in the number of chronic conditions among people with MLTC are likely to result in a large number of infrequently occurring combinations of conditions.

Clusters of patients according to their MLTC distribution were identified using latent class analysis (LCA). LCA is a soft clustering algorithm, which assigns individuals to a non-overlapping cluster in a data-driven approach. LCA divides the population under study into latent subpopulations (i.e., classes/clusters) [35, 36].

Models with 2 to 7 classes were fitted with 100 random starts. To determine the best/optimal cluster solution, we compared the performance of the different models through statistical criteria including entropy R2, the Akaike's Information Criteria (AIC), Bozdogan's Criterion (CAIC), the Bayesian Information Criteria (BIC), log-likelihood ratio, as well as considered clinical plausibility [35-37]. Additionally, we required that the smallest cluster to represent at least 1% of the study sub/population. Clinical judgement was also considered when the statistical criteria were inconclusive.

Survival analysis

After identifying the clusters of patients with MLTC, we conducted a survival analysis to assess the association between clusters with the highest 1-year rate of unplanned hospitalisation. Cluster membership was included as an independent variable, and time to first unplanned hospitalisation within one year was the outcome. Hazard ratios (HRs) were estimated using Cox proportional hazards regression and reported with their 95% confidence intervals (CIs), adjusted for age and sex.

All statistical analyses were conducted using R 4.4.2 for Windows (R Foundation for Statistical Computing, Vienna, Austria).

Results

Study population

Of 3,146,578 adults registered in CPRD Gold in January 2018, 239,279 (7.6%) were eligible for HES linkage and had at least two of the 58 pre-specified LTC recorded in their electronic health record. Following prevalence-based selection (≥5%) of 15 LTC and application of eligibility criteria, 237,156 individuals were included in the cluster analysis (Table 1 and Additional file 1: Table S1).

The distribution of demographic characteristics for the overall sample, stratified by area-level deprivation and ethnicity, is presented in Table 1. From the total MLTC population, 25.5% and 13.2% of patients resided in the least and most deprived areas, respectively. 214,086 (90.3%) were “White”, 5,368 (2.3%) were “Black”, and 10,395 (4.4%) were “Asian”. Overall, 57.7% of patients were female, with a mean age of 60.2 years.

When stratified by area-level deprivation, sex distribution was similar whilst mean age was higher among patients from the least deprived areas (63 years) compared with those from the most deprived areas (56 years). Moreover, patients from the least deprived areas were mostly White (94%), whereas this proportion was lower in the most deprived areas (79%).

Across ethnic groups, the mean age was higher for White patients (60 years) than for Asian (55 years) or Black (54 years) patients. Female comprised the majority of multimorbid individuals across all groups, with the highest proportion observed in the Black population (63%).

Asian and Black populations were mostly from greater socioeconomically deprived areas, with 27% and 49%, respectively, living in the most deprived Townsend quintile. In contrast, a higher proportion of White patients (27%) belonged to the least deprived Townsend quintile.

Table 1. Summary characteristics of patients with ≥2 long-term conditions, for the overall and stratified by quintiles of area deprivation and ethnic groups

Characteristic, n (%)	Overall	Socioeconomic area deprivation		Ethnicity		
	N=237,156	Least Deprived (n=60,590)	Most deprived (n=31,187)	White (n=214,086)	Black (n=5,368)	Asian (n=10,395)
Age, mean (SD)	60.2 (17.2)	62.8 (16.5)	56.4 (17.4)	60.7 (17.3)	54.3 (15.0)	54.8 (16.1)
Age into categories						
18-39 years	31,834 (13.4)	5,738 (9.5)	5,851 (18.8)	27,842 (13.0)	817 (15.2)	1,879 (18.1)
40-49 years	30,443 (12.8)	6,510 (10.7)	4,841 (15.5)	26,012 (12.2)	1,128 (21.0)	2,237 (21.5)
50-64 years	72,352 (30.5)	18,142 (29.9)	10,043 (32.2)	64,109 (29.9)	2,252 (42.0)	3,297 (31.7)
≥65 years	10,2527 (43.2)	30,200 (49.8)	10,452 (33.5)	96,123 (44.9)	1,171 (21.8)	2,982 (28.7)
Sex, Female	136,856 (57.7)	34,106 (56.3)	18,401 (59.0)	123,652 (57.8)	3,389 (63.1)	6,014 (57.9)
Townsend Score, mean (SD)	5.0 (2.8)	1.5 (0.5)	9.4 (0.5)	4.8 (2.7)	7.8 (2.4)	6.3 (2.9)
Townsend score into quintiles						
Q1-Least Deprived	60,590 (25.5)	-	-	56,822 (26.5)	275 (5.1)	1,610 (15.5)
Q2	50,103 (21.1)	-	-	46,659 (21.8)	414 (7.7)	1,509 (14.5)
Q3	49,826 (21.0)	-	-	45,879 (21.4)	575 (10.7)	1,828 (17.6)
Q4	45,381 (19.1)	-	-	39,952 (18.7)	1,453 (27.1)	2,595 (25.0)
Q5-Most deprived	31,187 (13.2)	-	-	24,712 (11.5)	2,650 (49.4)	2,850 (27.4)
Missing	<0.1	-	-	62 (<0.1)	<5 (<0.1)	<5 (<0.1)
Ethnicity						
White	214,086 (90.3)	56,822 (93.8)	24,712 (79.2)	-	-	-
Black	5,368 (2.3)	275 (0.5)	2,650 (8.5)	-	-	-
Asian	10,395 (4.4)	1,610 (2.7)	2,850 (9.1)	-	-	-
Missing	7,307 (3.1)	1,883 (3.1)	975 (3.1)	-	-	-

Distribution of the fifteen MLTC in the study population

The distribution of the 15 LTC for the overall and stratified by area deprivation and ethnicity groups, is presented in Table 2. The top four most prevalent LTC included obesity (54.2%), hypertension (45.9%), depression and/or anxiety (45.0%), and arthritis and/or osteoarthritis (43.4%). Similar individual MLTC distributions were observed across socioeconomic and ethnicity strata.

The prevalence of malignancy (14.8% versus 9.5%) and arthritis/osteoarthritis (47.0% versus 38.1%) was higher among patients in the least deprived areas than in those in the most deprived areas. In contrast, the prevalence of type 2 diabetes (15.0% versus 22.1%) and obesity (48.0% versus 62.4%) was higher among patients in the most deprived areas.

The prevalence of anaemia and type 2 diabetes was higher among patients from Black and Asian groups compared to White ethnic group. In addition, patients from Black ethnic group presented a higher prevalence of obesity (67.7% versus 54.2% and 48.0% among patients from White and Asian ethic groups, respectively).

Table 2. Prevalence distribution of long-term conditions (LTC) for the population with ≥2 conditions, stratified by area deprivation and ethnic groups (excludes conditions with less than 5% prevalence).

LTC, n (%)	Overall	Socioeconomic area deprivation		Ethnicity		
	N=237,156	Least Deprived (n=60,590)	Most deprived (n=31,187)	White (n=214,086)	Black (n=5,368)	Asian (n=10,395)
Total, median (IQR)	3.0 (2.0-4.0)	3.0 (2.0-4.0)	3.0 (2.0-5.0)	3.0 (2.0-5.0)	3.0 (2.0, 4.0)	3.0 (2.0, 4.0)
Anaemia	40,419 (17.0)	9,507 (15.7)	6,128 (19.6)	34,728 (16.2)	1,487 (27.7)	3,303 (31.8)
Anxiety and/or depressive disorder	10,6815 (45.0)	25,212 (41.6)	14,880 (47.7)	98,742 (46.1)	1,642 (30.6)	3,289 (31.6)
Any malignant (neoplastic disease[a])	30,146 (12.7)	8,993 (14.8)	2,969 (9.5)	28,846 (13.5)	357 (6.7)	561 (5.4)
Arthritis and/or osteoarthritis	102,984 (43.4)	28,490 (47.0)	11,894 (38.1)	95,096 (44.4)	1,658 (30.9)	3,996 (38.4)
Asthma	54,532 (23.0)	12,750 (21.0)	8,039 (25.8)	49,318 (23.0)	1,050 (19.6)	2,521 (24.3)
Cardiac arrythmia	29,704 (12.5)	8,449 (13.9)	3,146 (10.1)	28,202 (13.2)	313 (5.8)	751 (7.2)
Chronic kidney disease	35,320 (14.9)	8,865 (14.6)	4,529 (14.5)	32,434 (15.1)	806 (15.0)	1,364 (13.1)
Diabetes type 2	41,647 (17.6)	9,110 (15.0)	6,890 (22.1)	35,417 (16.5)	1,381 (25.7)	3,681 (35.4)
Dyslipidaemia/hyperlipidaemia	55,568 (23.4)	13,553 (22.4)	7,986 (25.6)	50,569 (23.6)	1,031 (19.2)	2,849 (27.4)
Hearing loss	34,974 (14.7)	10,119 (16.7)	3,539 (11.3)	32,945 (15.4)	304 (5.7)	951 (9.1)
Hypertension	108,835 (45.9)	285,66 (47.1)	13,781 (44.2)	98,770 (46.1)	2,664 (49.6)	4,445 (42.8)
Hypothyroidism	25,788 (10.9)	6,930 (11.4)	3,137 (10.1)	23,704 (11.1)	261 (4.9)	1,211 (11.6)
Insomnia	34,155 (14.4)	8,217 (13.6)	4,907 (15.7)	31,123 (14.5)	692 (12.9)	1,373 (13.2)
Obesity	128,622 (54.2)	29,069 (48.0)	19,460 (62.4)	116,036 (54.2)	3,632 (67.7)	4,991 (48.0)
Retinal disorder	21,494 (9.1)	5,785 (9.5)	2,490 (8.0)	18,975 (8.9)	530 (9.9)	1,469 (14.1)

[a]Non-melanoma was excluded; IQR: interquartile range.

Identifying clusters of patients with MLTC for overall population and by socioeconomic area deprivation and ethnicity

A summary of the main conditions characterising each cluster of patients with MLTC identified in the overall population and stratified by socioeconomic deprivation and ethnicity is presented in Table 3. We identified five distinct MLTC clusters across all stratified analyses. Of these, four were observed in the least deprived areas, three in the most deprived areas, and three within each ethnic group.

The “high-MLTC” cluster was characterised by a high prevalence of multiple LTCs; “inflammatory” by high prevalence of musculoskeletal disease and malignancy/ies; “mental health” by high prevalence of depression and anxiety; “obesity” by common obesity with lower co-occurring LTC; and “cardiometabolic” by a high prevalence of type 2 diabetes, dyslipidaemia/hyperlipidaemia, hypertension, and obesity, and a low or very low prevalence of all other LTCs (Table 3).

284 Table 3. Clusters of the Multiple Long-Term Conditions (MLTC) among the whole population, area deprivation and ethnic subgroups

Cluster	Whole population	Least Deprived	Most deprived	White ethnic group	Black ethnic group	Asian ethnic group
High-MLTC	Mainly composed of cardiometabolic factors (diabetes type 2[a], dyslipidaemia / hyperlipidaemia and hypertension, obesity) + retinal disorder Anaemia, cardiac arrythmia and chronic kidney disease	Mainly composed of cardiometabolic factors (diabetes type 2[a], dyslipidaemia / hyperlipidaemia and hypertension, obesity) + retinal disorder Cardiac arrythmia and chronic kidney disease	High prevalence in all conditions, except mental health	Mainly composed of cardiometabolic factors (diabetes type 2[a], dyslipidaemia / hyperlipidaemia and hypertension, obesity) + retinal disorder Anaemia, any malignancy, arthritis/osteoarthritis, cardiac arrythmia and chronic kidney disease	High prevalence in all conditions, except mental health	High prevalence in all conditions, except mental health
Inflammatory	Mainly composed of arthritis/osteoarthritis + any malignancy cardiac arrythmia, chronic kidney disease, dyslipidaemia / hyperlipidaemia, and hypertension	Mainly composed of arthritis/osteoarthritis + any malignancy cardiac arrythmia, chronic kidney disease, dyslipidaemia / hyperlipidaemia, and hypertension		Mainly composed of arthritis/osteoarthritis + any malignancy cardiac arrythmia, chronic kidney disease, dyslipidaemia / hyperlipidaemia, and hypertension		
Mental health	Anxiety/depression disorder, asthma and insomnia The remaining conditions showed low prevalence	Anxiety/depression disorder, asthma and insomnia The remaining conditions showed low prevalence	Anxiety/depression disorder, asthma and insomnia The remaining conditions showed low/very low prevalence	Anxiety/depression disorder, asthma and insomnia The remaining conditions showed low/very low prevalence	Anxiety/depression disorder, anaemia, asthma and insomnia The remaining conditions showed low/very low prevalence	Anxiety/depression disorder, anaemia, asthma and insomnia The remaining conditions showed low/very low prevalence
Obesity	Obesity[a] The remaining conditions showed low prevalence	Obesity[a] The remaining conditions showed low prevalence				
Cardiometabolic			Diabetes type 2, dyslipidaemia / hyperlipidaemia, hypertension and obesity The remaining conditions showed low/very low prevalence		Diabetes type 2[a], hypertension and obesity The remaining conditions showed low/very low prevalence	Diabetes type 2, dyslipidaemia / hyperlipidaemia and hypertension The remaining conditions showed low prevalence

285 [a]100% of patients in the specific cluster had the condition. Only conditions with at least 5% prevalence among patients aged 50 and over were clustered. Clusters were aligned across different
286 populations. Blank cells indicate that the conditions were not present or not clustered.

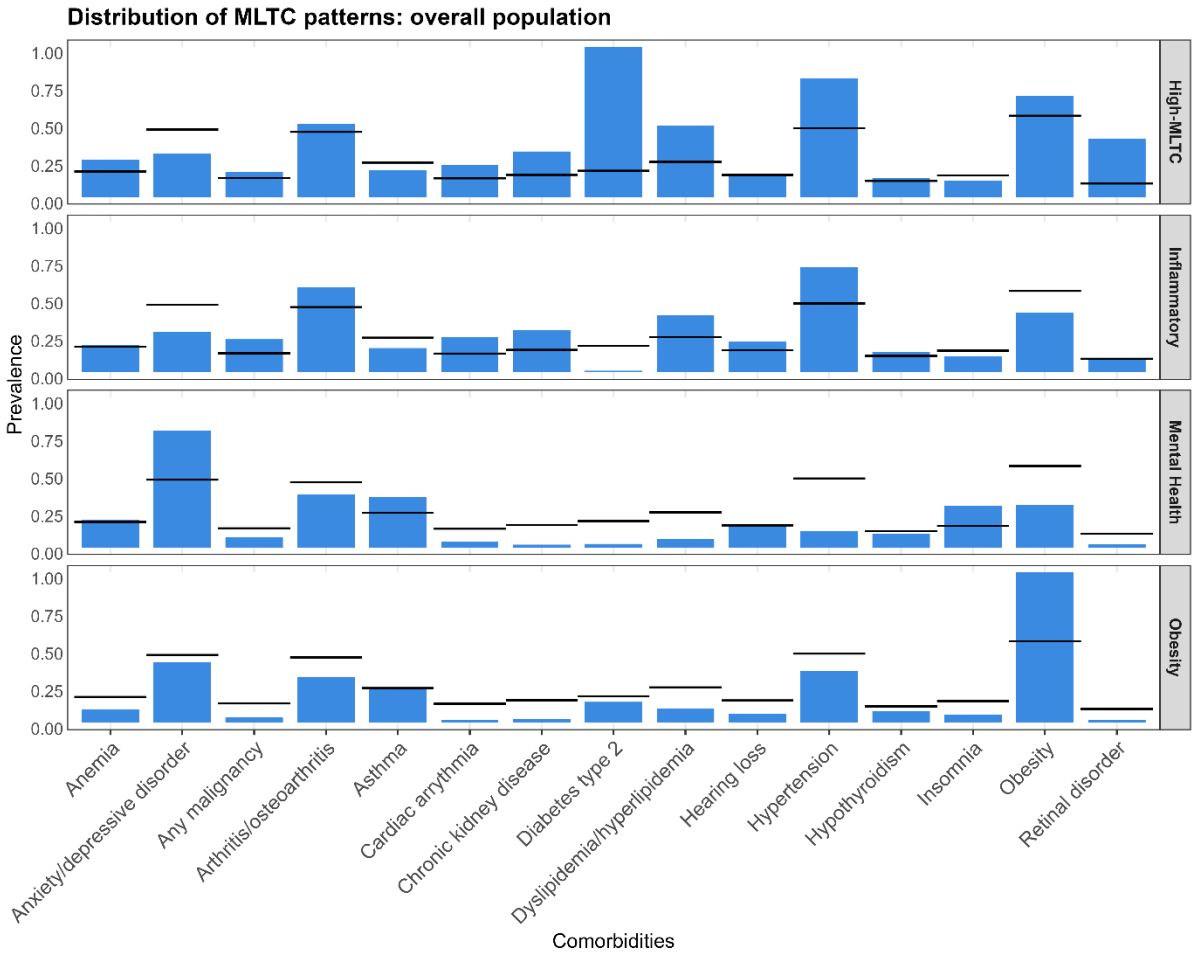
Overall population

After evaluating the fit statistics for the LCA and upon clinical judgement, we identified four distinct MLTC clusters in the overall population (Additional file 1: Tables S2-A, S3 and S4). The identified clusters were: “high-MLTC” (n=32,685; 13.8%), “inflammatory” (n=77,147; 32.5%), “mental health” (n=60,783; 25.6%) and “obesity” (n=66,541; 28.1%) (Table 3 and Figure 1).

People in the “inflammatory” cluster were oldest (mean age 71.2 years (SD:13.0)), while people in the “mental health” cluster were the youngest group (mean 49.8 years (SD: 16.1)) (Additional file 1: Table S5). Female were overrepresented in the “mental health” cluster (n=40,887; 67.3%) and underrepresented in the “high-MLTC” cluster (n=14,761; 45.2%). Additionally, the “high-MLTC” and “obesity” clusters were most prevalent among patients from Black ethnic group, and among patients from the most deprived areas.

Figure 1 Title: Distribution of Multiple Long-term Conditions (MLTC) across clusters for the overall population.

Figure 1 Legend: **Black lines** correspond to the proportion of the condition in the overall study population.



Least deprived areas group

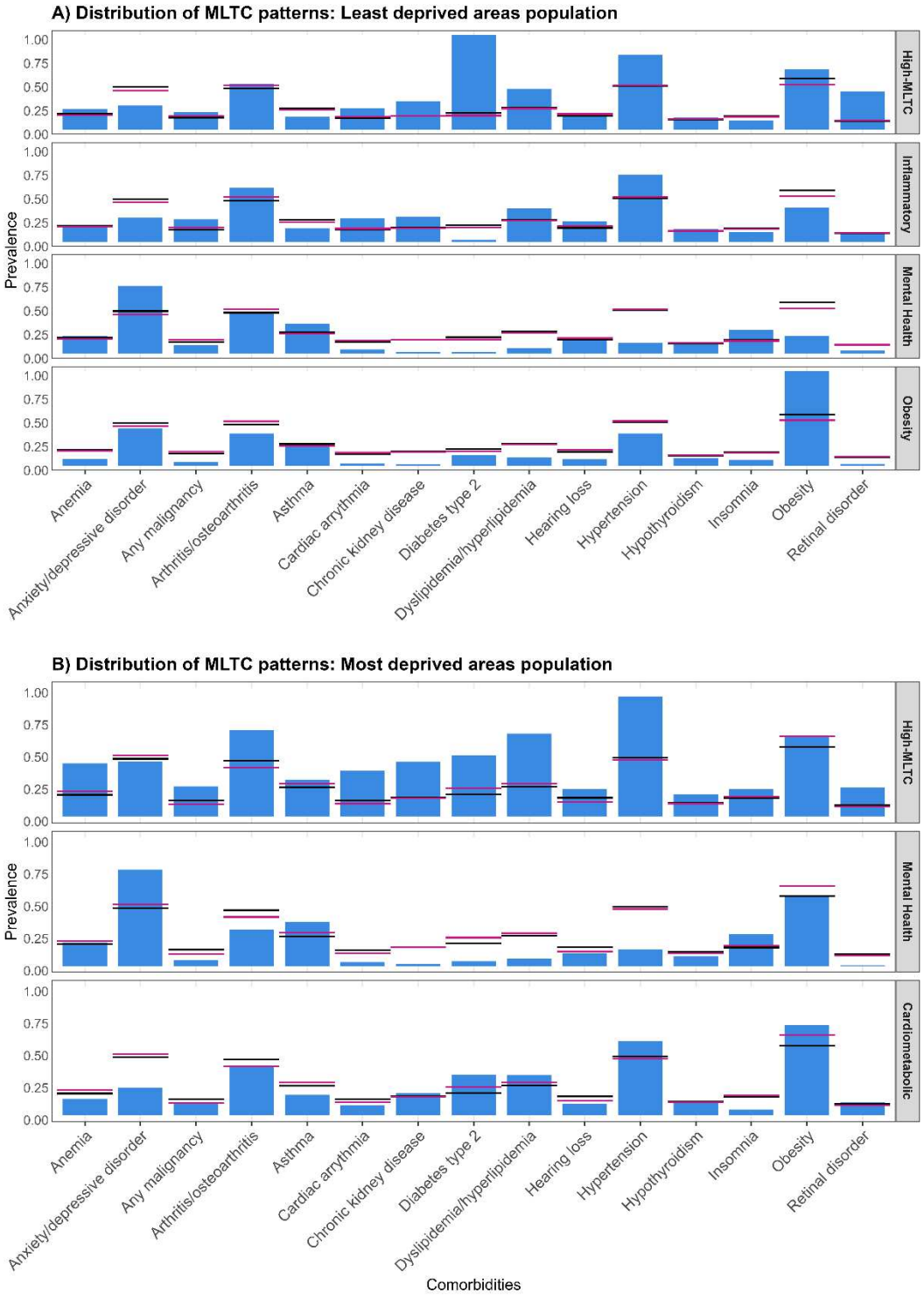
A four-cluster solution was considered the optimal/final solution in the group of individuals living in the least deprived areas (Additional file 1: Tables S2-B, S3 and S4).

The same clusters composition as for the overall population was identified: “high-MLTC” (n=7,316; 12.1%), “inflammatory” (n=21,701; 35.8%), “mental health” (n=15,908; 26.3%) and “obesity” (n=15,665; 25.9%) (Table 3 and Figure 2-A).

311 The distribution of demographic characteristics was also similar to that seen in the overall
312 population, except for ethnicity which was less diverse (Additional file 1: Table S6): more than
313 90% of the population was from the White ethnic group, and less than 1% patients were from
314 Black ethnic group.
315
316 Most deprived areas group
317 Three clusters were found in the group of patients living in the most deprived areas based on
318 model fit, classification quality, parsimony, and interpretability (Additional file 1: Tables S2-C,
319 S3 and S4). The clusters were named: “high-MLTC”, “mental health” and “cardiometabolic”
320 (Table 3 and Figure 2-B). The most prevalent cluster was “mental health”, with 44% (n=13,728)
321 of individuals in these strata (Additional file 1: Table S7). The “high-MLTC” was more prevalent
322 in the oldest age group (≥65 years; 75.5%), “mental health” among the youngest (18-39 years;
323 36.1%) and the “cardiometabolic” cluster among middle-aged group (50-64 years; 37.9%)
324 (Additional file 1: Table S7). The “cardiometabolic” cluster was more common among male,
325 and individuals from Black and Asian ethnic groups compared with the other clusters
326 (Additional file 1: Table S7).
327

Figure 2 Title: Distribution of Multiple Long-Term Conditions (MLTC) across clusters stratified by deprivation.

Figure 2 Legend: **Black** and **Pink** lines represent the proportion of each condition in the overall population and in the area-specific deprivation groups, respectively. A) Least deprived socioeconomic area (Townsend score quintile 1); B) Most deprived socioeconomic area (Townsend score quintile 5)



Ethnic group - White

A three-cluster solution was selected as the best/optimal (Additional file 1: Tables S2-D, S3 and S4), including: “high-MLTC” (n=32,485; 15.2%), “inflammatory” (n=82,278; 38.4%) and “mental health” (n=99,323; 46.4%) (Table 3 and Figure 3-A).

Individuals in the “high-MLTC” and “inflammatory” clusters were older (mean age (SD): 70.4 years (13.0) and 68.5 years (13.2), respectively), while those in the “mental health” cluster were the youngest (50.2 years (15.5)). Both the “inflammatory” and “mental health” clusters were more common in female (n=44,014; 53.5% and n=65,409; 65.9%, respectively) (Additional file 1: Table S8).

Ethnic group – Black

Three clusters were found in the Black ethnic group: “high-MLTC” (n = 499; 9.3%), “mental health” (n = 2,054; 38.3%) and “cardiometabolic” (n=2,815; 52.4%) (Table 3, Figure 3-B and Additional file 1: Tables S2-E, S3 and S4).

The cluster labelled as “high-MLTC” was characterised by a higher prevalence of all conditions compared with the overall population and with the overall Black ethnic population (Figure 3-B). The “mental health” cluster was characterised by a significant proportion of people with diagnoses of anxiety and depression, and the “cardiometabolic” cluster by people with diabetes type 2, hypertension, and obesity.

The characteristics of patients in the Black ethnic group, divided by MLTC clusters are presented in Additional file 1: Table S9. The “high-MLTC” cluster had a higher proportion of individuals aged 65 and over (n=348; 69.7%), and the “mental health” cluster was more prevalent among female (n=1,496; 72.8%).

Ethnic group - Asian

We identified three clusters as the best solution in this stratum. The decision was based on model fit statistics, clinically meaningful and distribution of comorbidity patterns (Table 3, Figure 3-C and Additional file 1: Tables S2-F, S3 and S4).

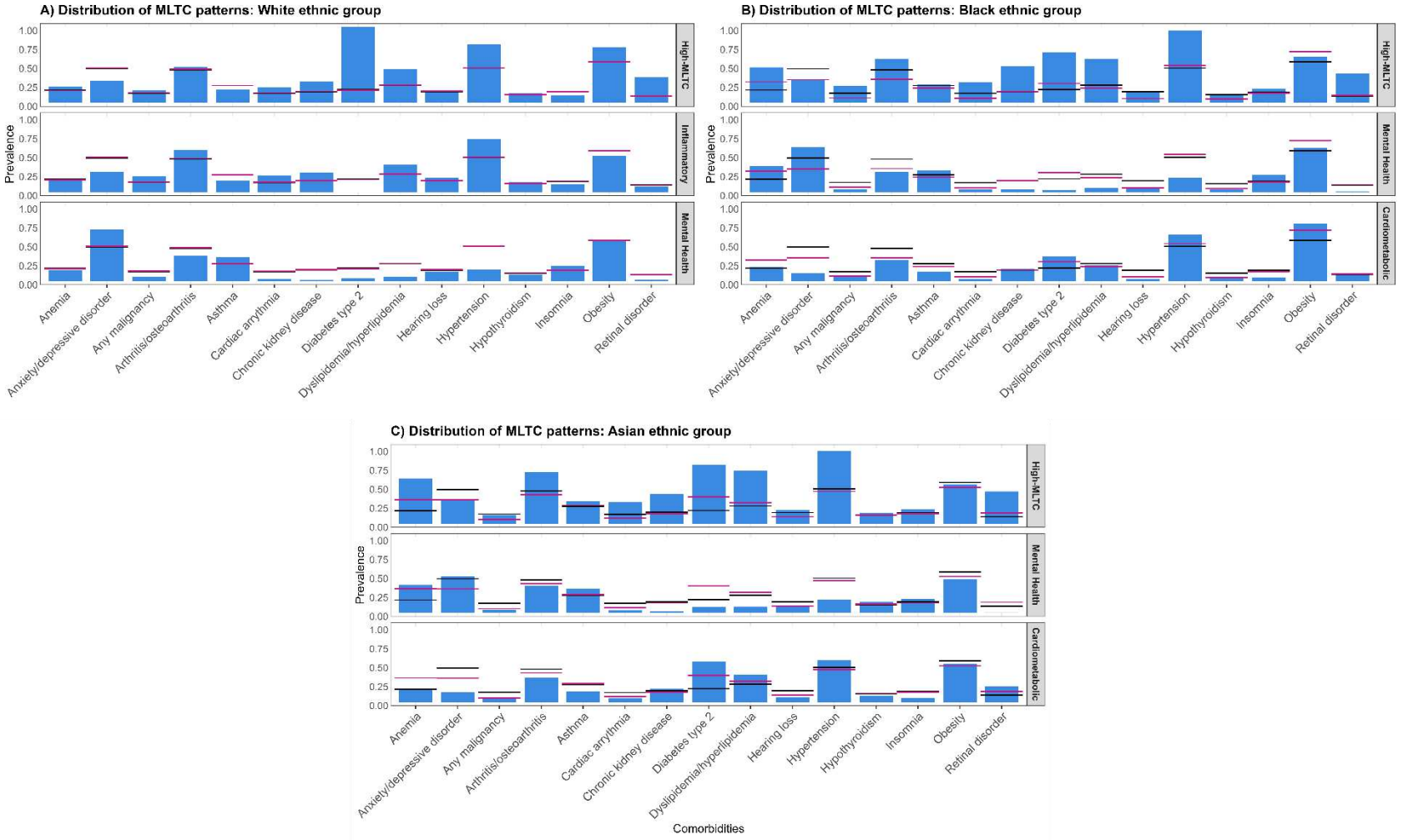
The clusters identified were: “high-MLTC” (characterised by mostly all LTC, except mental health problems; n=1,194-11.5%), “mental health” (n=4,863-46.8%), and “cardiometabolic” (n=4,338-41.7%) (dominated by type 2 diabetes, dyslipidaemia/hyperlipidaemia and hypertension, and retinal disorders) (Table 3 and Figure 3-C).

The “high-MLTC” cluster had the oldest patients (mean age 71.0 years (SD:12.0)), while the “mental health” cluster had the youngest (mean 47.0 years (SD: 14.3)) (Additional file 1: Table S10). Female were overrepresented in the “high-MLTC” (n=662; 55.4%) and “mental health” clusters (n=3,387; 69.6%) and underrepresented in the “cardiometabolic” cluster (n=1,965; 45.3%).

Figure 3 Title: Distribution of Multiple Long-Term Conditions (MLTC) clusters stratified by ethnicity.

Figure 3 Legend: **Black** and **Pink** lines represent the proportion of each condition in the overall population and in the ethnic-specific groups, respectively.

A) White; B) Black; C) Asian.



Association between MLTC clusters and risk of unplanned hospitalisation

The one-year cumulative incidence of unplanned hospitalisation ranged from 21.7% in the obesity cluster to 33.1% in the high-MLTC cluster, with similar patterns observed across deprivation and ethnic groups (Table 4).

The clusters “mental health” (adjusted HR 1.16 [95% CI: 1.13-1.19]), “inflammation” (adjusted HR 1.38 [95% CI: 1.35-1.42]) and ‘high-MLTC’ (adjusted HR 1.69 [95% CI: 1.65-1.74]) had increased risk of one-year unplanned hospitalisation compared with the cluster with the lowest-risk (“obesity”) in the overall population.

The highest risk of unplanned hospitalisation was also found in the “high-MLTC” cluster across all socioeconomic and ethnic groups. The risk was approximately doubled among patients from most deprived areas (adjusted HR 2.06 [95% CI: 1.94-2.18]) and among people from Black (adjusted HR 2.15 [95% CI: 1.83-2.54]) or Asian (adjusted HR 2.12 [95% CI: 1.90-2.37]) ethnic groups compared with the cluster with the lowest risk within each population.

An association between the “inflammatory” cluster and unplanned hospital admission was also observed in the least deprived areas (adjusted HR: 1.40 [95% CI: 1.34-1.47] compared with the “obesity” cluster) and among the White ethnic group (adjusted HR: 1.17 [95% CI: 1.15-1.20] compared with the “mental health” cluster).

The “mental health” cluster had adjusted HRs ranging from 1.09 (95% CI: 1.04-1.15) among individuals from least deprived areas to 1.19 (95% CI: 1.05-1.35) among those in the Asian ethnic group, compared with the cluster with the lowest risk within each specific population.

Table 4. Regression results for Multiple Long-Term Conditions (MLTC) clusters and unplanned hospitalisation, overall and stratified by area deprivation and ethnic groups.

MLTC cluster	No. of 1-year unplanned hospitalisation/total in cluster (%)	Hazard Ratio	95% confidence interval	p-value
Overall				
Obesity (ref)	14,441/66,541 (21.7)	1.00	-	-
Mental health	15,483/60,783 (25.5)	1.16	1.13-1.19	<0.001
Inflammatory	21,960/77,147 (28.5)	1.38	1.35-1.42	<0.001
High-MLTC	10,813/32,685 (33.1)	1.69	1.65-1.74	<0.001
Least deprived area				
Obesity (ref)	2,919/15,665 (18.6)	1.00	-	-
Mental health	3,309/15,908 (20.8)	1.09	1.04-1.15	<0.001
Inflammatory	5,698/21,701 (26.3)	1.40	1.34-1.47	<0.001
High-MLTC	2,175/7,316 (29.7)	1.65	1.56-1.75	<0.001
Most deprived area				
Cardiometabolic (ref)	3,470/13,223 (26.2)	1.00	-	-
Mental health	4,163/13,728 (30.3)	1.10	1.04-1.15	<0.001
High-MLTC	1,973/4,236 (46.6)	2.06	1.94-2.18	<0.001
White ethnic group				
Mental health problems (ref)	24,348/99,323 (24.5)	1.00		
Inflammatory	22,744/82,278 (27.6)	1.17	1.15-1.20	<0.001
High-MLTC	10,401/32,485 (32.0)	1.43	1.39-1.47	<0.001
Black ethnic group				
Cardiometabolic (ref)	666/2,815 (23.7)	1.00	-	-
Mental health/Anaemia	588/2,054 (28.6)	1.19	1.05-1.35	0.005
High-MLTC	221/499 (44.3)	2.15	1.83-2.54	<0.001
Asian ethnic group				
Cardiometabolic (ref)	977/4,338 (22.5)	1.00	-	-
Mental health/Anaemia	1,225/4,863 (25.2)	1.11	1.01-1.22	0.03
High-MLTC	514/1,194 (43.0)	2.12	1.90-2.37	<0.001

Note: all models were adjusted for age and sex.

Discussion

Summary of findings and comparison with the existing literature

This study significantly expands previous evidence in the field of MLTC by demonstrating that clusters of patients with MLTC vary across socioeconomic deprivation and ethnic groups and are differentially associated with unplanned hospitalisation risk, which constitutes an important shift for intervention research, clinical agenda and resources.

A small number of consistent MLTC patterns were identified across population strata, with "high-MLTC" and "mental health" clusters observed across all groups. However, the "high-MLTC" clusters differed in their characteristics across socioeconomically deprived and ethnic groups. Distinct patterns were also evident across sub-groups: the "inflammatory" cluster was restricted to the least deprived and White ethnic populations, while an "obesity"-related cluster was observed only in the least deprived group. In contrast, the "cardiometabolic" cluster was only identified in socioeconomically deprived and ethnically minoritised populations.

In addition, our study identified a strong association between membership in the "high-MLTC" cluster and increased risk of unplanned hospital admission, especially among patients from deprived areas and among minority ethnic groups.

These patterns may reflect cumulative exposure to structural and social determinants of health, including socioeconomic disadvantage, earlier onset of metabolic risk factors, and inequalities in access to preventive care. Barriers such as transport and subsistence costs, work or caring responsibilities, and knowledge of when to seek care may limit access to prevention and effective disease management, potentially contributing to differences in MLTC patterns across population groups [38, 39].

Previous systematic reviews have identified recurring clusters of MLTC that are consistently observed across different populations, and regardless of the method used, aligning with our findings [12, 14, 40, 41]. In particular, clusters of cardiometabolic conditions, such as hypertension, diabetes, and cardiovascular disease are among the most commonly reported. Similarly, mental health conditions and musculoskeletal diseases, including arthritis and related disorders, are frequently identified as distinct clusters.

Our results are comparable to previously identified patterns of multimorbidity, highlighting the robustness of our findings. However, comparison with other studies should be interpreted with caution. Variations in the definition of MLTC (i.e., two or more chronic conditions), the number and type of conditions included, data sources, settings, and clustering methodologies make direct comparisons challenging [11].

Our results showed some disparities in MLTC across strata. Individuals from deprived areas and minority ethnic groups were younger compared with those from more affluent areas or from the White ethnic group. These findings are consistent with previous studies [42-44], which showed that individuals from deprived areas tend to experience earlier onset of MLTC. Additionally, in the Asian ethnic group, we observed a high prevalence of anaemia, which is consistent with other studies where a higher prevalence was observed in Asians compared to non-Asians [45, 46].

Strengths and limitations

A major strength of this study is the use of a large primary care database, with comprehensive information on LTC, and unplanned hospital admissions among an adult population with at least two long-term conditions throughout England. The linkage with HES data helped to characterise the patterns of how patients with MLTC utilise healthcare resources. In addition,

448 the large sample size allowed analyses stratified by area-level socioeconomic status and
449 ethnicity, enabling the identification of groups-specific patterns of MLTC that could inform
450 workforce planning, resource allocation and the provision of specialist services for both
451 prevention and treatment. Nevertheless, given the observational nature of this study, these
452 findings provide insights into the characteristics of patients with MLTC and its association with
453 unplanned hospitalisation but cannot establish causality. A separate manuscript focuses
454 specifically on estimates of the healthcare and economic burden associated with the care of
455 people in the clusters identified in this study and examines differences across levels of care
456 and patient subgroups (cross-reference to sister paper by Fabiano et. al.).
457 Furthermore, CPRD Gold is broadly representative of the UK population [19]. Although
458 analyses of hospitalisation were restricted to patients in England who were eligible for linkage
459 to HES, previous work has shown that CPRD-HES linked data are generally representative of
460 the English population, with high data completeness and strong agreement with HES [47].
461 Additionally, area-level socioeconomic status and ethnicity were also subject to limitations:
462 the components used in the Townsend index may vary in their ability to capture deprivation
463 across different settings and may not fully reflect individual-level socioeconomic
464 circumstances. Regarding ethnicity, the data may lack granularity, as broad ethnic categories
465 can mask important heterogeneity within groups in terms of cultural, social, and health-
466 profiles. Future studies require cohorts that are more diverse in ethnic and socioeconomic
467 backgrounds.
468 Finally, cluster analysis was limited to 15 highly prevalent MLTC. Although this approach
469 excludes less common or socially patterned conditions, potentially leading to an
470 underestimation of the population prevalence of MLTC, it represents the greatest population

burden and can help with the identification of patients living with MLTC for clinical trials or interventions.

Implications for research and clinical practice

These findings suggest that MLTC patterns are socially patterned rather than randomly distributed, reinforcing the need to move beyond single-disease models towards cluster-informed, equity-focused approaches to care. The observed inequalities in MLTC patterns and complexity across socioeconomically deprived and minoritised ethnic populations suggests that MLTC services should be explicitly designed around inequality gradients in clinical and social need rather than average population need. This is particularly important, as population groups experiencing more complex MLTC patterns may also face substantial barriers to accessing and engaging with healthcare service.

For interventional research these findings are particularly relevant, as these can directly inform the identification of high-need target groups for recruitment into personalised, needs-based assessment and biopsychosocial interventions, which could result in more relevant outcomes for this population, with the aim of being implemented in practice. MLTC cluster-informed approaches can provide a pragmatic framework for stratification and targeting of integrated support aimed at improving physical and mental health outcomes and reducing avoidable hospital use and health inequalities. Future interventions should establish whether targeting support according to clusters of patients with MLTC, alongside socioeconomic and ethnic inequalities, leads to more equitable healthcare use and improved outcomes. This could provide an important evidence base for translating MLTC cluster research into more targeted and equitable models of care.

Conclusions

In conclusion, our study has identified different groups of patients, each with distinct clusters of MLTC with important implications for healthcare planning and delivery. High-burden MLTC clusters were observed across a large primary care population, but their characteristics differed across specific subgroups, indicating inequalities in the patterns and complexity of MLTC among socioeconomically disadvantaged and ethnic minority populations. Distinct MLTC clusters varied across socioeconomic and ethnic groups, including an “inflammatory” cluster specifically in the least deprived and White ethnic populations. Several of the identified clusters were associated with an increased risk of unplanned hospitalisation. More effective MLTC care will require healthcare systems designed around patterns of need rather than individual diseases, with the aim of reducing health inequalities and improving quality of life among people living with MLTC.

List of abbreviations

A&E: Accident and Emergency

AIC : Akaike's Information Criteria

APC : Admitted Patient Care

BIC: Bayesian Information Criteria

CAIC: Akaike's Information Bozdogan's Criterion

CIs: confidence intervals

CPRD Gold: Clinical Practice Research Datalink Gold

HES: Hospital Episodes Statistics

HRs: Hazard ratios

IQR: interquartile range

LCA: latent class analysis

MLTC: multiple long-term conditions

NHS: National Health Service

NIHR: National Institute for Health and Care Research

OMOP-CDM: Observational Medical Outcomes Partnership Common Data Model

OP: Outpatient

PPI: Patient and Public Involvement

SD: Standard deviation

UK: United Kingdom

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are available from CPRD, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

Competing interests

RPV has received consultancy fees from Kyowa Kirin Services and UCB Biopharma unrelated to this study. DPA is Principal Investigator for Real World Evidence studies sponsored by UCB Biopharma and Theramex, and a co-investigator for a study sponsored by Eli Lilly, and previous co-investigator for a study sponsored by Novartis. He has been a co-investigator in IMI-funded projects (EHDEN, EU-PEARL) and is currently involved in OPTIMA ONCOLOGY, while leading the IHI-funded GREG (ihi-greg.eu) programme. His research group has received public funding from NIHR, and the European Medicines Agency. He is also a member of the ENCePP Steering Group. LE is a full-time employee at Novartis Pharmaceuticals UK Ltd, however, her contribution to this study was made solely during her prior employment at the University of Oxford and is fully independent of her current role.

MTSS, XC, MPM, GF, AD, LC, MCP and DK have no conflict of interest.

Funding

This study was funded by the National Institute for Health and Care Research (NIHR) Programme Grants for Applied Research (PGfAR), Grant number NIHR206260. MTSS is funded by the NIHR, Award Reference Number: 303693. This study was supported by the Oxford NIHR

Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Authors' contributions

MTSS, RPV, LC, DPA, DK, MCP were involved in the study conception and design. MTSS, LE, GF, AD performed data acquisition and curation. MTSS, XC, LE, MPM developed the analytical framework and conducted the analyses. MTSS, XC, LE, MPM, GF, RPV, LC, DPA, DK, MCP led the initial interpretation of results. MTSS wrote the first draft of the manuscript. All authors provided critical input during the revision of the paper.

Acknowledgements

This study is based on data from the Clinical Practice Research Datalink obtained under license from the UK Medicines and Healthcare products Regulatory Agency (MHRA). The data are provided by patients and collected by the NHS as part of their care and support. The interpretation and conclusions contained in this study are those of the author/s alone, and not necessarily those of the MHRA, NIHR, NHS, or the Department of Health and Social Care. We would like to acknowledge all the data providers and general practices who make anonymised data available for research.

The authors gratefully acknowledge Hayley Doman and Sue Latter for their helpful review of an earlier draft of this manuscript.

AI use

During the preparation of this work, the authors used AI for language editing and for code review. AI was not used for data analysis, interpretation of results or any scholarly

judgement. The authors reviewed and edited all output and take full responsibility for the content of the publication.

References

1. **NIHR Strategic Framework for Multiple Long-Term Conditions** [https://www.nihr.ac.uk/nihr-strategic-framework-multiple-long-term-conditions]
2. **Department of Health Long Term Conditions Compendium of Information: Third Edition** [https://www.gov.uk/government/publications/long-term-conditions-compendium-of-information-third-edition]
3. **Age UK. 1m more living with multiple conditions by 2020. Press Realease.** [https://www.ageuk.org.uk/latest-press/archive/one-million-more-older-people-will-be-living-with-multiple-long-term-conditions/#:~:text=Today%20an%20estimated%20six%20million%20people%20aged%2060,a%20million%20to%20more%20than%207%20million%20.]
4. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B: **Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study**. *The Lancet* 2012, **380**(9836):37–43.
5. Whitty CJM, MacEwen C, Goddard A, Alderson D, Marshall M, Calderwood C, Atherton F, McBride M, Atherton J, Stokes-Lampard H *et al*: **Rising to the challenge of multimorbidity**. *BMJ (Clinical research ed)* 2020, **368**:l6964.
6. Chowdhury SR, Chandra Das D, Sunna TC, Beyene J, Hossain A: **Global and regional prevalence of multimorbidity in the adult population in community settings: a systematic review and meta-analysis**. *EClinicalMedicine* 2023, **57**:101860.
7. Valabhji J, Barron E, Pratt A, Hafezparast N, Dunbar-Rees R, Turner EB, Roberts K, Mathews J, Deegan R, Cornelius V *et al*: **Prevalence of multiple long-term conditions (multimorbidity) in England: a whole population study of over 60 million people**. *J R Soc Med* 2024, **117**(3):104–117.
8. Head A, Birkett M, Fleming K, Kypridemos C, O'Flaherty M: **Socioeconomic inequalities in accumulation of multimorbidity in England from 2019 to 2049: a microsimulation projection study**. *Lancet Public Health* 2024, **9**(4):e231–e239.
9. Sathanapally H, Sidhu M, Fahami R, Gillies C, Kadam U, Davies MJ, Khunti K, Seidu S: **Priorities of patients with multimorbidity and of clinicians regarding treatment and health outcomes: a systematic mixed studies review**. *BMJ Open* 2020, **10**(2):e033445.
10. Frolich A, Ghith N, Schiotz M, Jacobsen R, Stockmarr A: **Multimorbidity, healthcare utilization and socioeconomic status: A register-based study in Denmark**. *PLoS One* 2019, **14**(8):e0214183.
11. Ferris JK, Fiedeldey LK, Kim B, Clemens F, Irvine MA, Hosseini SH, Smolina K, Wister A: **A systematic review and meta-analysis of disease clusters in multimorbidity**. *Nat Commun* 2025, **17**(1):634.
12. Busija L, Lim K, Szoeke C, Sanders KM, McCabe MP: **Do replicable profiles of multimorbidity exist? Systematic review and synthesis**. *Eur J Epidemiol* 2019, **34**(11):1025–1053.

13. Hayanga B, Stafford M, Becares L: **Ethnic inequalities in multiple long-term health conditions in the United Kingdom: a systematic review and narrative synthesis**. *BMC Public Health* 2023, **23**(1):178.
14. Mikula-Noble N, Cormie V, McCowan RE, McCowan C: **Identifying clusters of multimorbid disease and differences by age, sex, and socioeconomic status: A systematic review**. *PLoS One* 2025, **20**(8):e0329794.
15. Eto F, Samuel M, Henkin R, Mahesh M, Ahmad T, Angdembe A, Hamish McAllister-Williams R, Missier P, N JR, M RB *et al*: **Ethnic differences in early onset multimorbidity and associations with health service use, long-term prescribing, years of life lost, and mortality: A cross-sectional study using clustering in the UK Clinical Practice Research Datalink**. *PLoS medicine* 2023, **20**(10):e1004300.
16. **Briefing: Emergency hospital admissions in England: which may be avoidable and how?** [https://www.health.org.uk/sites/default/files/Briefing_Emergency%2520admissions_web_final.pdf]
17. **Health and Care White Paper. Department of Health and Social Care. Working together to improve health and social care for all** [https://www.gov.uk/government/publications/working-together-to-improve-health-and-social-care-for-all]
18. Krauth SJ, Steell L, Ahmed S, McIntosh E, Dibben GO, Hanlon P, Lewsey J, Nicholl BI, McAllister DA, Smith SM *et al*: **Association of latent class analysis-derived multimorbidity clusters with adverse health outcomes in patients with multiple long-term conditions: comparative results across three UK cohorts**. *EClinicalMedicine* 2024, **74**:102703.
19. Sanchez-Santos MT, Axson EL, Dedman D, Delmestri A: **Data Resource Profile Update: CPRD GOLD**. *Int J Epidemiol* 2025, **54**(4).
20. Clinical Practice Research D: **CPRD GOLD January 2025**. In.: Clinical Practice Research Datalink; 2025.
21. Clinical Practice Research D: **CPRD GOLD Ethnicity Record December 2024 (Version 2024.12.001)**. In.: Clinical Practice Research Datalink; 2024.
22. Clinical Practice Research D: **CPRD GOLD Small Area data (patient) November 2024 (Version 2024.11.001)**. In.: Clinical Practice Research Datalink; 2024.
23. Clinical Practice Research D: **CPRD Aurum HES OP November 2024 (Version 2024.11.001)**. In.: Clinical Practice Research Datalink; 2024.
24. Clinical Practice Research D: **CPRD GOLD HES A&E August 2021 (Version 2021.08.001)**. In.: Clinical Practice Research Datalink; 2021.
25. Clinical Practice Research D: **CPRD GOLD HES APC November 2024 (Version 2024.11.001)**. In.: Clinical Practice Research Datalink; 2024.
26. Chan JE, Barnett R, Hodgson S, Chohan P, Mantovani G, Campbell J: **Linking primary care data from clinical practice research datalink to secondary care and other health-related patient data: update and implications**. *Int J Popul Data Sci* 2026, **11**(1):3069.
27. **Department of Health and Social Care, UK. Major conditions strategy : call for evidence** [https://www.gov.uk/government/calls-for-evidence/major-conditions-strategy-call-for-evidence/major-conditions-strategy-call-for-evidence]

28. Stokes J, Guthrie B, Mercer SW, Rice N, Sutton M: **Multimorbidity combinations, costs of hospital care and potentially preventable emergency admissions in England: A cohort study**. *PLoS medicine* 2021, **18**(1):e1003514.
29. Marengoni A, Nobili A, Pirali C, Tettamanti M, Pasina L, Salerno F, Corrao S, Iorio A, Marcucci M, Franchi C *et al*: **Comparison of disease clusters in two elderly populations hospitalized in 2008 and 2010**. *Gerontology* 2013, **59**(4):307–315.
30. Olanrewaju O, Trott M, Smith L, Lopez Sanchez GF, Carmichael C, Oh H, Schuch F, Jacob L, Veronese N, Soysal P *et al*: **Chronic physical conditions, physical multimorbidity, and quality of life among adults aged >/= 50 years from six low- and middle-income countries**. *Qual Life Res* 2023, **32**(4):1031–1041.
31. McLean G, Gunn J, Wyke S, Guthrie B, Watt GC, Blane DN, Mercer SW: **The influence of socioeconomic deprivation on multimorbidity at different ages: a cross-sectional study**. *The British journal of general practice : the journal of the Royal College of General Practitioners* 2014, **64**(624):e440–447.
32. Olutende MO, Mse E, Wanzala MN, Wamukoya EK: **The influence of socio-economic deprivation on multimorbidity: A systematic review**. *European Journal of Physical Education and Sport Science* 2021, **6**(12).
33. **UK Townsend Deprivation Scores from 2011 census data** [https://statistics.ukdataservice.ac.uk/dataset/2011-uk-townsend-deprivation-scores]
34. **CPRD Ethnicity Records Documentation** [https://www.cprd.com/sites/default/files/2025-09/CPRD_EthnicityRecord_Documentation_v1.2.pdf]
35. Nylund KL, Asparouhov T, Muthén BO: **Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study**. *Structural Equation Modeling: A Multidisciplinary Journal* 2007, **14**(4):535–569.
36. Sinha P, Calfee CS, Delucchi KL: **Practitioner's Guide to Latent Class Analysis: Methodological Considerations and Common Pitfalls**. *Critical care medicine* 2021, **49**(1):e63–e79.
37. Pineda Moncusi M, Strauss V, Robinson D, Prieto Alhambra D, Khalid S: **Unsupervised learning to understand patterns of comorbidity in 633,330 patients diagnosed with osteoarthritis**. In: *13th International Conference on Bioinformatics Models, Methods and Algorithms*. vol. 3: SciTePress; 2022: 121–129.
38. Bidmead E, Hayes L, Mazzoli-Smith L, Wildman J, Rankin J, Leggott E, Todd L, Bramhall L: **Poverty proofing healthcare: A qualitative study of barriers to accessing healthcare for low-income families with children in northern England**. *PLoS One* 2024, **19**(4):e0292983.
39. Fiscella K, Franks P, Gold MR, Clancy CM: **Inequality in quality: addressing socioeconomic, racial, and ethnic disparities in health care**. *JAMA* 2000, **283**(19):2579–2584.
40. Prados-Torres A, Calderon-Larranaga A, Hancco-Saavedra J, Poblador-Plou B, van den Akker M: **Multimorbidity patterns: a systematic review**. *J Clin Epidemiol* 2014, **67**(3):254–266.

41. Zhu Y, Edwards D, Mant J, Payne RA, Kiddle S: **Characteristics, service use and mortality of clusters of multimorbid patients in England: a population-based study**. *BMC Med* 2020, **18**(1):78.
42. Liu Y, Bankhead C, Wright Drakesmith C, Pope C, Gonzalez-Chica D, Swain S, CoMPuTE., Heneghan C, Perera-Salazar R, Zhu T: **Identifying profiles, trajectories, burden, social and biological factors in 3.3 million individuals with multimorbidity in England**. *medRxiv* 2025.
43. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B: **Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study**. *Lancet* 2012, **380**(9836):37–43.
44. Head A, Fleming K, Kypridemos C, Schofield P, Pearson-Stuttard J, O'Flaherty M: **Inequalities in incident and prevalent multimorbidity in England, 2004-19: a population-based, descriptive study**. *Lancet Healthy Longev* 2021, **2**(8):e489–e497.
45. Fischbacher C, Bhopal R, Patel S, White M, Unwin N, Alberti KG: **Anaemia in Chinese, South Asian, and European populations in Newcastle upon Tyne: cross sectional study**. *BMJ (Clinical research ed)* 2001, **322**(7292):958–959.
46. Ahmad T, Ullah A, Chelala C, Taylor SJC: **Ethnic and Socio-Economic Variations in Comorbidity and Mortality in Cancer Survivors: A UK Population-Based Observational Study**. *Cancers (Basel)* 2025, **17**(6).
47. Shiekh SI, Harley M, Ghosh RE, Ashworth M, Myles P, Booth HP, Axson EL: **Completeness, agreement, and representativeness of ethnicity recording in the United Kingdom's Clinical Practice Research Datalink (CPRD) and linked Hospital Episode Statistics (HES)**. *Popul Health Metr* 2023, **21**(1):3.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformationv2.docx