

Combining polygenic risk scores to understand genetic liability to physical-mental health multimorbidity in UK Biobank

Daniel Stow^{1,†,*}, Ruby S. M. Tsang^{2,†}, Ioanna K. Katzourou^{3,†}, Jack F. G. Underwood^{2,4}, The Lifespan multimorbidity research Collaborative (LINC)[§], Peter Holmans³, Inês Barroso⁵, Hilary Martin⁶, Marianne B. M. van den Bree^{3,4,‡}, Sarah Finer^{1,‡}, Nicholas J. Timpson^{2,‡}

¹Wolfson Institute of Population Health, Queen Mary University of London, E1 2AT, United Kingdom

²MRC Integrative Epidemiology Unit, Bristol Medical School, University of Bristol, BS1 5DS, United Kingdom

³Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, CF24 4HQ, United Kingdom

⁴Mental Health Innovation Institute, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, CF14 4XN, United Kingdom

⁵Medical School, University of Exeter, Exeter, EX2 5DW, United Kingdom

⁶Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, CB10 1SA, United Kingdom

*Corresponding author: Wolfson Institute of Population Health, Queen Mary University of London, 3rd Floor Abernethy Building, 2 Newark St, London E1 2AT, United Kingdom. E-mail: d.stow@qmul.ac.uk

[§]A list of authors and their affiliations appears in the acknowledgements section

[†]The authors wish it to be known that, in their opinion, the first 3 authors should be regarded as joint First Authors

[‡]The authors wish it to be known that, in their opinion, the last 3 authors should be regarded as joint Senior Authors

Abstract

Background: Multimorbidity, also known as multiple long-term conditions, is a major public health concern. Internalising and CardioMetabolic MultiMorbidity (ICM-MM) is a common form of mental-physical health multimorbidity, yet its genetic predisposition is largely unknown. We examined the polygenic nature of ICM-MM by assessing single trait-specific polygenic risk scores (PRS_{TRAIT}) and whether combining them could increase the proportion of variance in liability to ICM-MM explained by genetic variation.

Methods: We developed PRS_{TRAIT} using PRS-CS and summary statistics from the largest trait-specific GWAS excluding UK Biobank (UKB). We evaluated PRS_{TRAIT} on ICM-MM risk in 206 452 UKB participants ($n = 39\,311$ (19.0%) with ICM-MM) using logistic regression adjusted for gender and 10 genetic principal components, defining ICM-MM as lifetime occurrence of: ≥ 1 internalising (depression, anxiety, somatoform disorder) traits AND ≥ 1 cardiometabolic traits (type 2 diabetes, obesity, hypertension, dyslipidemia, chronic kidney disease). We used elastic net regression in a 50% training sample to generate ICM-MM-PRS_{TRAIT}: a weighted combination of PRS_{TRAIT} targeting ICM-MM.

Results: The strongest associations were between ICM-MM and PRS_{TRAIT} for depression and type 2 diabetes—both odds ratios (OR) 1.18, [95% confidence interval (CI) 1.17–1.20] per standard deviation increase in PRS_{TRAIT}. ICM-MM-PRS_{TRAIT} retained five PRS_{TRAIT}, with stronger associations (OR = 1.31, [95%CI 1.29–1.34]) than any PRS_{TRAIT} in the testing sample.

Discussion: Combining several PRS explains more variance in ICM-MM liability than single-trait PRSs alone. ICM-MM-PRS_{TRAIT} is a measure of genetic risk that could be used to examine premorbid stages of ICM-MM in external and youth cohorts, supporting awareness of earlier presentation and potentially avoidance or intervention.

Keywords Multimorbidity, MLTC, Polygenic risk scores, Genetic liability, UK Biobank

Introduction

Multimorbidity—sometimes referred to as multiple long-term conditions—the co-occurrence [1, 2] or lifetime occurrence [3] of two or more long-term health conditions, is a major public health concern, affecting over a third of the global population in mid-life [4, 5]. Multimorbidity prevalence is rising, [6, 7] leading to increased personal, [8, 9] societal, and healthcare costs [10, 11].

Many operationalisations of multimorbidity have been used in existing research [2, 10]. Here, we use a collaborative-specific operationalisation: The lifetime co-occurrence of internalising (depression, anxiety) and cardiometabolic conditions (hypertension, obesity, type 2 diabetes). This is a common form of mental and physical healthy multimorbidity, [12–15] and the interaction between mental health conditions and physical multimorbidity increases unplanned [16] and

Received: October 9, 2025. **Revised:** February 3, 2026. **Accepted:** February 10, 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

emergency department [17] hospital admissions, accounting for a substantive proportion of total secondary care costs in England [18].

Evidence for intervention and management of multimorbidity is growing [19, 20] but little is known about prevention or early life manifestations of multimorbidity [21]. Preventing or delaying onset of multimorbidity would have major benefits for affected individuals, as well as for health care provision and costs [22]. Genetic variation has been implicated in the development of internalising and cardiometabolic disorders, including rare and common factors [23–31]. These common, complex disorders have a polygenic architecture and an individual’s susceptibility to disease can partly be captured by polygenic risk scores. Polygenic risk scores (PRSs) aggregate the effect sizes of multiple single nucleotide polymorphisms (SNPs) from trait-relevant genome wide association study (GWAS) data, to create an individual-level measure of genetic risk for the trait [32]. Polygenic risk scores can be used to identify subgroups with increased risk of disease and predict case status in both case–control and population-based cohort studies [33].

Recent studies have shown that clinical manifestations of polygenic risk emerge early in life, for example, significant gradients of polygenic risk emerged in early childhood for obesity and cardiometabolic traits [34, 35]. Understanding how genetic risk of ICM-MM manifests in early life could support the development of tailored early intervention and prevention strategies to delay or prevent multimorbidity in later life, but the evidence on polygenic risk scores to support this undertaking are lacking.

The aims of this study were to generate a polygenic risk score for ICM-MM by i) evaluating associations between single trait-specific PRSs and lifetime ICM-MM risk, and ii) examining whether a combination of existing scores could increase the proportion of variance in liability to lifetime ICM-MM explained by genetic variation.

Results
Participants

Figure 1 shows the number of participants at each stage of the analysis. We included data from 206 452 eligible UKB participants (n = 111 953 [54.2%] female, mean age of recruitment 66.3 [sd = 8.0] years) with primary care electronic healthcare record data and genotype data passing QC. Table 1 shows the characteristics or participants included in the study. For condition specific prevalences and comparisons to the wider UKB cohort, please see Supplemental Table 1. For counts of condition combinations in participants with ICM-MM, please see Supplemental Fig. 1.

Identifying trait specific PRSs

Our searches identified no eligible pre-existing PRS for ICM-MM conditions. Characteristics of eligible GWAS for depression, [31] anxiety, [36] T2D, [37] CKD, [38] BMI, [39] LDL-Cholesterol, [40] (used here as a proxy for dyslipidaemia) and diastolic/systolic/pulse pressure [41] (used here as a proxy for hypertension) are in Supplemental Table 2. We could not identify any eligible GWAS for somatoform disorders.

Associations between trait-specific PRS and ICM-MM

Table 2 shows odds ratios (OR) for each PRS_TRAIT ICM-MM. The strongest associations with ICM-MM were for PRSs for type 2

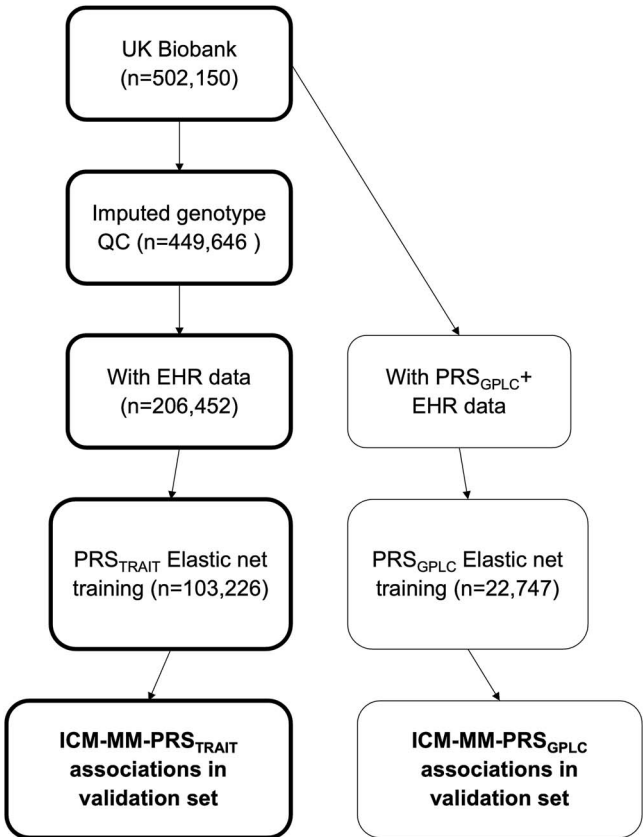


Figure 1 Flowchart of participants at each stage of analysis.

Table 1 Characteristics of participants included in the study.

	Participants meeting inclusion criteria
	(n = 206 452)
Gender	
Female	111 953 (54.2%)
Male	94 499 (45.8%)
Year of Birth	
Median [Q1,Q3]	1950 [1950,1960]
Townsend deprivation score	
Median [Q1,Q3]	−2.16 [−3.65,0.460]
Missing	313 (0.2%)
Ethnicity	
Asian or Asian British	4189 (2.0%)
Black or Black British	2181 (1.1%)
Chinese	550 (0.3%)
Do not know	79 (0.0%)
Mixed	1057 (0.5%)
Other ethnic group	1550 (0.8%)
Prefer not to answer	613 (0.3%)
White	196 077 (95.0%)
Missing	156 (0.1%)
Any internalising	
Prevalence	64 246 (31.1%)
Age onset (Median [Q1,Q3])	49.0 [40.0,58.0]
Any cardiometabolic	
Prevalence	113 297 (54.9%)
Age onset (Median [Q1,Q3])	58.0 [51.0,64.0]
ICM-MM	
Prevalence	39 311 (19.0%)
Age onset (Median [Q1,Q3])	59.0 [52.0,66.0]

Table 2 The association between LINC ICM-MM trait-specific PRSs and ICM-MM.

Original PRS _{TRAIT} target GWAS	AUC-PRS _{TRAIT} (95%CI)	Nagelkerke R ²	OR-PRS _{TRAIT} (95%CI)	P value
Depression	0.56 (0.56–0.56)	0.011	1.18 (1.17–1.20)	3.56x10 ⁻¹⁵³
T2D	0.56 (0.55–0.56)	0.010	1.18 (1.17–1.20)	2.18x10 ⁻¹¹⁷
BMI	0.55 (0.55–0.56)	0.009	1.13 (1.11–1.14)	4.47x10 ⁻⁷⁶
Systolic BP	0.55 (0.55–0.56)	0.009	1.14 (1.12–1.16)	3.38x10 ⁻⁷³
Diastolic BP	0.55 (0.55–0.56)	0.009	1.14 (1.13–1.16)	2.41x10 ⁻⁷⁵
Pulse pressure	0.55 (0.55–0.56)	0.009	1.14 (1.13–1.16)	8.78x10 ⁻⁷⁵
LDL-C	0.55 (0.54–0.55)	0.007	1.06 (1.05–1.07)	7.44x10 ⁻²⁵
Anxiety	0.55 (0.55–0.55)	0.007	1.11 (1.09–1.13)	1.05x10 ⁻³⁹
CKD	0.54 (0.54–0.55)	0.006	1.02 (1.01–1.04)	0.001

Odds ratios (OR) and AUC for each PRS_{TRAIT} from logistic regression adjusted for self-reported gender and 10 genetic PCs in 206 452 UKB participants. Nagelkerke R² for model fit, P value for PRS coefficient. PRS (polygenic risk score), AUC (area under the curve), CI (confidence interval), OR (odds ratio), T2D (type 2 diabetes), BMI (body mass index), BP (blood pressure), LDL-C (low density lipoprotein-C), CKD (chronic kidney disease)

diabetes (OR = 1.18, 95%CI:1.17–1.20) and major depression (OR = 1.18, 95%CI:1.17–1.20).

Supplemental Fig. 2 has metrics for the internalising and cardiometabolic conditions. Overall, the strongest associations were in the cardiometabolic cluster of conditions (e.g. ORs all > 1.3 for PRS_{TRAIT} for blood pressure measures vs cardiometabolic conditions) and strong associations in the cardiometabolic cluster of conditions attenuated in the ICM-MM models (e.g. association of PRS_{TRAIT} for T2D with cardiometabolic conditions was stronger than for the association PRS_{TRAIT} for T2D with ICM-MM).

Associations between ICM-MM specific PRS and ICM-MM

In the testing dataset ($n = 103\,226$), the OR for ICM-MM-PRS_{TRAIT} was 1.31 (95%CI:1.29–1.34) and the AUC was 0.57 (95%CI 0.57–0.58)—higher than the strongest PRS_{TRAIT}—ICM-MM association, which was for major depression OR = 1.19 (95%CI:1.16–1.21), AUC = 0.56 (95%CI:0.55–0.56). Details of the five retained PRS_{TRAIT}, and optimized weights are shown in Supplemental Fig. 3. The ICM-MM-PRS_{TRAIT} model had a significantly higher AUC than the model adjusted for major depression (DeLong's test p -value < 2.2e-16), and both scores improved predicted performance versus the null model (adjusted for gender and 10 genetic PCs only, AUC = 0.53, 95%CI = 0.53–0.54).

Supplemental analyses

Genomics PLC 'enhanced' polygenic risk scores (PGS_{GPLIC}) were available for $n = 45\,493$ UKB participants with primary care data. Correlations between PGS_{TRAIT} and PGS_{GPLIC} are shown in Supplemental Fig. 4—The highest correlation for PGS_{TRAIT} and PGS_{GPLIC} was for LDL-cholesterol ($r = 0.55$).

PGS_{GPLIC} for hypertension (OR = 1.25 95%CI:1.22–1.28) and ischaemic stroke (OR = 1.25, 95%CI:1.22–1.28) were most strongly associated with ICM-MM, and had the largest AUC (0.58, 95%CI: 0.57–0.59, and 0.58, 95%CI:0.57–0.58, respectively). For ORs and AUC for all PGS_{GPLIC} please see Supplemental Figs. 5 and 6.

Elastic net regression retained 13 of 51 PGS_{GPLIC}. Details of the retained PRS, and optimized weights for the linear combination in the $n = 22\,747$ testing set are shown in Supplemental Fig. 7. OR for ICM-MM-PGS_{GPLIC} was 1.32 (95%CI:1.28–1.37) and AUC was 0.59 (95%CI 0.58–0.60). The best performing single PGS_{GPLIC} in the testing set ($n = 22\,747$) was for ischaemic stroke (OR = 1.24, 95%CI: 1.20–1.28, AUC = 0.57, 95%CI:0.56–0.58). Both scores showed improved predicted

performance versus the null model (adjusted for gender and 10 genetic PCs only, AUC = 0.54, 95%CI = 0.53–0.55).

In this subset of data, Elastic net regression retained four of seven PRS_{TRAIT} (Supplemental Fig. 8) and OR for ICM-MM-PRS_{TRAIT} in this subset was 1.27 (95%CI:1.21–1.34) and AUC was 0.56 (95%CI:0.55–0.57). DeLong's test demonstrated that the models adjusted for ICM-MM-PRS_{GPLIC} and PRS_{GPLIC} for ischaemic stroke had significantly higher AUC than for ICM-MM-PRS_{TRAIT} ($P = 8.137\text{e-}09$ and $P = 0.027$ respectively).

Discussion

In this study, we created a novel PRS for ICM-MM, a common form of physical-mental health multimorbidity. Our work demonstrates that combining multiple existing PRSs yields stronger associations with lifetime ICM-MM risk than single trait-specific PRSs. Multimorbidity is a global public health concern but work on early-life manifestations of multimorbidity that could support identification and prevention are lacking. We will now use the novel ICM-MM-PRS in younger adult and youth cohorts [42] to understand how genetic risk of ICM-MM manifests itself in mid-and-early life: improving our understanding of condition ordering and the aetiology of this multimorbidity cluster, potentially identifying precursor signals (such as early-life trajectories of mental health or biomarkers related to cardiometabolic conditions), and highlighting early intervention targets.

The literature on multimorbidity is expanding rapidly, but there is a high degree of heterogeneity in phenotypes under study. This is due, in part, to the canonical definition of 'two or more long-term conditions', and study specific decisions about which long-term conditions to include [1, 2]. Many pairs of common chronic conditions demonstrate evidence of shared genetic risk [43], but our ICM-MM phenotype and findings are most similar to those of Baltramonaityte et al [44], who developed a PRS for a multimorbidity phenotype comprising coronary artery disease, T2D and depression. This PRS had a higher OR (1.91, 95%CI = 1.74–2.10) than ICM-MM-PGSs for a similar, but more restricted ($n = 3$ conditions) phenotype, and in a model that was additionally adjusted for age. Zhao et al. [45], also created a PRS for cardiometabolic multimorbidity (two or more of T2D, coronary heart disease and stroke) in UKB with an AUC of 0.62. This was higher than the AUC we found for the ICM-MM-PRSs on ICM-MM, but for a restricted phenotype ($n = 3$ conditions) with no internalising conditions and included age as a component of the model. Other studies that have developed PRS for multimorbidity phenotypes, did

not report any comparable metrics (in the case of Gezi et al [46]), or furthermore explicitly excluded some of the common cardiometabolic conditions we included in our ICM-MM definition (in the case of Zhang et al [47]).

Taken together with our findings, the available evidence suggests there is some utility to the use of polygenic risk scores as tools to explore how multimorbidity manifests itself throughout the life-course: across a range statistical metrics and similar definitions of multimorbidity, associations between PRSs and multimorbidity phenotypes are modest but persistent. The upper bound of the prediction accuracy of ICM-MM by PRSs likely to be driven by (SNP) heritability and frequency (as for individual traits), rather than the choice of method used to derive or combine the PRSs.

We highlight that caution is required around communication of the utility of polygenic risk scores: PRSs alone are not yet suitable for individual risk prediction or population risk stratification in public health, or clinical settings [48], and much of the risk of disease is determined by factors not captured by the genome [49]. Recent reviews [50, 51] have highlighted the importance of ‘place’ based factors including area-level deprivation and air quality as risk factors for a range of multimorbidity phenotypes, and there is emerging evidence of the role of early-life environmental exposures (including parental and family environment, and education) in the development of midlife hypertension-obesity multimorbidity [52]. Further work is required to elucidate the potential for gene x environmental interactions to influence downstream multimorbidity risk.

Limitations

We used data from UK Biobank because it is a population-based dataset with known higher prevalence of ICM-MM conditions, and rich phenotyping and genetic data. However UKB is subject to volunteer bias, low response rate, and is not representative of the UK population, especially with respect to age, socioeconomic status, or ancestry [53, 54]—all factors associated with multimorbidity risk, limiting the generalisability of our findings. UKB also contributed to many recent meta-GWAS that would have been ideal sources of summary statistics to construct PRS_{TRAIT}. Some but not all studies made available ‘leave out’ summary statistics, excluding UKB, meaning most of the input PRS_{TRAIT} are based on older studies with fewer participants (i.e. weaker instruments). To test the potential of stronger ‘on trait,’ as well as ‘cross trait’ input PRSs, we also carried out a supplemental analysis using 51 PRS created by Genomics PLC: PRS_{GPLC}. We found that ICM-MM-PRS_{GPLC} had a stronger association with ICM-MM (OR = 1.32, 95%CI:1.28–1.37, AUC = 0.59, 95%CI 0.58–0.60) than single PRS_{GPLC} for ischaemic stroke (the strongest predictor in this testing subset, AUC 0.57, 95%CI 0.56–.58) and ICM-MM-PRS_{TRAIT} (p for difference $P = 8.13 \times 10^{-09}$). One of the key limitations of PRS_{GPLC} in our analysis is the lack of PRSs for internalising traits, depression in particular—one of the PRSs most strongly associated with ICM-MM in our PRS_{TRAIT} analysis. We kept PRS_{TRAIT} and PRS_{GPLC} analyses separate because PRS_{GPLC} are only available ‘pre calculated’ in selected datasets and the exact methods required to recreate them are the intellectual property of Genomics PLC.

We used a ‘lifetime occurrence’ based approach to operationalising multimorbidity. This approach facilitates mechanistic understandings of disease co-occurrence over the whole life course [3], and was also a pragmatic solution to the problem of defining co-occurrence with episodic chronic conditions (such as depression and anxiety)

where ‘recovery’ is not well captured in electronic health records. This life-course based definition means some of the prevalence estimates we report are higher than other studies using a ‘co-occurrence’ definition.

By including obesity in the cardiometabolic cluster of conditions we recognise its status as a complex, chronic disease [55]. However, this means we were unable to examine its well-established role as a risk factor for many non-communicable disease, nor were we able to explore the mediating effects of BMI on causal relationships between disease pairings, such as in depression and T2D [56].

Somatoform disorders were included in our ICM-MM definition due to high genetic [57], diagnostic [58, 59], and symptomatic overlap with depression and anxiety. Our searches did not identify suitable GWAS for the somatoform disorders phenotype that is part of the ICM-MM definition. However, given the strong genetic correlations between somatic disorders and depression and anxiety in particular [57], it is likely that any potential somatoform PRS_{TRAIT} would likely not be retained in the elastic net regression, as was the case for PRS_{TRAIT} for anxiety.

The conditions that make up the ICM-MM definition are diagnosed and managed primarily in primary care settings in the UK. We restricted our cohort to individuals with available linked primary care data (supplemented with Hospital Episode Statistics data), which reduced our sample size, but enhanced cross-cohort reproducibility and the quality of phenotype ascertainment [60]. The limitations of routinely collected electronic healthcare records are well reported (see Farmer et al for an example) [61] and our use of EHR data may have led to incomplete capture of some ICM-MM conditions. Existing comparisons between UKB EHR data and data from the Secure Anonymised Information Linkage (SAIL) databank (a nationally representative sample of primary care data in Wales) suggest that mental health conditions may be underrepresented in UKB [62]. Our use of primary care data broadly retained the characteristics of the UKB sample. However, there was a reduction in the proportion of participants who identified as Black or Black British ethnicity in the sample (Supplemental Table 1).

PRSmix uses a training/test split to minimise data leakage. However, because the genetic PCs we used to correct for stratification in the model were calculated for the whole cohort simultaneously [63], there is the potential for some data leakage as information from the PCs is shared across the training and test datasets. The Elastic Net regression implemented in PRSmix results in a weighted linear combination of PRS. Since multimorbidity requires a combination of two or more traits to be present, it is possible that modelling non-linear combinations of PRS might improve prediction of MM. Future work could explore other approaches to combining multiple single trait PRSs [64, 65] or use of machine learning methods to produce non-linear combinations of input PRSs, as well as exploring specific pairwise combinations of ICM-MM conditions.

Conclusion

Combining multiple polygenic risk scores improves the proportion of variance in liability to ICM-MM explained by genetic variation over single trait specific scores. We have generated a novel polygenic risk score for ICM-MM that can be used to explore manifestations of ICM-MM in individuals at higher genetic risk in external and early-life cohorts. Future work should investigate early-life manifestations of high genetic risk of ICM-MM, as well as environmental risk factors

and gene x environment interactions to potentially improve early life recognition and interventions for ICM-MM.

Materials and methods

This work was conducted as part of the Lifespan Multimorbidity Research Collaborative (LINC) [42], which studies development of internalising and cardiometabolic multimorbidity (ICM-MM) across the lifespan in five longitudinal population-based cohorts with genetic data.

Patient and public involvement

Members of the public with lived experiences of multiple long-term conditions have been involved throughout the course of the LINC study. For this study they assisted primary and secondary care doctors, and other members of the research team to select the conditions that make up the ICM-MM phenotype (see below ‘phenotype definition’) and informed the research questions and aims.

Ethics

UK Biobank has approval from the North West Multi-centre Research Ethics Committee (reference 16/NW/0274) as a Research Tissue Bank. Participants provided electronic signed consent at recruitment.

Data source

UK Biobank (UKB) is a cohort comprising over 500 000 people living in the United Kingdom. Participants aged 40 to 69 years old were recruited between 2006 and 2010. Participants attended a baseline assessment and were also invited to attend repeat assessments over the following ten years. Our study was conducted under application number 79704.

Participants

Participant recruitment processes and eligibility criteria for UKB are described in detail in Bycroft et al., 2018 [63]. We restricted our analysis to the subset of individuals with primary care electronic healthcare record (EHR) data (~230 000, or 45% of the UKB cohort) [66].

Phenotype definition

LINC [42] investigates a study-specific form of multimorbidity: internalizing and cardiometabolic multimorbidity (ICM-MM), defined as the lifetime occurrence of at least one internalising condition (anxiety, depression, somatoform disorder) AND at least one cardiometabolic condition (chronic kidney disease, dyslipidemia, hypertension, obesity, type 2 diabetes). The ICM-MM phenotype was developed by primary and secondary care clinicians in partnership with members of the public with lived experiences of long term physical and mental health conditions. The cardiometabolic conditions were selected because they occur relatively early on the pathways to cardiovascular disease with divergence in risk starting early in life [67–69], rendering them suitable for study across the lifespan.

In this study, we used UKB electronic health record data (primary care and NHS Digital Hospital Episode Statistics[HES]) to identify participants with ICM-MM using existing Read code (primary care) and ICD-10 (secondary care—HES) codelist resources [70, 71]

(See supplemental codelist information and data files). The presence of any one (or more) code(s) was taken as a diagnosis of the condition.

PRS construction

Constructing trait specific PRSs

Figure 2 summarises the methods to identify the polygenic risk scores and test their associations with ICM-MM. We searched existing literature and the PGS catalogue [72, 73] to identify PRS for the eight ICM-MM traits: ‘PRS_{TRAIT}’. To avoid overfitting in our UKB analysis, we restricted our searches to PRS constructed using non-UKB data.

If we could not find an existing PRS for ICM-MM conditions meeting these criteria, we searched for GWAS summary statistics from non-UKB studies, or meta-GWAS where authors reported or made available summary statistics excluding UKB. If our searches identified multiple eligible PRS or GWAS for an ICM-MM condition, we selected the one with the largest number of participants. Once appropriate GWAS summary statistics had been obtained, we used PRS-CS [74] and the European ancestry 1000-genomes external linkage disequilibrium (LD) reference panel to calculate PRS. PRS-CS uses a Bayesian algorithm to infer posterior SNP effect sizes via continuous shrinkage (CS), avoiding the reduction in effective sample size for downstream analysis that arises from linkage disequilibrium pruning and p-value thresholding methods [74]. We used PLINK 2.0 [75, 76] to calculate per-individual PRS_{TRAIT} using the inferred posterior SNP effect sizes and the UKB genotype/array data.

UKB imputed genotype quality control

Imputed genotype array data were used for PRS construction—genotypes were imputed using the Haplotype Reference Consortium (HRC) and UK10K haplotype resource. The imputed genotypes are aligned to the + strand of the reference and the positions are in GRCh37 coordinates. Further details of the imputation and standard quality-control performed on UKB’s genotype data are available in Bycroft et al, 2018 [63]. We further removed variants with a low INFO score (<0.9), high missingness (>0.05), low minor allele frequency (<0.01) or variants departing from Hardy–Weinberg equilibrium ($P < 10^{-6}$), and individuals with high missingness (>0.05) or sex discordance using PLINK 2.0. Sex chromosomes were excluded. We used Kinship-based Inference for GWAS (KING) [77] to identify individuals related to the second-degree ($r^2 > 0.0884$), and removed one individual from each related pair at random. After quality control, 449 646 individuals and 6 899 626 variants were retained.

Constructing an ICM-MM specific PRS

We used elastic net regression, implemented in PRSmix [78] to produce a linear, weighted combination of ICM-MM trait-specific PRS_{TRAIT} (ICM-MM-PRS_{TRAIT}) that best predicted ICM-MM. PRSmix methods are described in detail in Truong et al., 2024 [78]. Briefly, PRSmix employs elastic net regression on a random subset of training data (specified here as 50%) to identify the optimal, weighted linear combination of input PRSs for predicting the presence of the target trait. The input model was additionally adjusted for self-reported gender and 10 genetic principal components (available from UKB data field 22 009). PRSmix produces information on predictive performance (odds ratios [OR] and area under the curve [AUC]) for the new, weighted combination of PRS in the remaining 50% testing set, compared to i) the ‘null’ input model (adjusted for gender, and 10 genetic principal

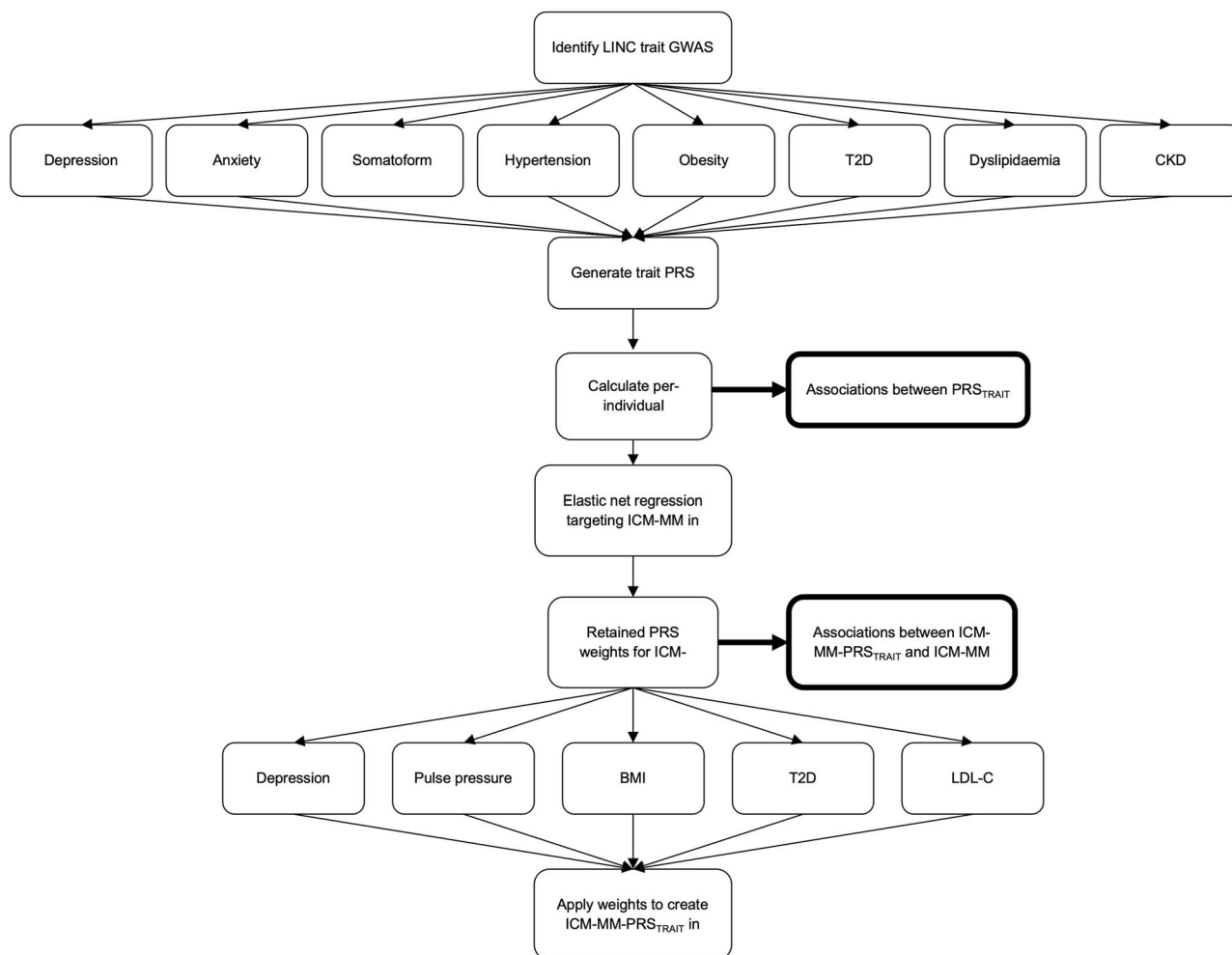


Figure 2 Methods flowchart for generating and testing the PRSs. GWAS(genome wide association study), BMI (body mass index, a proxy for obesity), T2D (type 2 diabetes), LCL-C (LDL-cholesterol, a proxy for dyslipidaemia), CKD (chronic kidney disease), PRS (polygenic risk score), ICM-MM (internalising and CardioMetabolic MultiMorbidity).

components only), and ii) the best performing *single* PRS. All PRSs were standardised (z-transformation) before running PRSmix.

Statistical analysis

We used logistic regression to examine the association between each PRS_{TRAIT} , lifetime internalising or cardiometabolic traits and ICM-MM. All models were adjusted for self-reported gender and 10 genetic principal components to account for population stratification and ancestry (available from UKB data field 22 009). We also report area under the curve (AUC) from receiver operating characteristic curves for these logistic models to provide estimates of discrimination between cases and controls. We corrected these analyses for multiple testing via Bonferroni correction. All analyses were performed using R Statistical Software [79] on the UKB Research Access Platform. For ROC analyses we used the pROC package [80].

Supplemental analysis

Given the genetic correlation between the conditions used by LINC to define ICM-MM (see above ‘phenotype definition’) and other cardiometabolic conditions such as cardiovascular disease and ischaemic stroke; and to further test the potential to combine existing polygenic

risk scores to predict ICM-MM, we carried out a supplemental analysis using polygenic risk scores calculated by Genomics PLC (PRS_{GPLC})—available under category 300 in the UKB data showcase. SNP weights are not publicly available, nor is it possible to identify the specific GWAS used to derive each PRS, but the ‘enhanced’ PRS_{GPLC} have been calculated directly for a subset of participants UKB and in select additional external cohorts. PRS_{GPLC} have been constructed for a range of traits, including many of the ICM-MM cardiometabolic traits, but not for any of the internalising ICM-MM traits (see Thompson et al [81] for further details).

More detailed methods for how the Genomics PLC PRSs were derived may be found in Thompson et al, 2022 [82], and Thompson et al., 2024 [81]. Briefly: Genomics PLC used existing GWAS and proprietary methods to generate two sets of PRSs for a range of binary and quantitative traits in UKB. The ‘standard’ PRS set was generated for 39 binary and quantitative and binary traits for all 500 K UKB participants. The ‘standard’ set uses per-trait meta-analysis GWAS from studies *excluding* UKB (i.e. trained externally). An ‘enhanced’ set of PRSs was also generated for a mixed-ancestry UKB subgroup comprising $n \sim 120$ k participants, using scores from the original per-trait meta GWAS (used to develop the ‘standard’ scores), meta analysed with GWAS from the ‘white British unrelated’ subset of UKB

($n \sim 337$ k) for 51 binary and quantitative traits (i.e. external training data + UKB subset training data). Genomics PLC have calculated PRSs for participants in external cohorts using the weights derived from the ‘enhanced’ set. If the external cohort contributed information to the ‘standard’ score meta-GWAS, Genomics-PLC re-ran their PRS generation pipeline, excluding the cohort from the meta-GWAS.

We calculated the correlations between each individual PRS_{TRAIT} and ‘enhanced’ PRS_{GPLC} to assess how similar these were for shared target traits, as well as to explore correlations across different traits. We then repeated the logistic regression, ROC, and PRSmix analyses outlined previously using PRS_{GPLC} to create ICM-MM-PRS_{GPLC}. We also repeated the PRSMix analysis with PRS_{TRAIT} as inputs using the same training and testing subset used to create ICM-MM-PRS_{GPLC}, and compared the AUC for ICM-MM-PRS_{TRAIT} and ICM-MM-PRS_{GPLC} using DeLong’s test [83]. All models were adjusted for gender and 10 genetic principal components.

Acknowledgements

Full list of LINC members: Marianne B. M. van den Bree, George Kirov, Michael J. Owen, James T. R. Walters, Peter A. Holmans, Jane Lynch, Ioanna K. Katzourou, Lauren Bengler, David Owen, Lowri O’Donovan (Cardiff University, UK). David A. van Heel, Sarah Finer, Daniel Stow (Queen Mary University of London, UK). Golam M. Khan-daker, Nicholas J. Timpson, John A. A. MacLeod, Julie P. Clayton, Ruby S. M. Tsang, Jane Sprackman, Shahid Khan (University of Bristol, UK). Inês Barroso, Rupert A. Payne (University of Exeter, UK). Mark Mon-Williams, Megan L. Wood, Nabila N Ali (University of Leeds, UK). Hilary C. Martin (Wellcome Sanger Institute, UK). Morteza Vaez, Lam Opal Huang, Thomas Werge, Andrés Ingason (Institute of Biological Psychiatry, Denmark). For further details of the LINC project, please visit our website: <https://www.cardiff.ac.uk/lifespan-multimorbidity-research-collaborative>

Author contributions

Daniel Stow (Conceptualization, Data curation, Formal analysis, Investigation, Validation, Visualization, Writing—original draft, Writing—review & editing), Ruby Tsang (Conceptualization, Data curation, Formal analysis, Investigation, Validation, Writing—original draft, Writing—review & editing), Ioanna Katzourou (Conceptualization, Data curation, Formal analysis, Validation, Writing—original draft, Writing—review & editing), Jack Underwood (Investigation, Writing—review & editing), Peter Holmans (Conceptualization, Formal analysis, Methodology, Supervision, Writing—review & editing), Ines Barroso (Conceptualization, Funding acquisition, Supervision, Writing—review & editing), Hilary Martin (Conceptualization, Funding acquisition, Methodology, Supervision, Writing—review & editing), Marianne van den Bree (Conceptualization, Funding acquisition, Supervision, Writing—review & editing), Sarah Finer (Conceptualization, Funding acquisition, Supervision, Writing—review & editing), Nicholas Timpson (Conceptualization, Funding acquisition, Supervision, Writing—review & editing).

Supplementary material

Supplementary Material is available at *Human Molecular Genetics* online.

Conflict of interest statement: All named authors report no conflicts of interest.

Funding

This work was supported by the Tackling Multimorbidity at Scale Strategic Priorities Fund programme [grant number MR/W014416/1] delivered by the Medical Research Council and the National Institute for Health Research in partnership with the Economic and Social Research Council and in collaboration with the Engineering and Physical Sciences Research Council. JU is funded by a Wellcome Trust GW4-CAT Clinical Doctoral Fellowship (222849/Z/21/Z).

References

1. Barnett K, Mercer SW, Norbury M. *et al.* Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet*. 2012;**380**:37–43. [https://doi.org/10.1016/S0140-6736\(12\)60240-2](https://doi.org/10.1016/S0140-6736(12)60240-2).
2. Ho ISS, Azcoaga-Lorenzo A, Akbari A. *et al.* Measuring multimorbidity in research: Delphi consensus study. *BMJ Med* 2022;**1**:e000247. <https://doi.org/10.1136/bmjmed-2022-000247>.
3. Langenberg C, Hingorani AD, Whitty CJM. Biological and functional multimorbidity—from mechanisms to management. *Nat Med* 2023;**29**:1649–1657.
4. Chowdhury SR, Chandra Das D, Sunna TC. *et al.* Global and regional prevalence of multimorbidity in the adult population in community settings: a systematic review and meta-analysis. *EClinicalMedicine*. 2023;**57**:101860.
5. Nguyen H, Manolova G, Daskalopoulou C. *et al.* Prevalence of multimorbidity in community settings: a systematic review and meta-analysis of observational studies. *J Comorb* 2019;**9**:2235042X19870934. <https://doi.org/10.1177/2235042X19870934>.
6. Garin N, Koyanagi A, Chatterji S. *et al.* Global multimorbidity patterns: a cross-sectional, population-based, multi-country study. *J Gerontol A Biol Sci Med Sci* 2016;**71**:205–214.
7. Kingston A, Robinson L, Booth H. *et al.* Projections of multimorbidity in the older population in England to 2035: estimates from the population ageing and care simulation (PACSim) model. *Age Ageing* 2018;**47**:374–380. <https://doi.org/10.1093/ageing/afx201>.
8. Duguay C, Gallagher F, Fortin M. The experience of adults with multimorbidity: a qualitative study. *J Comorb* 2014;**4**:11–21.
9. Tran PB, Ali A, Ayesha R. *et al.* An interpretative phenomenological analysis of the lived experience of people with multimorbidity in low- and middle-income countries. *BMJ glob. Health*. 2024;**9**:e013606. <https://doi.org/10.1136/bmjgh-2023-013606>.
10. The Academy of Medical Sciences. *Multimorbidity: a priority for global health research*, The Academy of Medical Sciences, London, 2018.
11. Tran PB, Kazibwe J, Nikolaidis GF. *et al.* Costs of multimorbidity: a systematic review and meta-analyses. *BMC Med* 2022;**20**:234.
12. Ferro MA. Understanding multimorbidity early in life takes a step forward. *Lancet Public Health* 2023;**8**:e662–e663. [https://doi.org/10.1016/S2468-2667\(23\)00174-3](https://doi.org/10.1016/S2468-2667(23)00174-3).
13. Kuan V, Denaxas S, Patalay P. *et al.* Identifying and visualising multimorbidity and comorbidity patterns in patients in the English National Health Service: a population-based study. *Lancet Digit Health* 2023;**5**:e16–e27. [https://doi.org/10.1016/S2589-7500\(22\)00187-X](https://doi.org/10.1016/S2589-7500(22)00187-X).
14. MacRae C, Mercer SW, Henderson D. *et al.* Age, sex, and socioeconomic differences in multimorbidity measured in four ways: UK

- primary care cross-sectional analysis. *Br J Gen Pract* 2023;**73**:e249–e256. <https://doi.org/10.3399/BJGP.2022.0405>.
15. Steell L, Krauth SJ, Ahmed S. *et al*. Multimorbidity clusters and their associations with health-related quality of life in two UK cohorts. *BMC Med* 2025;**23**:1. <https://doi.org/10.1186/s12916-024-03811-3>.
 16. Soley-Bori M, Ashworth M, Bisquera A. *et al*. Impact of multimorbidity on healthcare costs and utilisation: a systematic review of the UK literature. *Br J Gen Pract* 2021;**71**:e39–e46. <https://doi.org/10.3399/bjgp20X713897>.
 17. Gaulin M, Simard M, Candas B. *et al*. Combined impacts of multimorbidity and mental disorders on frequent emergency department visits: a retrospective cohort study in Quebec. *Canada CMAJ* 2019;**191**:E724–E732. <https://doi.org/10.1503/cmaj.181712>.
 18. Stokes J, Guthrie B, Mercer SW. *et al*. Multimorbidity combinations, costs of hospital care and potentially preventable emergency admissions in England: a cohort study. *PLoS Med* 2021;**18**:e1003514. <https://doi.org/10.1371/journal.pmed.1003514>.
 19. Smith SM, Wallace E, O'Dowd T. *et al*. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. *Cochrane Database Syst Rev* ;2021. <https://doi.org/10.1002/14651858.CD006560.pub4>.
 20. Zhou Y, Dai X, Ni Y. *et al*. Interventions and management on multimorbidity: an overview of systematic reviews. *Ageing Res Rev* 2023;**87**:101901.
 21. Skou ST, Mair FS, Fortin M. *et al*. Multimorbidity. *Nat Rev Dis Primers* 2022;**8**:48. <https://doi.org/10.1038/s41572-022-00376-4>.
 22. Head A, Fleming K, Kypridemos C. *et al*. Multimorbidity: the case for prevention. *J Epidemiol Community Health* 2021;**75**:242–244. <https://doi.org/10.1136/jech-2020-214301>.
 23. Dennis J, Sealock J, Levinson RT. *et al*. Genetic risk for major depressive disorder and loneliness in sex-specific associations with coronary artery disease. *Mol Psychiatry* 2021;**26**:4254–4264.
 24. Gillespie NA, Gentry AE, Kirkpatrick RM. *et al*. Determining the stability of genome-wide factors in BMI between ages 40 to 69 years. *PLoS Genet* 2022;**18**:e1010303. <https://doi.org/10.1371/journal.pgen.1010303>.
 25. Hagenaars SP, Coleman JRI, Choi SW. *et al*. Genetic comorbidity between major depression and cardio-metabolic traits, stratified by age at onset of major depression. *Am J Med Genet B Neuropsychiatr Genet* 2020;**183**:309–330. <https://doi.org/10.1002/ajmg.b.32807>.
 26. Katzourou IK, Consortium L, Barroso I. *et al*. Contributions of common and rare genetic variation to different measures of mood and anxiety disorder in the UK biobank. *BJPsych Open* 2025;**11**:e97. <https://doi.org/10.1192/bjo.2025.43>.
 27. Katzourou IK, Consortium L, Barroso I. *et al*. Neurodevelopmental copy number variants increase risk of internalising and cardiometabolic multimorbidity: findings from UK biobank. *medRxiv*. 2025. <https://doi.org/10.1101/2025.05.13.25326577>.
 28. So H-C, Chau CK-L, Cheng Y-Y. *et al*. Causal relationships between blood lipids and depression phenotypes: a Mendelian randomisation analysis. *Psychol Med* 2021;**51**:2357–2369.
 29. Tang B, Yuan S, Xiong Y. *et al*. Major depressive disorder and cardiometabolic diseases: a bidirectional Mendelian randomisation study. *Diabetologia*. 2020;**63**:1305–1311. <https://doi.org/10.1007/s00125-020-05131-6>.
 30. Tyrrell J, Mulugeta A, Wood AR. *et al*. Using genetics to understand the causal influence of higher BMI on depression. *Int J Epidemiol* 2019;**48**:834–848. <https://doi.org/10.1093/ije/dyy223>.
 31. Wray NR, Ripke S, Mattheisen M. *et al*. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 2018;**50**:668–681. <https://doi.org/10.1038/s41588-018-0090-3>.
 32. Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* 2020;**15**:2759–2772. <https://doi.org/10.1038/s41596-020-0353-1>.
 33. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med* 2020;**12**:44. <https://doi.org/10.1186/s13073-020-00742-5>.
 34. Khara AV, Chaffin M, Wade KH. *et al*. Polygenic prediction of weight and obesity trajectories from birth to adulthood. *Cell*. 2019;**177**:587–596.e9. <https://doi.org/10.1016/j.cell.2019.03.028>.
 35. Richardson TG, O'Nunain K, Relton CL. *et al*. Harnessing whole genome polygenic risk scores to stratify individuals based on cardiometabolic risk factors and biomarkers at age 10 in the lifecourse-brief report. *Arterioscler Thromb Vasc Biol* 2022;**42**:362–365. <https://doi.org/10.1161/ATVBAHA.121.316650>.
 36. Meier SM, Tronetti K, Purves KL. *et al*. Genetic variants associated with anxiety and stress-related disorders: a genome-wide association study and mouse-model study: a genome-wide association study and mouse-model study. *JAMA Psychiatry* 2019;**76**:924–932. <https://doi.org/10.1001/jamapsychiatry.2019.1119>.
 37. Mahajan A, Taliun D, Thurner M. *et al*. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 2018;**50**:1505–1513. <https://doi.org/10.1038/s41588-018-0241-6>.
 38. Pattaro C, Teumer A, Gorski M. *et al*. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun* 2016;**7**:10023.
 39. Locke AE, Kahali B, Berndt SI. *et al*. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;**518**:197–206. <https://doi.org/10.1038/nature14177>.
 40. Willer CJ, Schmidt EM, Sengupta S. *et al*. Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013;**45**:1274–1283. <https://doi.org/10.1038/ng.2797>.
 41. Keaton JM, Kamali Z, Xie T. *et al*. Genome-wide analysis in over 1 million individuals of European ancestry yields improved polygenic risk scores for blood pressure traits. *Nat Genet* 2024;**56**:778–791. <https://doi.org/10.1038/s41588-024-01714-w>.
 42. Linc. Lifespan Multimorbidity Research Collaborative (LINC), 2025. <https://www.cardiff.ac.uk/lifespan-multimorbidity-research-collaborative/research>.
 43. Murrin O, Mounier N, Voller B. *et al*. A systematic analysis of the contribution of genetics to multimorbidity and comparisons with primary care data. *EBioMedicine*. 2025;**113**:105584.
 44. Baltramonaityte V, Pingault J-B, Cecil CAM. *et al*. A multivariate genome-wide association study of psycho-cardiometabolic multimorbidity. *PLoS Genet* 2023;**19**:e1010508. <https://doi.org/10.1371/journal.pgen.1010508>.
 45. Zhao C, Ma T, Cheng X. *et al*. Genome-wide association study of cardiometabolic multimorbidity in the UK biobank. *Clin Genet* 2024;**106**:72–81.
 46. Gezi A, Van der Auwera S, Mäkinen H. *et al*. Unique genetic and risk-factor profiles in clusters of major depressive disorder-related multimorbidity trajectories. *Nat Commun* 2024;**15**:7190.
 47. Zhang Y, Jiang X, Mentzer AJ. *et al*. Topic modeling identifies novel genetic loci associated with multimorbidities in UK biobank. *Cell Genom* 2023;**3**:100371. <https://doi.org/10.1016/j.xgen.2023.100371>.
 48. Hingorani AD, Gratton J, Finan C. *et al*. Performance of polygenic risk scores in screening, prediction, and risk stratification: secondary analysis of data in the polygenic score Catalog. *BMJ Med* 2023;**2**:e000554. <https://doi.org/10.1136/bmjmed-2023-000554>.

49. Sud A, Horton RH, Hingorani AD. *et al.* Realistic expectations are key to realising the benefits of polygenic scores. *BMJ*. 2023;**380**:e073149. <https://doi.org/10.1136/bmj-2022-073149>.
50. Ingram E, Ledden S, Beardson S. *et al.* Household and area-level social determinants of multimorbidity: a systematic review. *J Epidemiol Community Health* 2021;**75**:232–241. <https://doi.org/10.1136/jech-2020-214691>.
51. Zheng C, MacRae C, Rowley-Abel L. *et al.* The impact of place on multimorbidity: a systematic scoping review. *Soc Sci Med* 2024;**361**:117379.
52. Stannard S, Owen RK, Berrington A. *et al.* Early life domains as predictors of obesity and hypertension comorbidity: findings from the 1970 British cohort study (BCS70). medRxiv. 2024. <https://doi.org/10.1101/2024.05.13.24307277>
53. Schoeler T, Speed D, Porcu E. *et al.* Participation bias in the UK biobank distorts genetic associations and downstream analyses. *Nat Hum Behav* 2023;**7**:1216–1227. <https://doi.org/10.1038/s41562-023-01579-9>.
54. van Alten S, Domingue BW, Faul J. *et al.* Reweighting UK biobank corrects for pervasive selection bias due to volunteering. *Int J Epidemiol* 2024;**53**. <https://doi.org/10.1093/ije/dyae054>.
55. Ralston J, Brinsden H, Buse K. *et al.* Time for a new obesity narrative. *Lancet* 2018;**392**:1384–1386. [https://doi.org/10.1016/S0140-6736\(18\)32537-6](https://doi.org/10.1016/S0140-6736(18)32537-6).
56. Bala R, Handley D, Gillett A. *et al.* Evidence of bidirectional relationship between type 2 diabetes and depression; a Mendelian randomization study. *Mol Psychiatry* 2025;**30**:5013–5023.
57. Fominykh V, Jaholkowski P, Shadrin AA. *et al.* Genome-wide study of somatic symptom and related disorders identifies novel genomic loci and map genetic architecture. medRxiv. 2025. <https://doi.org/10.1101/2025.07.16.25331639>.
58. Eto F, Samuel M, Henkin R. *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 Med* 2023;**20**:e1004300. <https://doi.org/10.1371/journal.pmed.1004300>.
59. Tesfaye M, Jaholkowski P, Shadrin AA. *et al.* Identification of novel genomic loci for anxiety symptoms and extensive genetic overlap with psychiatric disorders. *Psychiatry Clin Neurosci* 2024;**78**:783–791.
60. Clifton L, Liu X, Collister JA. *et al.* Assessing the importance of primary care diagnoses in the UK biobank. *Eur J Epidemiol* 2024;**39**:219–229. <https://doi.org/10.1007/s10654-023-01095-0>.
61. Farmer R, Mathur R, Bhaskaran K. *et al.* Promises and pitfalls of electronic health record analysis. *Diabetologia*. 2018;**61**:1241–1248. <https://doi.org/10.1007/s00125-017-4518-6>.
62. Hanlon P, Jani BD, Nicholl B. *et al.* Associations between multimorbidity and adverse health outcomes in UK biobank and the SAIL databank: a comparison of longitudinal cohort studies. *PLoS Med* 2022;**19**:e1003931. <https://doi.org/10.1371/journal.pmed.1003931>.
63. Bycroft C, Freeman C, Petkova D. *et al.* The UK biobank resource with deep phenotyping and genomic data. *Nature*. 2018;**562**:203–209. <https://doi.org/10.1038/s41586-018-0579-z>.
64. Maier RM, Zhu Z, Lee SH. *et al.* Improving genetic prediction by leveraging genetic correlations among human diseases and traits. *Nat Commun* 2018;**9**:989. <https://doi.org/10.1038/s41467-017-02769-6>.
65. Rohde PD, Nyegaard M, Kjolby M. *et al.* Multi-trait genomic risk stratification for type 2 diabetes. *Front Med (Lausanne)* 2021;**8**:711208.
66. UK Biobank Primary Care Linked Data. 2024. https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/primary_care_data.pdf
67. Beesdo K, Knappe S, Pine DS. Anxiety and anxiety disorders in children and adolescents: developmental issues and implications for DSM-V. *Psychiatr Clin North Am* 2009;**32**:483–524. <https://doi.org/10.1016/j.psc.2009.06.002>.
68. Berenson GS, Srinivasan SR, Hunter SM. *et al.* Risk factors in early life as predictors of adult heart disease: the Bogalusa heart study. *Am J Med Sci* 1989;**298**:141–151.
69. McGill HC Jr, McMahan CA, Zieske AW. *et al.* Association of Coronary Heart Disease Risk Factors with microscopic qualities of coronary atherosclerosis in youth. *Circulation*. 2000;**102**:374–379.
70. Eto F, Samuel M, Finer S. *MULTIPLY-Initiative: clinical codelists Version 1.1*, v1.1 edn. Zenodo, 2023. <https://zenodo.org/badge/latestdoi/316011491>.
71. Baksh RA, Pape SE, Chan LF. *et al.* Multiple morbidity across the lifespan in people with down syndrome or intellectual disabilities: a population-based cohort study using electronic health records. *Lancet Public Health* 2023;**8**:e453–e462. [https://doi.org/10.1016/S2468-2667\(23\)00057-9](https://doi.org/10.1016/S2468-2667(23)00057-9).
72. Lambert SA, Gil L, Jupp S. *et al.* The polygenic score Catalog as an open database for reproducibility and systematic evaluation. *Nat Genet* 2021;**53**:420–425. <https://doi.org/10.1038/s41588-021-00783-5>.
73. Lambert SA, Wingfield B, Gibson JT. *et al.* Enhancing the polygenic score Catalog with tools for score calculation and ancestry normalization. *Nat Genet* 2024;**56**:1989–1994. <https://doi.org/10.1038/s41588-024-01937-x>.
74. Ge T, Chen C-Y, Ni Y. *et al.* Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 2019;**10**:1776.
75. Chang CC, Chow CC, Tellier LC. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;**4**:7.
76. Purcell S, Chang C. *PLINK*, 2.0. www.cog-genomics.org/plink/2.0/.
77. Manichaikul A, Mychaleckyj JC, Rich SS. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010;**26**:2867–2873. <https://doi.org/10.1093/bioinformatics/btq559>.
78. Truong B, Hull LE, Ruan Y. *et al.* Integrative polygenic risk score improves the prediction accuracy of complex traits and diseases. *Cell Genom* 2024;**4**:100523. <https://doi.org/10.1016/j.xgen.2024.100523>.
79. R CoreTeam. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2025. <https://www.R-project.org/>.
80. Robin X, Turck N, Hainard A. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;**12**:77.
81. Thompson DJ, Wells D, Selzam S. *et al.* A systematic evaluation of the performance and properties of the UK biobank polygenic risk score (PRS) release. *PLoS One* 2024;**19**:e0307270. <https://doi.org/10.1371/journal.pone.0307270>.
82. Thompson DJ, Wells D, Selzam S. *et al.* UK biobank release and systematic evaluation of optimised polygenic risk scores for 53 diseases and quantitative traits. medRxiv. 2022. <https://doi.org/10.1101/2022.06.16.22276246>.
83. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;**44**:837–845.