nature medicine

Article

https://doi.org/10.1038/s41591-026-04353-2

Data-driven prioritization of high-risk individuals for weight loss interventions

Received: 4 April 2025

Accepted: 17 March 2026

Published online: 30 April 2026

Check for updates

Kamil Demircan[1,2], Julia Carrasco-Zanini[1], Alice Williamson[1,2,3], Carl Beuchel[2], Linsey Jackson[4], Werner Römisch-Margl[5], Aleksander L. Hansen[6], Sarah Finer[7], David A. van Heel[1,8], Genes & Health Research Team*, Gabi Kastenmüller[5], Matthew Coghlan[4], Ida Moeller[4], Nicholas J. Wareham[9], Maik Pietzner[1,2] & Claudia Langenberg[1,2,10]

New obesity medications have demonstrated efficacy in trials, but their real-world deployment is partly limited by the absence of approaches that identify individuals for treatment based on risks for obesity-related complications. Here we present a risk prediction model to guide prioritization of high-risk individuals. In a population-based sample of ~200,000 individuals with a body mass index (BMI) exceeding 27 kg m^{-2}, our machine learning framework identified the 20 most informative features, from among thousands tested, that predict future onset of 18 complications of obesity, providing information beyond BMI. An integrated model (OBSCORE) successfully stratified individuals into risk groups based on incidence over 10 years: for example, 5.7%, 1.8%, 0.9%, 0.4% and 0.1% for cardiovascular mortality. We demonstrate generalizability of the model in independent populations of European and non-European ancestry and, in SURMOUNT-1 trial participants, show that weight loss was similar across baseline OBSCORE risk groups and that predicted risks decreased following treatment with tirzepatide. In summary, OBSCORE provides a framework for prioritizing high-risk individuals with overweight or obesity based on their risk of obesity-related complications, complementing BMI-based frameworks.

Obesity is a chronic disease that, if untreated, may lead to complications that range from metabolic diseases like type 2 diabetes (T2D) and cardiovascular diseases, to mechanical complications such as sleep apnea and others[1–5]. Annually, these complications account for ~5 million deaths and 130 million disability-adjusted life years, an approximately threefold increase compared to 1990[1]. Advances in obesity pharmacotherapy including semaglutide and tirzepatide targeting glucagon-like peptide-1 (GLP-1) and glucose-dependent insulinotropic polypeptide (GIP) show high efficacy in treating obesity[6–8] and some of its complications[9–13]. As 60–70% of adults in Western countries are classified with overweight or obesity, pharmacological treatment of all individuals is a major challenge for healthcare systems[14,15].

The identification and prioritization of individuals who are at high risk for the complications of obesity could help to guide effective allocation of resources including pharmacological, dietary, behavioral or surgical interventions but is currently hindered by important

[1]Precision Healthcare University Research Institute, Queen Mary University of London, London, UK. [2]Computational Medicine, Berlin Institute of Health at Charité-Universitätsmedizin Berlin, Berlin, Germany. [3]Friede Springer Cardiovascular Prevention Center at Charité-Universitätsmedizin Berlin, Berlin, Germany. [4]Eli Lilly and Company, Indianapolis, IN, USA. [5]Institute of Computational Biology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany. [6]Steno Diabetes Center Copenhagen, Herlev, Denmark. [7]Wolfson Institute of Population Health, Queen Mary University of London, London, UK. [8]Blizard Institute, Queen Mary University of London, London, UK. [9]IMS Epidemiology, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge, UK. [10]Max Planck Institute for Molecular Genetics, Berlin, Germany. *A list of authors and their affiliations appears at the end of the paper. e-mail: claudia.langenberg@bih-charite.de

https://doi.org/10.1038/s41591-026-04353-2

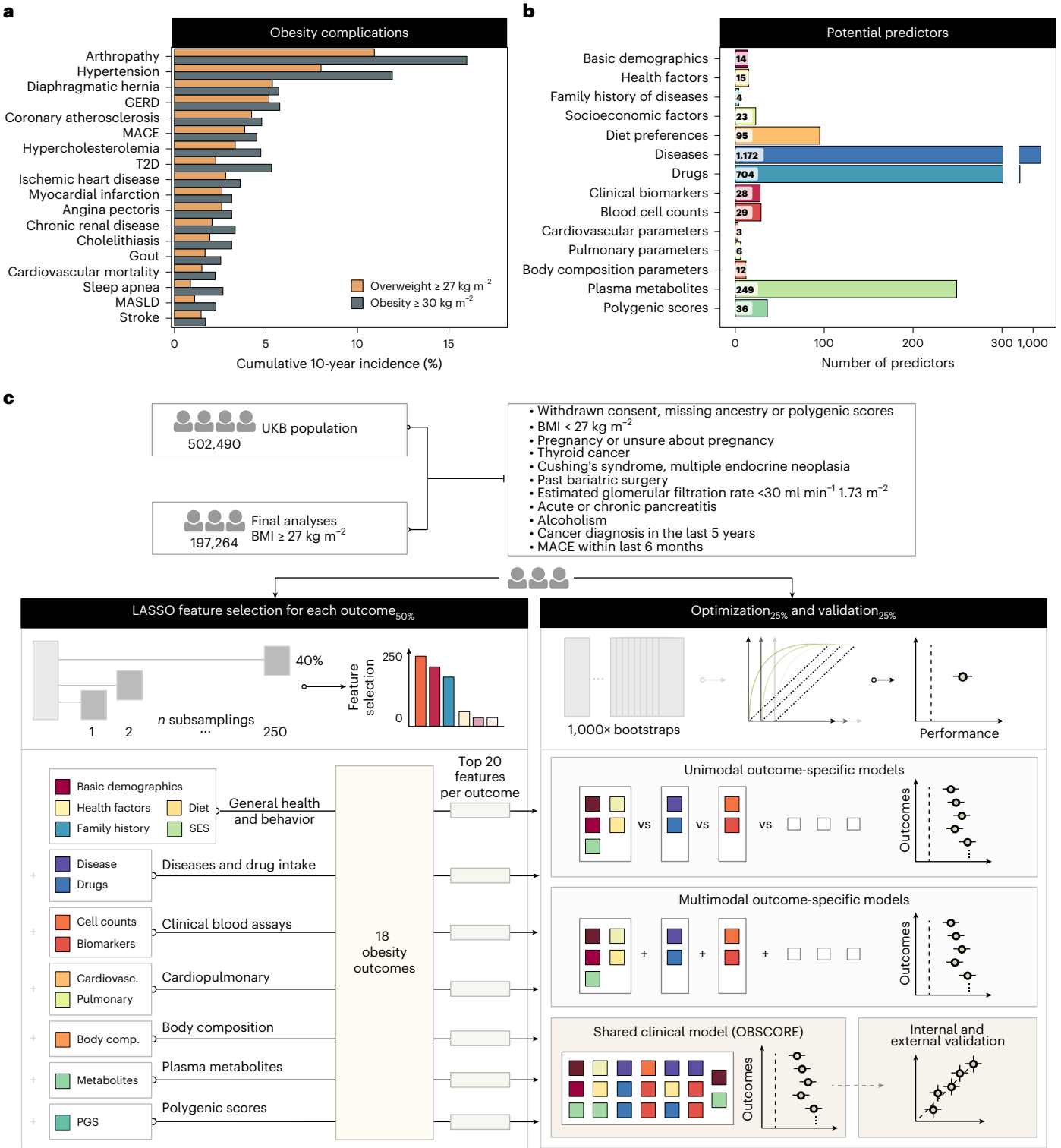


Fig. 1 | Outcomes, predictor pool and study design. a, Ten-year cumulative incidence of each outcome investigated in this study. **b**, Bar plot displaying the number of one-hot-encoded features considered per category. **c**, Exclusion criteria. The final dataset was split into feature selection sets (50%) and optimization (25%) and validation (25%) sets for each outcome. LASSO regression was used to select the top 20 predictive features for each set of predictors for each obesity-related outcome. We analyzed the predictive value of (1) each individual domain, (2) multimodal models, and (3) a clinically available model with a core set of shared features. The shared model was externally validated in the EPIC-Norfolk and G&H studies for available outcomes. Panel **c** created in BioRender; Blight, L. https://biorender.com/2y7grrp (2026). Cardiovasc., cardiovascular; comp., composition; MASLD, metabolic dysfunction-associated steatotic liver disease; SES, socioeconomic status.

challenges. The limitations of using BMI alone as a risk measure for individuals are well recognized, and a 2025 commission report advocated for a definition of clinical obesity that considers health status and obesity subtypes beyond BMI, based on contemporary population data[16]. Yet the parameters that need to be considered to accurately identify 'clinical obesity' and predict its complications remain unclear,

and no framework for guiding the allocation of interventions, considering differential risk, has been developed in a data-driven, risk-based manner. Large-scale, population-based studies with comprehensive baseline assessment and linkage to healthcare records now facilitate the analysis of diverse data modalities in contemporary populations[17,18], rather than relying on historical data from selected study participants[19]. We have previously shown that a simple, interpretable machine learning framework can identify sparse sets of informative predictors among thousands of features for diverse diseases[20–22], but this has not yet been applied to build a risk-based tool that could inform a data-driven obesity management strategy.

Here we develop and present a risk prediction tool for prioritizing high-risk individuals for interventions based on a single integrated model (OBSCORE) with 20 clinical parameters that predict the future onset of 18 important cardiovascular, metabolic and mechanical complications of obesity. We derive this model in a contemporary population of nearly 200,000 adults potentially eligible for obesity medication, by empirically integrating a comprehensive set of more than 2,300 features.

Results

Of ~500,000 UK Biobank (UKB) participants, we included 197,264 individuals mirroring the inclusion criteria for recent weight loss medication trials[7] (Fig. 1). We included 97,421 participants living with overweight (BMI 27–30 kg m^{-2}) and 99,843 living with obesity (BMI ≥ 30 kg m^{-2}), with a total of 95,034 women and 102,230 men (median age at recruitment 58 years) (Extended Data Table 1). A total of 9.3% of the included participants had T2D at baseline (13% and 5.7% in individuals with obesity and overweight (BMI ≥ 27 kg m^{-2}), respectively) and 4.4% had a medical history of a major adverse cardiac event (MACE), including myocardial infarction or stroke. We investigated the incidence of 18 obesity-related complications as outcomes, considering the incidence of cardiovascular, metabolic and mechanical complications of obesity, as well as cardiovascular mortality over a 10-year follow-up period, excluding prevalent disease for each outcome and events occurring within the first 6 months of follow-up (Fig. 1a). Arthropathy was the most frequently occurring outcome, affecting 13.4% of the cohort, with hypertension occurring in 9.7%. A total of 9,337 deaths occurred within the follow-up period. We considered a total of 2,390 predictors from various categories, ranging from clinically available parameters such as basic demographics and clinical blood assays to more complex data such as plasma metabolite levels and polygenic scores (Fig. 1b), of which 2,078 were used as predictors after pruning for highly correlated features ($r < 0.9$). We conducted a two-step machine learning procedure to build and validate sparse models with up to 20 features and analyzed the predictive value of (1) each individual data modality (for example, general health and behavior or among others), (2) multimodal models with stepwise addition of all modalities, and (3) a clinical model with a single set of 20 shared features for prediction across all 18 outcomes (Fig. 1c). The total number of incident cases overall and in training, optimization and validation sets is depicted in Supplementary Tables 1 and 2. We externally validated the shared model in the European Investigation into Cancer (EPIC)-Norfolk[23] and Genes & Health (G&H) studies.

Performance of unimodal outcome-specific models

We compared the discriminative performance of models including the top 20 features from each data modality (that is, feature set) individually (Fig. 2, Extended Data Fig. 1 and Supplementary Table 3). We observed a wide range of performance by individual domains for the prediction of the complications of obesity, with general health and behavior (median concordance index (C-index) 0.713, range 0.593–0.794) and clinical blood biomarkers (median C-index 0.695, range 0.557–0.850) achieving the best performance and polygenic scores achieving the lowest performance (median C-index 0.564; range 0.506–0.623). Sensitivity analyses including all non-random features of individual domains in the models instead of only the 20 top features demonstrated highly similar performance (Supplementary Fig. 1). We observed no apparent increase in C-indices for most domains apart from increases for the diseases and medication domain (median delta C-index 0.02, range 0.005–0.033).

Performance of multimodal outcome-specific models

We next analyzed discriminative performance when adding different data domains, by selecting the top 20 features from increasingly comprehensive pools of predictors for each outcome separately in the following order: (1) indicators of general health and behavior, (2) diseases and drug intake, (3) clinical blood biomarkers, (4) cardiopulmonary parameters, (5) body composition, (6) plasma metabolites and (7) polygenic scores (Fig. 3 and Supplementary Table 4).

For most outcomes, the integration of current and past diagnoses, current medication and clinical blood assays on top of general health and behavior parameters led to increases in C-indices (increase in median mean C-index 0.025). This emphasizes the large predictive value of information mostly available in clinical records. However, further increasing the predictor pool to select from cardiopulmonary parameters (change in median mean C-index −0.0006), body composition (increase in median mean C-index 0.002), plasma metabolites (change in median mean C-index −0.0003 or polygenic scores (increase in median mean C-index 0.003) provided limited additional gains in predictive accuracy for most outcomes. Notably, although body composition markers are well-known predictors of the complications of obesity, the addition of these measures resulted in relatively small increases in performance across all outcomes. The largest increase in performance with the addition of body composition was observed for sleep apnea (0.027). Augmenting the potential predictor pool to all modalities of data, predictive performance varied substantially across outcomes (median C-index 0.75, range 0.634–0.864). The highest performance was observed for gout (0.864 (95% CI 0.852–0.876)), chronic kidney disease (0.863 (95% CI 0.852–0.872)) and T2D (0.854 (95% CI 0.846–0.863)), largely driven by clinical blood assays. By contrast, the lowest performance was observed for diaphragmatic hernia and gastroesophageal reflux disease (GERD), with maximum C-indices of 0.634 (95% CI 0.623–0.646) and 0.643 (95% CI 0.63–0.654), respectively. Outcome-specific features most frequently selected across outcomes are displayed in Extended Data Fig. 2 for clinically available models (excluding plasma metabolites and polygenic scores) and Extended Data Fig. 3 for extended models.

Performance of the data-driven shared clinical model OBSCORE

Implementing separate predictive models for each health outcome is impractical in clinical settings and would not result in an integrated assessment of overall risk. We investigated shared predictors among top features based on the training set for the 18 outcome-specific models, with the goal of identifying common predictors that could inform a unified model applicable to multiple outcomes[20,24]. We prioritized rapidly available clinical features over more sophisticated markers such as serum metabolites and polygenic scores, owing to only small prognostic value, and identified a shared model (OBSCORE) with a set of 20 largely uncorrelated features (Fig. 4a,b, Extended Data Fig. 4 and Supplementary Table 5).

In the validation sets, OBSCORE achieved discriminative performance comparable with the extended outcome-specific models (that is, those including top-20 features from all data modalities), with an overall median deviation in C-index across outcomes of 0.012 (range −0.006–0.024) (Extended Data Fig. 5).

We next compared OBSCORE to a model consisting of age, sex and BMI, as well as the well-recognized definition of metabolically healthy obesity (MHO). Considering the unavailability of similar

https://doi.org/10.1038/s41591-026-04353-2

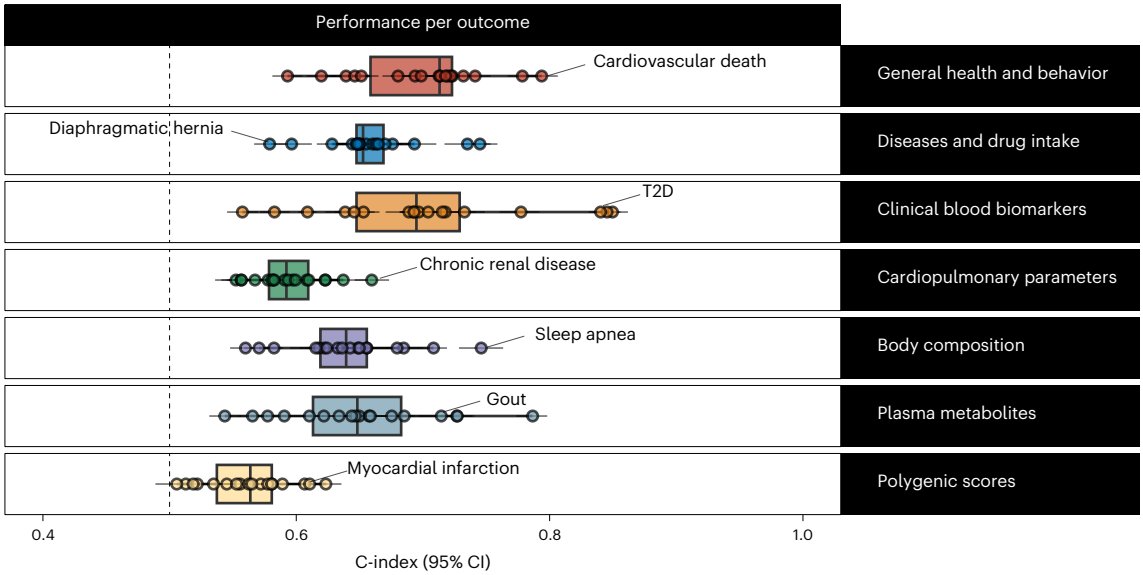


Fig. 2 | Performance of unimodal outcome-specific models. Outcome-specific models were built using the LASSO and Ridge Cox algorithms, selecting from up to 20 features from different categories of predictors individually: that is, indicators of general health and behavior, diseases and drug intake, clinical blood biomarkers, cardiopulmonary parameters, body composition, plasma metabolites and polygenic scores. Data points display the mean C-index with error bars showing 95% confidence intervals (CIs) across 1,000 bootstraps. Vertical dashed line at C-index = 0.5 denotes a random predictor. Boxplots show the median (center line) and the 25th and 75th percentiles (lower and upper box limits, respectively), and whiskers extend to the most extreme observed values within 1.5× the interquartile range from the lower and upper quartiles. Data points beyond this range are outliers. A selection of outcomes is labeled. Sample sizes are provided in Supplementary Table 2.

obesity-complication focused scores, we compared the performance of OBSCORE with features from established risk predictors such as SCORE2 and atherosclerotic cardiovascular disease (ASCVD) risk estimator (Fig. 4b and Supplementary Table 6). OBSCORE outperformed the BMI-based approach across all outcomes (Fig. 4b). Across 17 outcomes, OBSCORE outperformed ASCVD features (median difference in C-index of 0.027, range −0.004–0.146). Although ASCVD features performed comparably to OBSCORE for some cardiovascular outcomes, such as hypertension and coronary atherosclerosis, we observed a considerable difference for death from cardiovascular causes (difference in mean C-index 0.035). Particularly for non-cardiovascular outcomes, there were stark differences: for example, for T2D (mean C-index difference = 0.146) and sleep apnea (C-index difference = 0.079). Performance did not differ when excluding early incident cases up to 2 years after baseline in a sensitivity analysis (Extended Data Fig. 6). In subgroup analyses, model performance was similar comparing European to non-European ancestries or groups of lower versus higher Townsend deprivation index (Extended Data Fig. 7). We tested the generalizability and performance of OBSCORE through external validation in the prospective EPIC-Norfolk cohort study. We used coefficients derived in UKB—that is, without model retraining—for 18 of the 20 features available in EPIC-Norfolk to predict 14 outcomes (range of case numbers = 48–217). We observed strong agreement in performance ($r = 0.83$; $P = 2.7 \times 10^{-4}$), indicating high generalizability of the model (Fig. 4c). We further tested the generalizability of our model to individuals of South Asian ancestry from the G&H study, a population at high risk of developing T2D early and at lower levels of BMI compared to white Europeans. Here, 15 available OBSCORE features predicted incident T2D well (mean C-index = 0.78) and achieved a greater improvement (difference in mean C-index) of 0.196, compared to 0.166 in UKB over the basic model including age, sex and BMI. This indicates generalizability of the features selected, even in populations with different profiles and study methods.

Beyond C-indices, screening metrics including detection rates across different false positive rate (FPR) thresholds ranging from 5% to 40% were similar between the shared clinical and outcome-specific extended models (Supplementary Fig. 2). The probability of a high-risk individual, identified through OBSCORE, developing obesity-associated complications (that is, post-test probabilities) at a 10% FPR ranged from 4.8% to 28.7% (median = 10%) across the 18 outcomes studied, compared to 3.9–24.5% (median = 8.5%) for high-risk individuals identified by ASCVD features. For example, post-test probabilities at 10% FPR for OBSCORE were as high as 18.5% for T2D, an approximately twofold increase compared to post-test probabilities from other models (Fig. 4d,e). Pretest and post-test probabilities at 10% FPR for the other outcomes are displayed in Supplementary Figs. 3 and 4. We evaluated calibration in the held-out validation sets. Most outcomes were well calibrated, although OBSCORE overestimated risk for some outcomes (Supplementary Fig. 5). Applying Platt's logistic regression improved calibration across outcomes.

Risk stratification by OBSCORE

The utility of predictive models lies in their ability to stratify individuals by absolute risk, enabling targeted interventions and informed clinical decision-making. To evaluate this stratification ability, we analyzed the observed numbers of events across 10 years for estimated risk groups, as depicted in Fig. 5a. We observed large proportions of participants with overweight instead of obesity, even in the highest fifth of risk scores (Fig. 5a), highlighting limitations of BMI-only–based approaches to prioritize obesity complication risk. Across outcomes, the proportion of individuals with BMI 27–30 kg m^{-2} in the highest risk decile ranged from 19% to 45% (median = 39.9%). Notably, despite being among the most proximal outcomes of obesity, the proportion of BMI 27–30 kg m^{-2} among individuals at highest risk for T2D was 30%. To further investigate the large differences in risk despite similar BMI, we investigated the distributions of estimated risk stratified by overweight BMI 27–30 kg m^{-2}, class 1, class 2 and class 3 obesities (Fig. 5b). Within each stratum, we observed wide-ranging risk differences for most outcomes, despite equal BMI category. This overlap was smaller for outcomes more proximally related to body weight, such as metabolic dysfunction-associated steatotic liver disease. OBSCORE stratified risk across all outcomes (that is, rate ratios were consistently more than 1

https://doi.org/10.1038/s41591-026-04353-2

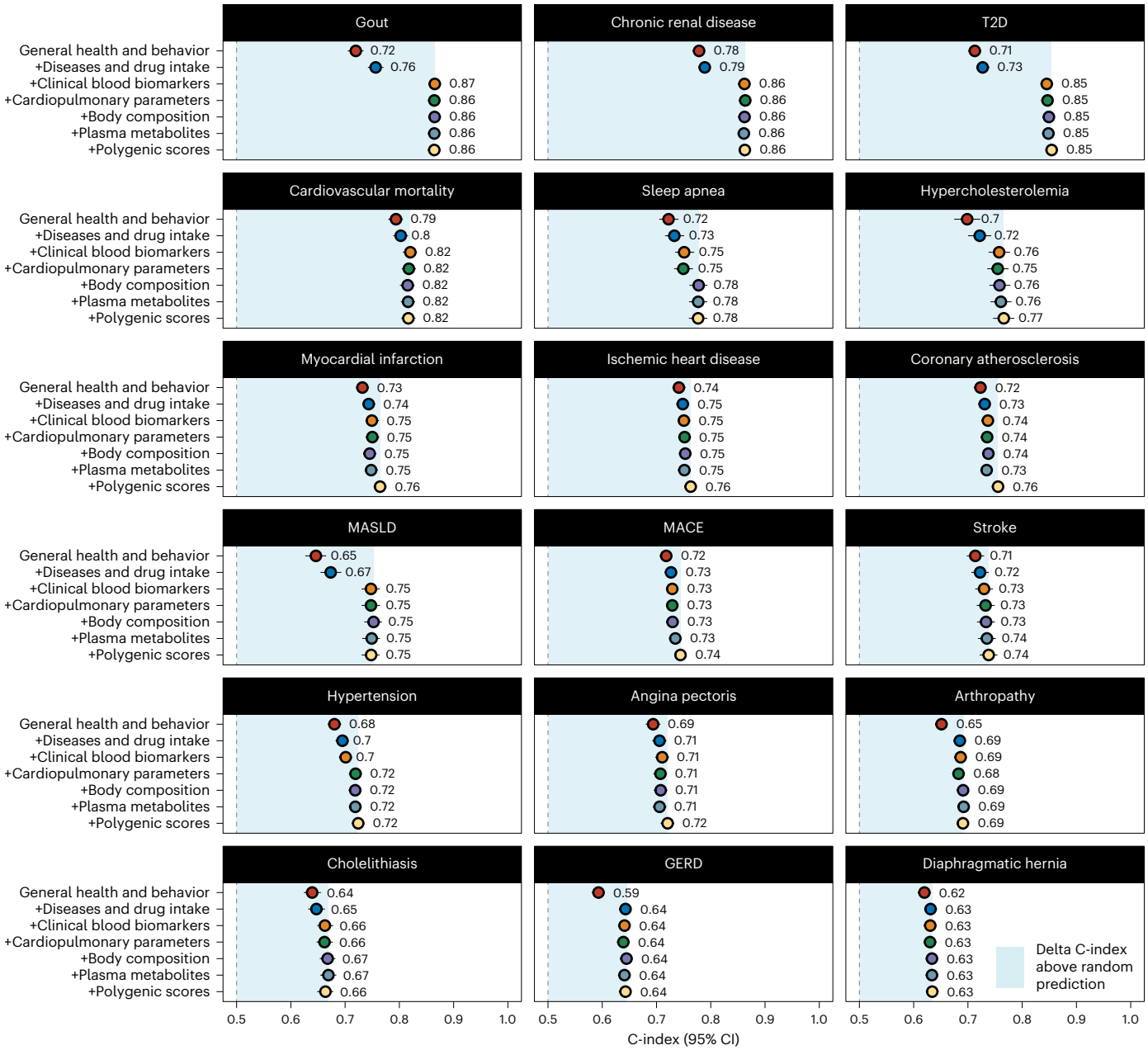


Fig. 3 | Performance of multimodal outcome-specific models. Outcome-specific models were built using LASSO and Ridge Cox models, selecting the top 20 features from increasingly comprehensive pools of predictors, for each outcome separately. The predictors were added stepwise in the following order: (1) indicators of general health and behavior, (2) diseases and drug intake, (3) clinical blood biomarkers, (4) cardiopulmonary parameters, (5) body composition, (6) plasma metabolites and (7) polygenic scores. Data points display the mean C-index with error bars showing 95% CIs across 1,000 bootstrap tests. Vertical dotted line at C-index = 0.5 denotes a random predictor. Blue shaded area depicts the maximum increase in C-index over random prediction. Sample sizes are provided in Supplementary Table 2.

between the top and bottom 20%). For 12 out of 18 outcomes, OBSCORE demonstrated substantial stratification (rate ratio > 10). Rate ratios were as high as 89-fold for chronic renal disease, 36 for gout and 42 for T2D. Notably, the rate ratio for death by cardiovascular cause was 47, with a 10-year absolute event rate of 5.7% in 20% of the individuals with the highest risk. Cumulative 10-year incidences for cardiovascular mortality across all quintiles are shown in Fig. 5c. We note that the rate ratios of this magnitude warrant particular attention, as we did not compare to participants within the reference BMI range, but all participants, even in the lowest 20%, were living with overweight (BMI 27–30 kg m^{-2}) or obesity. Moreover, rate ratios comparing individuals in the top to bottom 50% of OBSCORE were higher (range = 1.9–14.9) than those comparing individuals with a BMI of >30 kg m^{-2} to those with a BMI of 27–30 kg m^{-2} (range = 1.05–3.07) (Extended Data Fig. 8). We demonstrate clinical heterogeneity by exemplary illustrating the risk of developing 18 outcomes for three selected individuals with the same age, sex and BMI (Fig. 5d). We developed an interactive web interface (https://omicscience.org/apps/obscore/) to enable interrogation and exploration of model-derived individual-level risk estimates.

OBSCORE in a randomized controlled trial of tirzepatide

To test whether changes in adiposity measures by tirzepatide treatment were similar across OBSCORE risk groups and whether OBSCORE captured treatment-induced changes in risk, we applied our model with 12 of the 20 features available to individual-level data from the SURMOUNT-1 trial[7]. We observed comparable reductions

https://doi.org/10.1038/s41591-026-04353-2

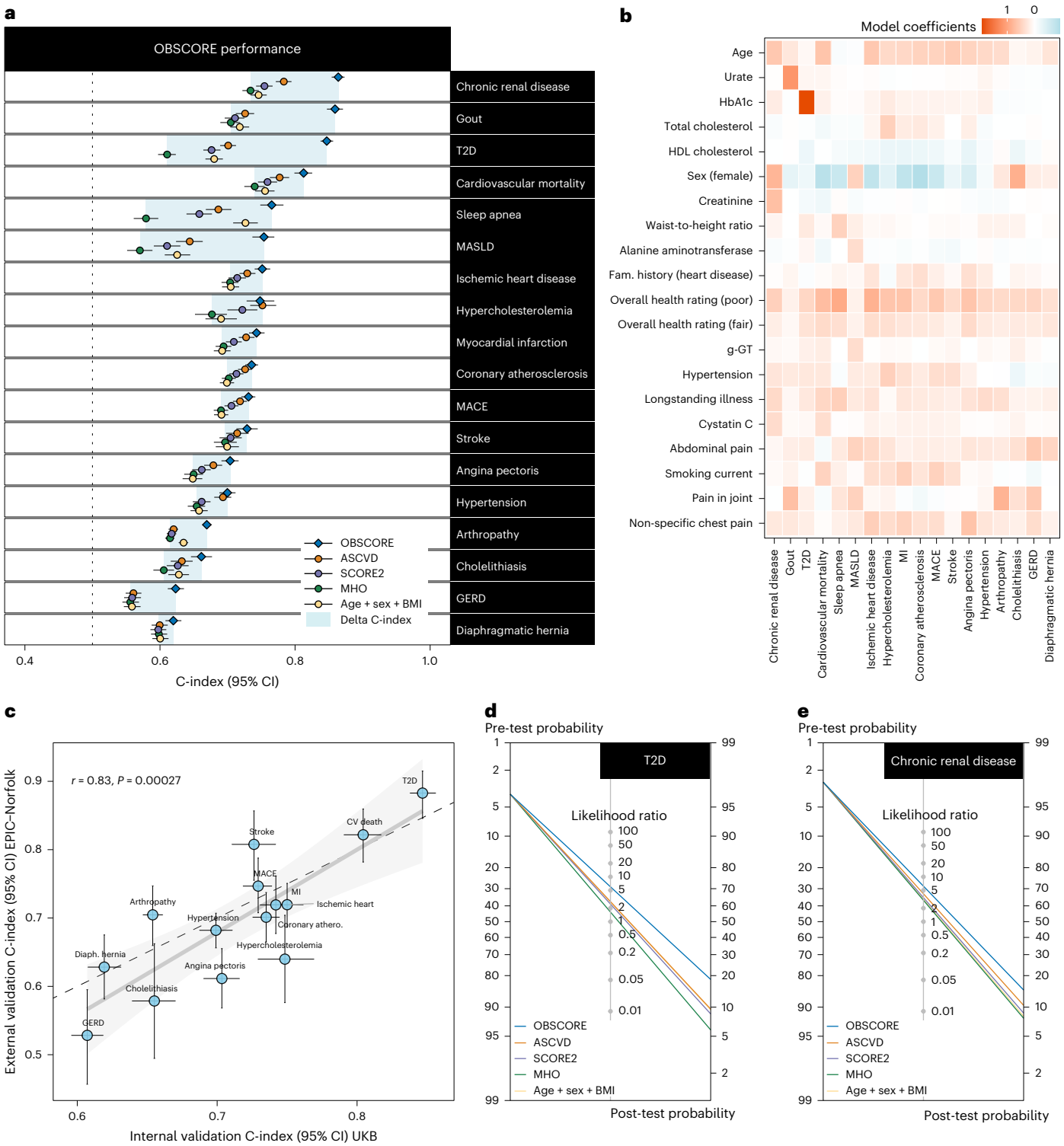


Fig. 4 | Features and performance of OBSCORE. a, Performance of OBSCORE in comparison with age, sex and BMI, as well as features from other risk models, such as MHO, SCORE2 risk score and ASCVD risk estimator. Data points display the mean C-index, with error bars showing 95% CIs estimated from 1,000 bootstrap resamples. Sample sizes are provided in Supplementary Table 2. **b**, Top 20 features identified and used in OBSCORE and their coefficients across all outcomes. **c**, External validation of OBSCORE (with 18 features) performance in the EPIC-Norfolk study for 14 outcomes. Data points display the mean C-index with 95% CIs estimated from 1,000 bootstrap samples. The solid line represents the fitted linear regression model, with shaded bands indicating 95% CIs. Two-sided Pearson's correlation test was used to derive the raw *P* value and the correlation coefficient (*r*). **d**, Fagan's nomogram for T2D displaying pretest and post-test probabilities using different models. **e**, Fagan's nomogram for chronic renal disease displaying pretest and post-test probabilities using different models. Pretest probability is calculated based on incidence in the study, and post-test probability is calculated based on likelihood ratio for each model at 10% FPR. Athero., atherosclerosis; diaph., diaphragmatic; fam., family.

https://doi.org/10.1038/s41591-026-04353-2

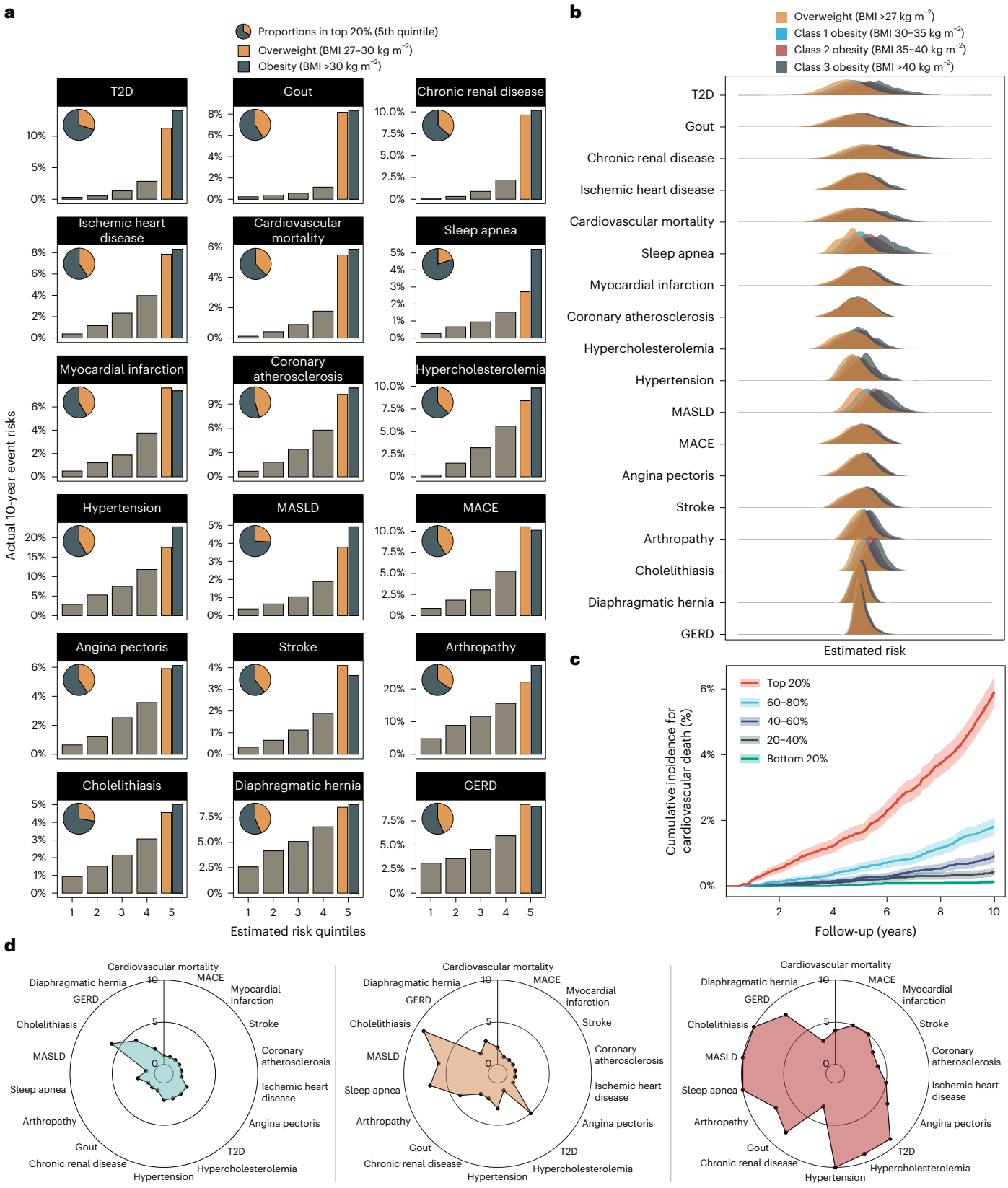


Fig. 5 | Risk stratification by OBSCORE. a, Ten-year observed risk rates across outcomes, according to groups (each containing 20% of individuals) of risk states as estimated by OBSCORE. The bar of the highest 20% (5th quintile) shows event rates stratified by individuals with overweight (BMI 27–30 kg m^{-2}) and obesity BMI >30 kg m^{-2}. Embedded pie charts depict the proportions of individuals with overweight versus obesity in the highest 20%. **b**, Distribution of risk estimates in individuals stratified by overweight and class 1, class 2 and class 3 obesities. **c**, Cumulative incidence plot for death from cardiovascular causes according to the bottom 20%, 20–40%, 40–60%, 60–80% and top 20% of estimated risk. 1 − Kaplan–Meier estimates are shown, and error bands represent 95% CIs. **d**, Radar plots depict the predicted risks of three individuals with the same age, sex and BMI for the 10-year development of 18 obesity-associated complications. Risk deciles are 10% risk groups as compared with the underlying reference group in the UKB.

https://doi.org/10.1038/s41591-026-04353-2

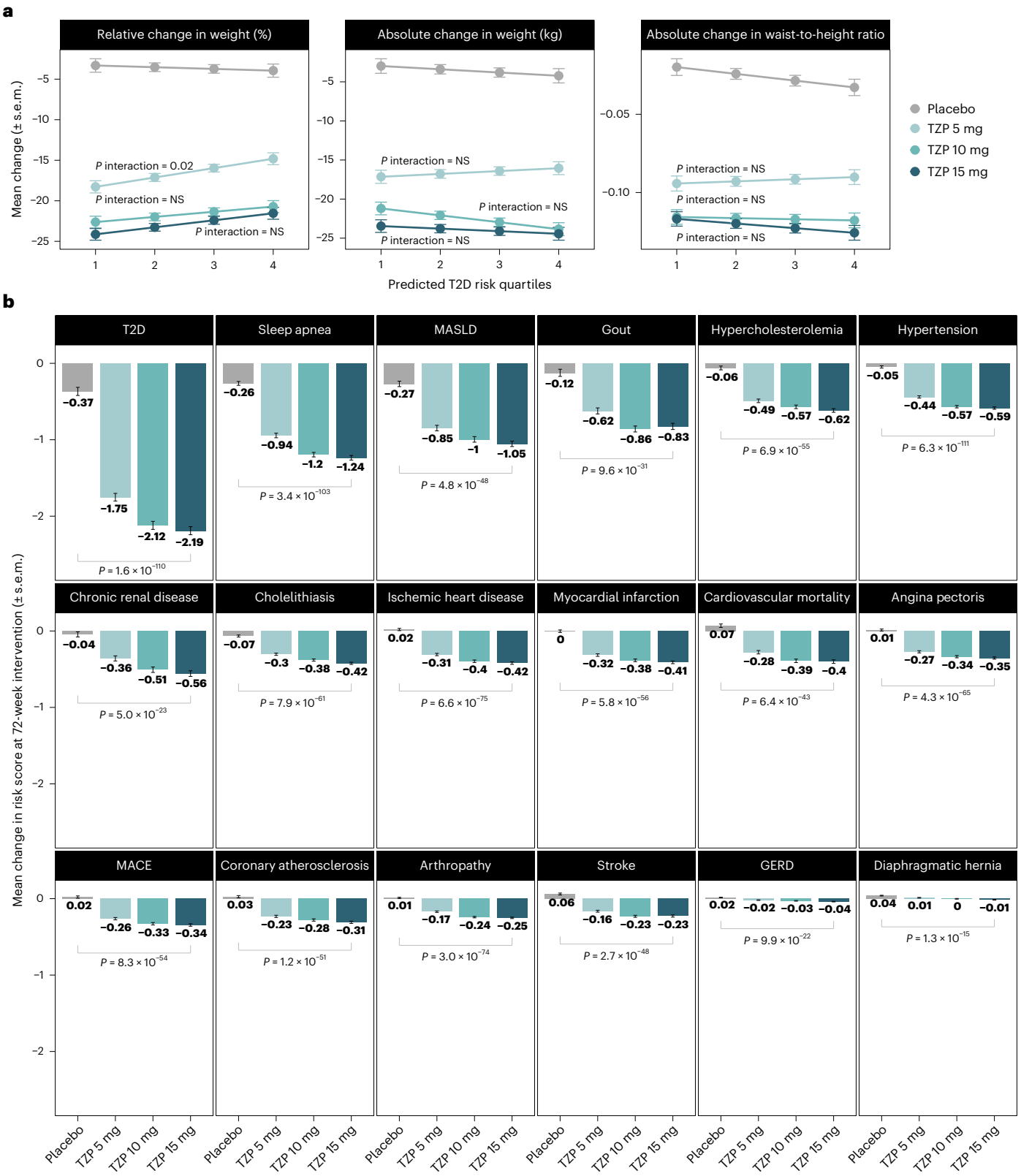


Fig. 6 | OBSCORE in the SURMOUNT-1 trial. a, EMMs of change in percent body weight, body weight and waist-to-height ratio along with ±s.e.m. at week 72 after treatment based on predicted T2D risk quartiles in the SURMOUNT-1 trial. Unadjusted interaction *P* values (treatment × risk quartiles) were obtained from linear regression. $n_{placebo}$ = (Q1 = 89, Q2 = 103, Q3 = 101, Q4 = 95), $n_{TZP\,5mg}$ = (Q1 = 116, Q2 = 104, Q3 = 126, Q4 = 116), $n_{TZP\,10mg}$ = (Q1 = 121, Q2 = 127, Q3 = 112, Q4 = 119), $n_{TZP\,15mg}$ = (Q1 = 125, Q2 = 117, Q3 = 112, Q4 = 121). **b**, EMMs of change in predicted risk scores (linear predictors) across 18 outcomes at week 72 (±s.e.m.) in placebo and treatment arms. $n_{placebo}$ = 388, $n_{TZP\,5mg}$ = 462, $n_{TZP\,10mg}$ = 479, $n_{TZP\,15mg}$ = 475. Unadjusted *P* values for two-sided pairwise comparisons between 15 mg Ttrzepatide (TZP) and placebo are shown. NS, not significant.

in body weight and waist-to-height ratio with tirzepatide (72 weeks) across OBSCORE-predicted risk groups at baseline, with only one of 162 comparisons reaching the multiple-testing-corrected significance threshold (Supplementary Table 7). Absolute changes in weight and waist-to-height ratio were on average greater in individuals with higher predicted OBSCORE risk, whereas relative changes, such as in weight loss, were slightly smaller—for example, for T2D (Fig. 6a)—consistent with findings in SURMOUNT-2[25]. At baseline, OBSCORE-predicted risks in the treatment versus placebo arms did not differ (Extended Data Fig. 9), but they were significantly lower in all tirzepatide arms after the 72-week follow-up across all outcomes (Fig. 6b and Supplementary Table 8), indicating that OBSCORE captured treatment-associated changes in predicted risk.

Discussion

In this study of ~200,000 individuals with overweight (BMI 27–30 kg m^{-2}) or obesity, we systematically investigated the performance of different data modalities, individually or in combination, for predicting the 10-year incidence of 18 obesity-associated complications. Our agnostic framework identified a shared predictive model (OBSCORE) that can effectively predict these complications and may inform risk-based allocation of obesity medication as well as other dietary, behavioral or surgical interventions.

New obesity medications targeting hunger signaling via GLP-1 and in combination with GIP and glucagon represent a major advance in obesity management. Comparing with bariatric surgery, recent data indicate that short-term weight management and outcomes for complications may be similar[26]. Hence, these pharmacological interventions pose a promising option to manage overweight and obesity at population scale, globally[27]. The ultimate goal of these interventions is to prevent obesity-associated outcomes. However, there is currently no consensus for identifying individuals at highest risk for obesity-associated complications, who could be prioritized for allocation of the new medications. Besides clinicians and clinical routine, this uncertainty also affects other key stakeholders, such as policymakers, healthcare systems and insurance providers, with implications for how access is determined[28]. A framework for prioritization is important for several reasons, including the current high costs[28,29], global shortages[30], lack of data-driven support tools for clinicians and targeting of drugs to those at highest absolute risk. Currently, recommendations for the allocation of these therapies remain underdeveloped. For instance, owing to scarcity of data and data-driven approaches providing a rationale for prioritization, the US health insurance program Medicare limited the coverage of semaglutide to a specific set of individuals with obesity and known history of heart disease or stroke[31,32]. OBSCORE, however, identified many individuals with overweight (BMI 27–30 kg m^{-2}) with other risk factors at similarly high risk for complications. For individuals identified as high-risk despite having a BMI less than 27 kg m^{-2}, it remains to be determined whether targeted interventions provide clinical benefit, as current trials mostly focused on individuals with overweight or obesity. Whether different cut-offs for BMI should be considered is particularly important for certain populations who have a higher risk of obesity-related complications at lower BMI cut-offs. Moreover, such criteria often neglect non-cardiovascular outcomes, although for some of them, benefit has been demonstrated[27]. Further, challenging the commonly practiced BMI-based prioritization, our results highlight how large intraindividual discordance for cardiovascular/non-cardiovascular complications can be, even among individuals with a similar BMI. Overall, our data-driven approach aims to suggest an agnostically derived set of predictors that may be considered for identifying individuals at high risk for a broad set of obesity complications.

BMI thresholds are the currently implemented, simple mechanism for establishing eligibility for behavioral and pharmacological obesity management intervention. Recent consort statements[16] have highlighted the need to develop frameworks that better align eligibility with adiposity-associated clinical risks. Our findings suggest that OBSCORE could be useful for refining existing BMI-based criteria by identifying individuals with overweight and obesity and with a risk profile comparable to or greater than individuals currently eligible based on BMI cut-offs. Ideally, future studies will further validate and calibrate the model in additional populations and with sensitive and specific follow-up for diverse obesity-related outcomes. We envisage OBSCORE to be implemented as a data-driven support tool for referral that complements existing frameworks, rather than requiring a drastic change of regulatory indications or reimbursement policies. Within systems such as NICE in the UK, OBSCORE could be incorporated at the level of technology appraisal and clinical guidelines. To operationalize targeting of 'high-risk' OBSCORE individuals, processes to ensure the availability and clinical documentation of parameters required for standardized, quantitative risk estimation would need to be implemented. The fact that OBSCORE is based on a minimal set of 20 features, most of which are readily available in a clinical setting, may facilitate implementation in electronic health record (EHR)-based decision support, for (semi)automated identification of eligible patients at the point of care.

This approach is a complementary perspective to concepts such as 'obesities'[33] and MHO[34,35] or 'discordant obesity'[36,37]. Each of these concepts aims to identify and explain facets of heterogeneity in obesity, and they have advanced our understanding and clinical view of obesity. However, these approaches so far have considered a set of prespecified cross-sectional features to define heterogeneity. By contrast, our approach has empirically derived most informative features, based on their longitudinal associations with outcomes, to identify clinical obesity, as recently highlighted[16]. Nevertheless, future work should determine whether OBSCORE's sparse combination of features could be further refined in the context of cost-effectiveness and patient outcomes.

Using individual-level data from the SURMOUNT-1 trial, we demonstrated that tirzepatide-related changes in weight and waist-to-height ratio were evident and comparable across OBSCORE risk groups at baseline, suggesting that treatment efficacy demonstrated in the trial[7] may be generalizable to individuals prioritized by our model. In combination with their higher predicted risk of important obesity-related outcomes, this further highlights the potential opportunity for targeted intervention in this group.

Some limitations should be acknowledged when interpreting our results. Our study included individuals of middle to older ages. Considering the dramatic rise of obesity among adolescents and younger adults[14], future research must evaluate the generalizability of our model to those age groups and other populations not represented in this study. However, we observed good performance across non-European ancestries and across Townsend deprivation groups. The UKB represents a healthier population compared to the general population, leading to an underestimation of event rates[38]. Although this impacts absolute risk estimates, it is unlikely to substantially affect the model-building process or model performance. Hence, although discriminative performance was strong and can stratify the population into statistically defined high and low risk individuals, ultimate reliability on absolute risks for clinical decision-making will require future research on (1) informed absolute risk cut-offs for several of these diseases yet needed to be established by the clinical community and (2) calibration with larger-scale and diverse cohorts or trials with stricter case ascertainment. Leveraging EHR data to ascertain outcomes may have introduced potential biases[39]. Although the predictors identified in our analysis are clinically plausible and partially consistent with other models, the study design was not intended to establish causal predictors, and hence these should be interpreted with caution. Some outcomes with high incidence rates, such as arthropathy and GERD, displayed lower C-indices, necessitating further studies with other data modalities to establish stronger predictors.

The comprehensive and unbiased approach marks an important strength of our study. We leveraged a large and diverse set of more than 2,300 predictors in ~200,000 individuals with overweight and obesity. Our approach represents one of the largest studies of individuals with overweight or obesity with data on 249 plasma metabolites. Moreover, we covered a range of domains including easily accessible information on general health and behavior, a deep characterization of medical history with more than 1,000 current diagnoses and more than 700 different medications, most of the clinically relevant blood biomarkers and blood cell counts, clinically accessible cardiopulmonary measurements, different body composition parameters and established polygenic scores for 36 different diseases or traits. Our use of an interpretable machine learning framework refined through our prior work[20–22] enabled us to distill this large set into a sparse subset. These predictors were selected in a data-driven manner and with minimal manual curation, and are both clinically plausible and implementable in real-world settings. We were able to externally validate OBSCORE for 14 out of 18 outcomes tested in independent studies, showing overall generalizability of OBSCORE and retained performance of the models, although remaining outcomes need further testing. Notably, external validation showed robust results even though there were differences in methods between the studies: for example, despite EPIC-Norfolk participants being recruited ~15–20 years earlier. This consistency emphasizes the potential for real-world applications.

In conclusion, we developed and validated a risk prediction tool to identify individuals at high risk of developing obesity-related complications. These findings may inform data-driven, risk-centered tailoring of obesity management strategies.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-026-04353-2.

References

1. Zhou, X. D. et al. Burden of disease attributable to high body mass index: an analysis of data from the Global Burden of Disease Study 2021. *EClinicalMedicine* **76**, 102848 (2024).
2. Ong, K. L. et al. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet* **402**, 203–234 (2023).
3. Kivimäki, M. et al. Body-mass index and risk of obesity-related complex multimorbidity: an observational multicohort study. *Lancet Diab. Endocrinol.* **10**, 253–263 (2022).
4. Lingvay, I., Cohen, R. V., Roux, C. W. L. & Sumithran, P. Obesity in adults. *Lancet* **404**, 972–987 (2024).
5. Huang, J. et al. Genomics and phenomics of body mass index reveals a complex disease network. *Nat. Commun.* **13**, 7973 (2022).
6. Wilding, J. P. H. et al. Once-weekly semaglutide in adults with overweight or obesity. *N. Engl. J. Med.* **384**, 989–1002 (2021).
7. Jastreboff, A. M. et al. Tirzepatide once weekly for the treatment of obesity. *N. Engl. J. Med.* **387**, 205–216 (2022).
8. Jastreboff, A. M. et al. Triple-hormone-receptor agonist retatrutide for obesity - a phase 2 trial. *N. Engl. J. Med.* **389**, 514–526 (2023).
9. Loomba, R. et al. Tirzepatide for metabolic dysfunction-associated steatohepatitis with liver fibrosis. *N. Engl. J. Med.* **391**, 299–310 (2024).
10. Malhotra, A. et al. Tirzepatide for the treatment of obstructive sleep apnea and obesity. *N. Engl. J. Med.* **391**, 1193–1205 (2024).
11. Perkovic, V. et al. Effects of semaglutide on chronic kidney disease in patients with type 2 diabetes. *N. Engl. J. Med.* **391**, 109–121 (2024).
12. Bliddal, H. et al. Once-weekly semaglutide in persons with obesity and knee osteoarthritis. *N. Engl. J. Med.* **391**, 1573–1583 (2024).
13. Deanfield, J. et al. Semaglutide and cardiovascular outcomes in patients with obesity and prevalent heart failure: a prespecified analysis of the SELECT trial. *Lancet* **404**, 773–786 (2024).
14. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in underweight and obesity from 1990 to 2022: a pooled analysis of 3663 population-representative studies with 222 million children, adolescents, and adults. *Lancet* **403**, 1027–1050 (2024).
15. Shi, I. et al. Semaglutide eligibility across all current indications for US adults. *JAMA Cardiol.* https://doi.org/10.1001/jamacardio.2024.4657 (2024).
16. Rubino, F. et al. Definition and diagnostic criteria of clinical obesity. *Lancet Diabetes Endocrinol.* https://doi.org/10.1016/S2213-8587(24)00316-4 (2025).
17. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
18. Bick, A. G. et al. Genomic data in the All of Us Research Program. *Nature* **627**, 340–346 (2024).
19. The Lancet Diabetes & Endocrinology. Redefining obesity: advancing care for better lives. *Lancet Diabetes Endocrinol.* https://doi.org/10.1016/S2213-8587(25)00004-X (2025).
20. Carrasco-Zanini, J. et al. Proteomic prediction of diverse incident diseases: a machine learning-guided biomarker discovery study using data from a prospective cohort study. *Lancet Digit. Health* **6**, e470–e479 (2024).
21. Carrasco-Zanini, J. et al. Proteomic signatures improve risk prediction for common and rare diseases. *Nat. Med.* **30**, 2489–2498 (2024).
22. Carrasco-Zanini, J. et al. Proteomic signatures for identification of impaired glucose tolerance. *Nat. Med.* **28**, 2293–2300 (2022).
23. Day, N. et al. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br. J. Cancer* **80**, 95–103 (1999).
24. Buergel, T. et al. Metabolomic profiles predict individual multidisease outcomes. *Nat. Med.* **28**, 2309–2320 (2022).
25. Garvey, W. T. et al. Tirzepatide once weekly for the treatment of obesity in people with type 2 diabetes (SURMOUNT-2): a double-blind, randomised, multicentre, placebo-controlled, phase 3 trial. *Lancet* **402**, 613–626 (2023).
26. Abubaker, J. et al. 208-OR: Comparable Improvement in NAFLD, weight loss, and lipid metabolism in rats treated with dual GLP-1–GIP receptor agonist (tirzepatide) and gastric sleeve. *Diabetes* **72**, 208-OR (2023).
27. Drucker, D. J. Expanding applications of therapies based on GLP1. *Nat. Rev. Endocrinol.* https://doi.org/10.1038/s41574-024-01066-9 (2024).
28. Wright, D. R., Guo, J. & Hernandez, I. A prescription for achieving equitable access to antiobesity medications. *JAMA Health Forum* **4**, e230493–e230493 (2023).
29. Anderer, S. Expanded Medicare coverage of antiobesity drugs may cost billions each year. *JAMA* **332**, 1133–1133 (2024).
30. Mahase, E. GLP-1 shortages will not resolve this year, EMA warns, amid concern over off-label use. *Brit. Med. J.* **385**, q1448 (2024).
31. Hernandez, I., Wright, D. R., Guo, J. & Shrank, W. H. Medicare Part D coverage of anti-obesity medications: a call for forward-looking policy reform. *J. Gen. Intern. Med.* **39**, 306–308 (2024).
32. Paddu, N. U., Lawrence, B., Wong, S., Poon, S. J. & Srivastava, G. Weight maintenance on cost-effective antiobesity medications after 1 year of GLP-1 receptor agonist therapy: a real-world study. *Obesity* **32**, 2255–2263 (2024).
33. Piché, M.-E., Tchernof, A. & Després, J.-P. Obesity phenotypes, diabetes, and cardiovascular diseases. *Circ. Res.* **126**, 1477–1500 (2020).

34. Schulze, M. B. & Stefan, N. Metabolically healthy obesity: from epidemiology and mechanisms to clinical implications. *Nat. Rev. Endocrinol.* **20**, 633–646 (2024).
35. Zembic, A., Eckel, N., Stefan, N., Baudry, J. & Schulze, M. B. An empirically derived definition of metabolically healthy obesity based on risk of cardiovascular and total mortality. *JAMA Netw. Open* **4**, e218505–e218505 (2021).
36. Coral, D. E. et al. Subclassification of obesity for precision prediction of cardiometabolic diseases. *Nat. Med.* https://doi.org/10.1038/s41591-024-03299-7 (2024).
37. Coral, D. E. et al. A phenome-wide comparative analysis of genetic discordance between obesity and type 2 diabetes. *Nat. Metab.* **5**, 237–247 (2023).
38. Fry, A. et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
39. Boyd, A. D. et al. Potential bias and lack of generalizability in electronic health record data: reflections on health equity from the National Institutes of Health Pragmatic Trials Collaboratory. *J. Am. Med. Inform. Assoc.* **30**, 1561–1566 (2023).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

Genes & Health Research Team

Claudia Langenberg[1,2,10], Sarah Finer[7] & David A. van Heel[1,8]

A full list of members and their affiliations appears in the Supplementary Information.

Methods

All studies were conducted according to all relevant ethical regulations and principles of the Declaration of Helsinki. UKB was granted ethical approval from the North West Centre for Research Ethics Committee (11/NW/0382), in the form of a Research Tissue Bank (RTB) approval. The EPIC-Norfolk study has been approved by the Norfolk Research Ethics Committee, under the reference number 05/Q0101/191. G&H was approved by the London Southeast NRES Committee of the Health Research Authority (14/LO/1240). SURMOUNT-1 was conducted according to the principles of the Declaration of Helsinki and Good Clinical Practice guidelines and was approved by independent ethics committees or institutional review boards at each of the 119 sites across nine countries[7].

Statistics and reproducibility

The UKB is a prospective, population-based observational study of ~500,000 individuals residing in the UK, aged 40–69 at enrollment, who were recruited between 2006 and 2010 across >20 centers[17]. All participants provided informed consent, and study procedures were performed according to the Helsinki declaration. This study was conducted with data accessed under the application numbers 44448 and 30418.

No statistical method was used to predetermine sample size. No randomization or blinding was applied in this observational study. To increase generalizability and reflect a scenario close to clinical reality, we applied exclusion criteria closely matching those of the SURMOUNT-1 trial[7] to define the study population (Fig. 1c). However, we did not exclude individuals with prevalent T2D because this population has been investigated separately in SURMOUNT-2[25], and we did not exclude individuals with weight change >5 kg within the last 90 days because this information was not reliably obtainable[7]. We included participants with a BMI equal or above 27 kg m^{-2} with or without prevalent comorbidities, and we excluded participants with contraindications for dual agonist therapies, with self-reported pregnancy or uncertainty about pregnancy or with diagnoses of thyroid cancer[7]. Additionally, we excluded participants with other potential causes of obesity: that is, owing to Cushing's syndrome or reported intake of tricyclic antidepressants, atypical antipsychotics, mood stabilizers or paroxetine. We excluded participants with a history of bariatric surgery, estimated glomerular filtration rate < 30 ml min^{-1} per 1.73 m^2, acute or chronic pancreatitis, alcoholism, recent cancer diagnosis (within the last 5 years) or recent MACE within the last 6 months. After applying these criteria, 197,264 participants remained for the final analyses. Analyses are based on distinct samples and did not include technical or biological replicates. All tests were two-sided.

Exposure

To establish a data-driven discovery of predictors of obesity-associated complications, we collated a comprehensive dataset of 2,390 exposure variables by leveraging multimodal data[40] (Supplementary Table 9).

We aggregated data on general characteristics of participants, including data on age, sex (self-reported), self-reported overall health, socioeconomic status, deprivation indices and related factors. We included behavior parameters, such as dietary habits and activity levels. To integrate medical history of the participants, we have gathered a comprehensive set of prevalent medical conditions for each participant from different data sources including self-reported conditions, primary care records where available (~45% of the study population), hospital episode statistics and cancer registries, as outlined previously[41]. We mapped these medical conditions onto 'phecodes'—that is, medical concept terms—as described[41]. We use the term disease/diagnosis rather 'phecode' for accessibility.

We included information on regular medication and health supplements from self-report, which were categorized based on the Anatomical Therapeutic Chemical classification system as described previously[41], yielding a total of 704 individual Anatomical Therapeutic Chemical codes. We have included a wide-ranging set of clinically established blood biomarkers and blood cell counts, the measurement of which have been described before[42]. We collated basic cardiopulmonary measurements including heart rate, blood pressure and spirometry parameters. In addition to basic body composition parameters (for example, weight, height, BMI, waist circumference, hip circumference and waist-to-hip and waist-to-height ratios), we considered bioelectrical impedance measures. We included metabolites measured with the Nightingale platform using the nuclear magnetic resonance method[43,44]. We included polygenic scores for a diverse set of 36 different diseases and traits, as provided by the UKB[45].

After filtering highly correlated variables ($r > 0.9$), whereby variables were kept either due to lower missingness or clinical availability, a total of 2,078 features were included in the predictor pool. We applied one-hot encoding to categorical variables, transforming them into binary indicator variables. We standardized continuous variables using *Z*-score transformation, where each value was centered by subtracting the variable mean and scaled by dividing by its standard deviation. We used multiple imputation to impute missing data using the miceRanger package (version 1.5.0), excluding variables with missingness above 25%, and generated a single dataset with five iterations[46].

Outcomes

We computed prediction models for 18 different obesity outcomes[3–5]. We set the recruitment date in UKB as the baseline of the observation period. Follow-up was censored at 10 years based on the time of event for the outcome or date of death as retrieved from the death register, whichever occurred first. Incident complications were defined based on phecode definitions as described in detail previously[41], from primary care where available, hospital episode statistics and self-report, except for T2D, for which we adapted a previously described definition to correctly distinguish between the different types, and for myocardial infarction and stroke, for which we adapted algorithmically defined outcomes provided by UKB[47] (Supplementary Table 10). For each outcome investigated, we removed prevalent cases (additionally using tests that were available at baseline, such as blood pressure measurements for hypertension). To minimize bias arising from delayed diagnosis, we excluded incident complications occurring within the first 6 months of follow-up. The number of cases based on availability of primary care data are shown in Supplementary Table 11.

Statistical analysis

Feature selection, optimization and internal validation. We employed a two-step machine learning procedure to identify sets of predictors consisting of the 20 most important features for each of the 18 complications, as recently applied to identify proteomic signatures predicting diseases[20,21]. This procedure involves first applying a least absolute shrinkage and selection operator (LASSO) model to identify the most relevant predictors from a large pool of features. In the second step, these selected predictors are used in a regularized Cox proportional hazards model for optimization and validation, providing a robust framework for feature selection and outcome prediction.

We randomly divided the cohort of eligible participants into three sets for each outcome: (1) feature selection (50%), (2) optimization (25%), and (3) validation (25%). In the feature selection set for each outcome, we applied LASSO with 250 random subsamples, each containing a fraction of 40% of the feature selection set. For each outcome, we compared different pools of possibly selectable features: that is, either by comparing each category (for example blood biomarkers, diseases and drugs and plasma metabolites, among others) of features individually or by increasing the pool of features by adding categories in a stepwise manner. For each outcome and from each set of feature pools, we took forward features with the top 20 highest feature selection scores—that is, those selected most frequently across subsamples—to be used in the

optimization and validation sets. Our approach, refined over iterations for similar feature selection use cases[20–22], is conceptionally equivalent to stability selection strategies[48]. In the optimization and test sets, we used a Cox proportional hazards model with L_2 regularization (ridge). We performed fivefold cross-validation on the optimization set to identify the optimal lambda parameter using a fine grid. For validation, we implemented a bootstrap procedure with 1,000 bootstrap samples to assess model performance on the independent test dataset. In each bootstrap sample, we calculated the C-index to serve as a discriminatory index for predictive performance. We took forward the average C-index and 95% CIs using the 2.5th and 97.5th percentiles of the bootstrap distribution.

Derivation and validation of a shared model (OBSCORE) for all outcomes. In a clinical setting, implementation of a single model with a core set of features that predict multiple complications would be most practical and hence more feasible to implement. Considering that multimorbid obesity—that is, the occurrence of multiple complications in a single individual–is common[3], the clinical scenario can be reframed as a multilabel classification problem. We handled this problem as separate binary classification tasks, building separate models for each outcome individually, as explained above. From each model, we extracted the top 20 clinically available features that were consistently selected: that is, with non-zero coefficients in the training set. In the optimization sets, after further pruning for correlated features ($r > 0.6$), we tested the feature importance of these features using the leave-one-out covariate importance procedure[49], and we subsequently ranked these features by mean importance across all 18 outcomes and took forward the top 20. This set of shared features represented a core group of predictors that showed relevance across multiple conditions. We derived 18 separate models to predict each outcome, each retrained using these 20 features in the optimization sets and tested in the validation sets.

External validation. To test for generalizability of the shared OBSCORE model, we additionally performed external validation in the EPIC-Norfolk study[20,23]. EPIC-Norfolk is a prospective cohort study of more than 25,000 middle-aged individuals from the general population of Norfolk, England.

In the EPIC-Norfolk study, mortality was retrieved from the UK Office of National Statistics, as described before[20,21,23]. Hospital episode statistics and certificates of death were retrieved from the NHS digital database, with National Health Service numbers. Health records were coded by trained nosologists according to the International Statistical Classification of Diseases and Related Health Problems, 9th (ICD-9) or 10th Revision (ICD-10), and codes were merged owing to the long-term follow-up. Events were determined in case of disease as underlying cause of death or as the reason for hospitalization. To allow for comparisons, we used the same ICD-10/ICD-9 definitions for baseline disease exclusion and incident disease definitions, except for T2D (ICD-9: 250, ICD-10: E10–E14) and stroke (ICD-9: 433–435, ICD-10: I63, I65, I66). For external validation, we included 2,112 individuals from the EPIC-Norfolk cohort with a BMI ≥ 27 and with complete data on predictor variables (18 out of 20 variables were available, except for cystatin C and joint pain). We conducted a complete case analysis. The variables underwent the same standardization processes as in the UKB, including Z-score transformation of continuous variables and logarithmic transformation of skewed variables. We were able to externally validate 14 outcomes with more than 20 incident cases during a 10-year follow-up: angina pectoris (155 cases), arthropathy (122), cholelithiasis (48), coronary atherosclerosis (143), cardiovascular death (94), diaphragmatic hernia (128), GERD (78), hypercholesterolemia (60), hypertension (112), ischemic heart disease (217), MACE (109), myocardial infarction (106), stroke (59) and T2D (114). We used coefficients (model weights) from the OBSCORE model trained with 18 available predictors and tested performance in external validation: that is, without retraining the models. We evaluated performance using C-indices in analogy to UKB analyses, calculated over 1,000 bootstrap samples. We excluded prevalent cases based on diagnoses or early incident cases during follow-up (within the first 6 months).

To further assess generalizability across ancestries, we performed external validation for incident T2D in the G&H study, a population at high risk for T2D and with sufficient incident events to support reliable analyses for this outcome. G&H is a prospective study enrolling individuals from self-reported British Pakistani and British Bangladeshi backgrounds and aged 16 or over since 2015. We curated and harmonized routine UK NHS EHR data from primary care and secondary care and death information from the Office for National Statistics[50]. We extracted first occurrences of diagnoses and closest quantitative measures from EHRs to study baselines. T2D was ascertained by EHR-based diagnoses as described before[51], and we used HbA1c > 48 mmol mol^{-1} to additionally exclude prevalent undiagnosed cases. A total of 1,740 individuals with complete data on 15 OBSCORE features, with a South Asian BMI cut-off of ≥22 kg m^{-2}, which corresponds to a BMI of ~27 kg m^{-2} in white populations, as informed by ref. 52, and without prevalent T2D at baseline (study enrollment), were included for external validation. A total of 337 cases occurred during a median follow-up of 6.3 years, excluding incident cases within occurring within 6 months after baseline. We used coefficients from the OBSCORE model trained with the 15 available features (excluding urate, cystatin C, family history of heart disease, joint and non-specific chest pain) in the UKB for external validation.

Performance comparison with established models. We compared this model with the best-performing outcome-specific models—a model containing age, sex and BMI—and to features of other established classifiers of obesity: for example, metabolically healthy obesity (age, sex and binary variable based on systolic blood pressure < 130 mmHg and no antihypertensive medication, waist-to-hip ratio < 0.95 for women or < 1.03 for men and no prevalent T2D) as well as the SCORE2 model (age, sex, smoking status, total cholesterol, high-density lipoprotein cholesterol, systolic blood pressure) and ASCVD risk score (age, sex, smoking status, ancestry, systolic blood pressure, diastolic blood pressure, total cholesterol, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, T2D, antihypertensive medication, cholesterol-lowering medication, acetylsalicylic acid intake)[35,53,54].

Calculation of performance metrics. To test how the shared clinical model performs at risk stratification, we computed the linear predictors for each outcome, resembling individual risk assigned by the prognostic model. We categorized the estimated risk by 20% (quintiles) of individuals for each outcome and calculated and visualized the actual observed 10-year event rate for each 20% group. We calculated risk ratios between the highest 20% and the lowest 20%. Moreover, we calculated rate ratios between the top 50% higher and lower risk groups and compared them to the rate ratios of individuals with obesity versus overweight. Using the estimated risk, we computed additional performance parameters including detection rates across FPRs ranging from 5% to 40%. FPRs were calculated as $\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN+FP}}$, where TN is true negatives; and detection rates (DRs) were calculated as $\mathrm{DR} = \frac{\mathrm{TP}}{\mathrm{FN+TP}}$, where TP is true positives and FN is false negatives. Likelihood ratios (LRs) were computed as $\mathrm{LR} = \frac{\mathrm{DR}}{\mathrm{FPR}}$. To assess the real-world translational potential of the models, we calculated pre-test and post-test probabilities. Pre-test probabilities were derived from cumulative complication incidence in the study, and pre-test odds were calculated as $\text{pretest odds} = \frac{\text{pretest probability}}{(1-\text{pretest probability})}$. Post-test probabilities at 10% FPR were calculated as $\text{post} - \text{test probability} = \frac{(\text{pretest odds} \times \mathrm{LR})}{1+(\text{pretest odds} \times \mathrm{LR})}$. We assessed calibration visually with calibration plots and by calculating ratios and differences between estimated and observed risks (from Kaplan–Meier estimates) in the held-out validation set.

OBSCORE integration in the randomized controlled SURMOUNT-1 trial. SURMOUNT-1 is a randomized controlled trial of participants with BMI ≥30 kg m^{-2} or ≥27 kg m^{-2} with at least one weight-related comorbidity treated with tirzepatide (5 mg, 10 mg, 15 mg) versus placebo (NCT04184622)[7]. We applied the OBSCORE risk model to individuals participating in SURMOUNT-1. A total of 12 features of OBSCORE (age, sex, waist-to-height ratio, hypertension, total cholesterol, high-density lipoprotein cholesterol, HbA1c, alanine aminotransferase, creatinine, cystatin C, urate, smoking status) were available at baseline and at 72 weeks. Coefficients from models retrained with these 12 features were applied to participants from SURMOUNT-1 with available data ($n_{placebo}$ = 388, $n_{TZP\,5mg}$ = 462, $n_{TZP\,10mg}$ = 479, $n_{TZP\,15mg}$ = 475) to calculate risk scores at baseline and at week 72. We first investigated the heterogeneity of treatment effects (body weight change, waist-to-height ratio change) across baseline predicted risk strata (quartiles) using linear regression models, whereby risk strata were coded numerically to test for linear trends. For each adiposity measure and risk score, we fit a model with treatment contrasts (placebo as reference). The model includes main effects for predicted risk strata and treatment arm, and their interaction term that tests whether treatment effects (versus placebo) change linearly across risk quartiles. EMMs for each treatment arm within each stratum were calculated using the emmeans package, with equal weight given to each stratum regardless of sample size. To test for changes in predicted risk from baseline to week 72 across all 18 outcomes, a one-way analysis of variance across treatment arms was employed, followed by unadjusted pairwise contrasts of the group means.

Subgroup and sensitivity analyses. To investigate performance of OBSCORE in different subgroups, we evaluated the performance of OBSCORE separately in individuals of genetically inferred non-European (n = 9,668) and European (sample size matched) ancestries as well as in individuals with the lower 50% and higher 50% of the Townsend deprivation index (median of the study population −2.05), without retraining in the subgroups. To test whether discriminative performance may be influenced by early incident cases that represent delayed diagnosis, we investigated C-indices of OBSCORE when excluding early incident cases additionally at 1 year, 1.5 years and 2 years after baseline for each outcome.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All individual-level data are publicly available to bona fide researchers upon application from the UKB (https://www.ukbiobank.ac.uk/), EPIC-Norfolk (https://www.epic-norfolk.org.uk/for-researchers/data-sharing/data-requests/) and G&H (https://www.genesandhealth.org/). SURMOUNT-1: Lilly provides access to all individual participant data collected during the trial, after anonymization, with the exception of pharmacokinetic or genetic data. Data are available on request 6 months after the indication studied has been approved in the USA and EU and after primary publication acceptance, whichever is later. No expiration date of data requests is currently set once data are made available. Access is provided after a proposal has been approved by an independent review committee identified for this purpose and after receipt of a signed data sharing agreement. Data and documents, including the study protocol, statistical analysis plan, clinical study report, blank or annotated case report forms, will be provided in a secure data sharing environment. For details on submitting a request, see the instructions provided at www.vivli.org. Source data are provided with this paper.

Code availability

Statistical code is available at https://github.com/comp-med/obesity-prognosis.

References

40. Pietzner, M. et al. Machine learning-guided deconvolution of plasma protein levels. *Mol. Syst. Biol.* **21**, 1822–1844 (2025).
41. Pietzner, M. et al. Complex patterns of multimorbidity associated with severe COVID-19 and long COVID. *Commun. Med. (Lond.)* **4**, 94 (2024).
42. Sinnott-Armstrong, N. et al. Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet.* **53**, 185–194 (2021).
43. Nightingale Health Biobank Collaborative Group. Metabolomic and genomic prediction of common diseases in 700,217 participants in three national biobanks. *Nat. Commun.* **15**, 10092 (2024).
44. Zoodsma, M. et al. A genetic map of human metabolism across the allele frequency spectrum. *Nat. Genet.* **57**, 2445–2455 (2025).
45. Thompson, D. J. et al. A systematic evaluation of the performance and properties of the UK Biobank Polygenic Risk Score (PRS) Release. *PLoS ONE* **19**, e0307270 (2024).
46. Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O. & Hemingway, H. Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study. *Am. J. Epidemiol.* **179**, 764–774 (2014).
47. Gardner, E. J. et al. Damaging missense variants in IGF1R implicate a role for IGF-1 resistance in the etiology of type 2 diabetes. *Cell Genom.* https://doi.org/10.1016/j.xgen.2022.100208 (2022).
48. Bodinier, B., Filippi, S., Nøst, T. H., Chiquet, J. & Chadeau-Hyam, M. Automated calibration for stability selection in penalised regression and graphical models. *J. R. Stat. Soc. Ser. C Appl. Stat.* **72**, 1375–1393 (2023).
49. Verdinelli, I. & Wasserman, L. Decorrelated variable importance. *J. Mach. Learn. Res.* **25**, 1–27 (2024).
50. Jacobs, B. M. et al. Genetic architecture of routinely acquired blood tests in a British South Asian cohort. *Nat. Commun.* **15**, 8929 (2024).
51. Hodgson, S. et al. Genetic basis of early onset and progression of type 2 diabetes in South Asians. *Nat. Med.* **31**, 323–331 (2025).
52. Caleyachetty, R. et al. Ethnicity-specific BMI cutoffs for obesity based on type 2 diabetes risk in England: a population-based cohort study. *Lancet Diab. Endocrinol.* **9**, 419–426 (2021).
53. Score working group and ESC Cardiovascular risk collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur. Heart J.* **42**, 2439–2454 (2021).
54. Goff, D. C. Jr et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation* **129**, S49–S73 (2014).

Acknowledgements

We acknowledge the Scientific Computing Department of the IT Division at the Charité – Universitätsmedizin Berlin for providing computational resources. We are grateful to the participants and study teams of UKB, EPIC-Norfolk, Genes & Health and the SURMOUNT-1 trial. This work was supported by a BHF (British Heart Foundation) – DZHK (German Centre for Cardiovascular Research) grant (grant Prot4CVD 81X2100281), a UKRI/NIHR Strategic Priorities Award in Multimorbidity Research for the Multimorbidity Mechanism and Therapeutics Research Collaborative (grant MR/V033867/1) grant to C.L. and funded by the European Union (ERC Consolidator Grant D-MAPS to C. Langenberg, grant 101231435). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. The work was supported by a Walter Benjamin Fellowship (grant 547107463) by the DFG (German Research Foundation) to K.D., the Friede Springer

Cardiovascular Prevention Center at Charité - Universitätsmedizin Berlin, Germany (A.W.) and the de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) and ELIXIR-DE (Forschungszentrum Jülich and grant nos. W-de.NBI-001, W-de.NBI-004, W-de.NBI-008, W-de.NBI-010, W-de.NBI-013, W-de.NBI-014, W-de.NBI-016, W-de.NBI-022) for webserver development support. We thank the EPIC-Norfolk investigators, the Study Co-ordination team, the Epidemiology Field, Data and Laboratory teams. The EPIC-Norfolk study (https://doi.org/10.22025/2019.10.105.00004) has received funding from the Medical Research Council (grants MR/N003284/1 and MC_UU_00006/1 to N.J.W.) and Cancer Research UK (grant C864/A14136 to N.J.W.). Genes & Health is/has recently been core-funded by Wellcome (grants WT102627, WT210561), the Medical Research Council (UK) (grants M009017, MR/X009777/1, MR/X009920/1), Higher Education Funding Council for England Catalyst, Barts Charity (grant 845/1796), Health Data Research UK (for London substantive site) and research delivery support from the NHS National Institute for Health Research Clinical Research Network (North Thames). We acknowledge the support of the National Institute for Health and Care Research Barts Biomedical Research Centre (grant NIHR203330); a delivery partnership of Barts Health NHS Trust, Queen Mary University of London, St George's University Hospitals NHS Foundation Trust and St George's University of London. Genes & Health is/has recently been funded by Alnylam Pharmaceuticals, Genomics PLC; and a Life Sciences Industry Consortium of AstraZeneca PLC, Bristol-Myers Squibb Company, GlaxoSmithKline Research and Development Limited, Maze Therapeutics Inc., Merck Sharp & Dohme LLC, Novo Nordisk A/S, Pfizer Inc., and Takeda Development Centre Americas Inc. We thank Social Action for Health, Centre of The Cell, members of our Community Advisory Group and staff who have recruited and collected data from volunteers. We thank the NIHR National Biosample Centre (UK Biocentre), the Social Genetic & Developmental Psychiatry Centre (King's College London), Wellcome Sanger Institute and Broad Institute for sample processing, genotyping, sequencing and variant annotation. We thank Barts Health NHS Trust, NHS Clinical Commissioning Groups, East London NHS Foundation Trust, Bradford Teaching Hospitals NHS Foundation Trust, Public Health England (D. Wyllie), Discovery Data Service/Endeavour Health Charitable Trust (D. Stables), Voror Health Technologies Ltd (S. Don), NHS England for GDPR-compliant data sharing backed by individual written informed consent. This work uses data provided by patients and collected by the NHS as part of their care and support. We are grateful to the participants and study teams of UKB, EPIC-Norfolk, Genes & Health and the SURMOUNT-1 trial. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

Conceptualization: K.D., C.L. Conceptualization (specific elements): K.D., J.C.-Z., A.W., L.J., M.C., I.M., M.P., C.L. Data curation and software: K.D., J.C.-Z., C.B., L.J., W.R.-M., M.P. Formal analysis: K.D., J.C.-Z., C.B., L.J., M.P. Methodology: K.D., J.C.-Z., A.W., C.B., L.J., M.C., I.M., A.L.H., N.J.W., M.P., C.L. Visualization: K.D., J.C.-Z., C.B., L.J. Funding acquisition: K.D., A.W., N.J.W., S.F., D.A.v.H., M.P., C.L. Project administration: G.K., N.J.W., M.P., C.L. Supervision: J.C.-Z., M.P., C.L. Writing—original draft: K.D., J.C.-Z., M.P., C.L. Writing—review and editing: K.D., J.C.-Z., L.J., A.W., C.B., A.L.H., W.R.-M., G.K., S.F., D.A.v.H., M.C., I.M., N.J.W., M.P., C.L.

Competing interests

L.J., M.C. and I.M. are employees and stockholders of Eli Lilly and Company. Eli Lilly provided no funding for this study and had no influence on the overall analytical plan, conduct or interpretation of main results. L.J. performed analyses of the SURMOUNT-1 trial, according to an analytical plan jointly designed by C.L., K.D., M.C., I.M. and L.J. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41591-026-04353-2.

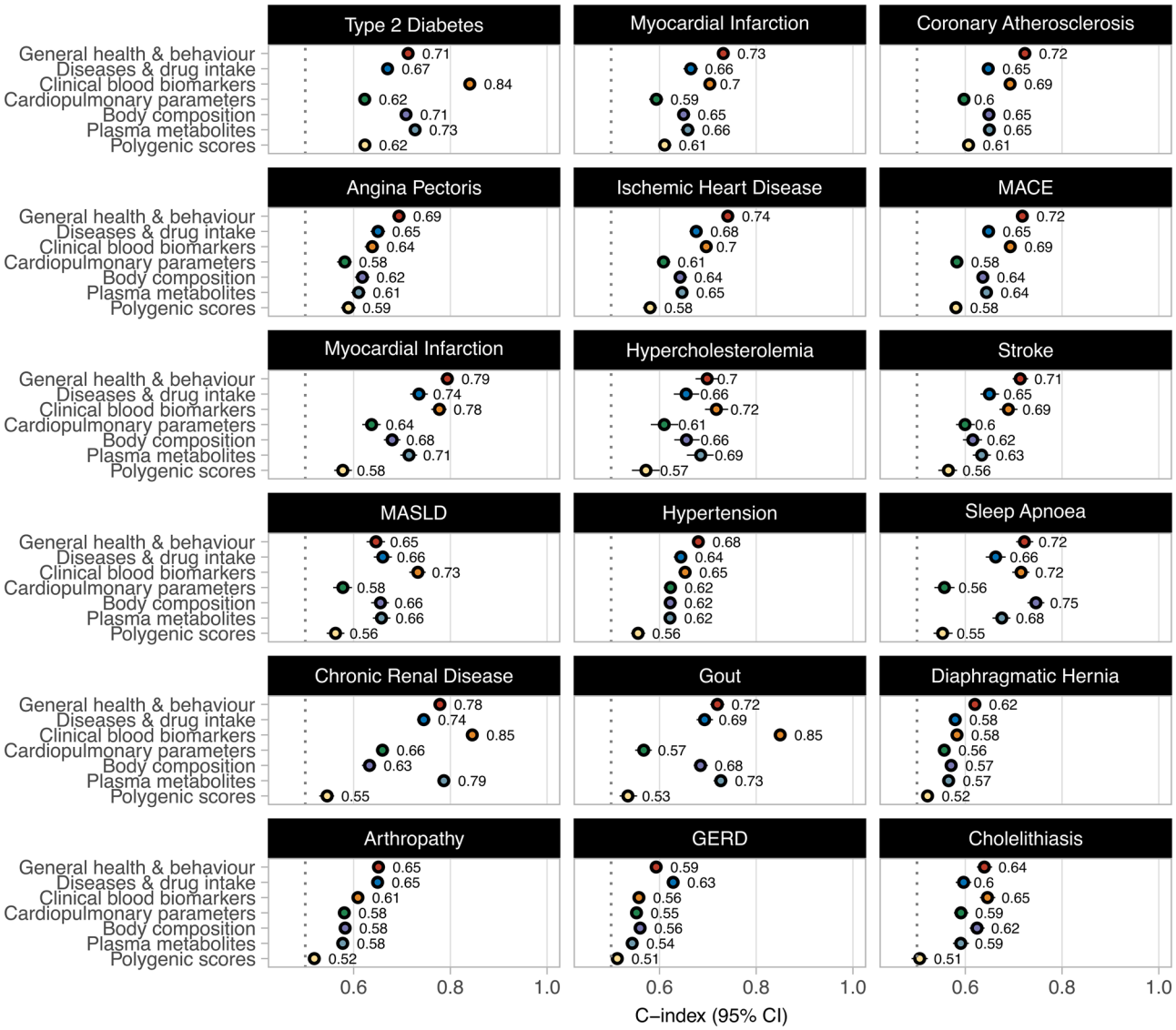
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-026-04353-2.

Correspondence and requests for materials should be addressed to Claudia Langenberg.

Peer review information *Nature Medicine* thanks Louis Aronne, Ilaria Gandin and Francesco Giorgino for their contribution to the peer review of this work. Primary Handling Editor: Ashley Castellanos-Jankiewicz, in collaboration with the *Nature Medicine* team.

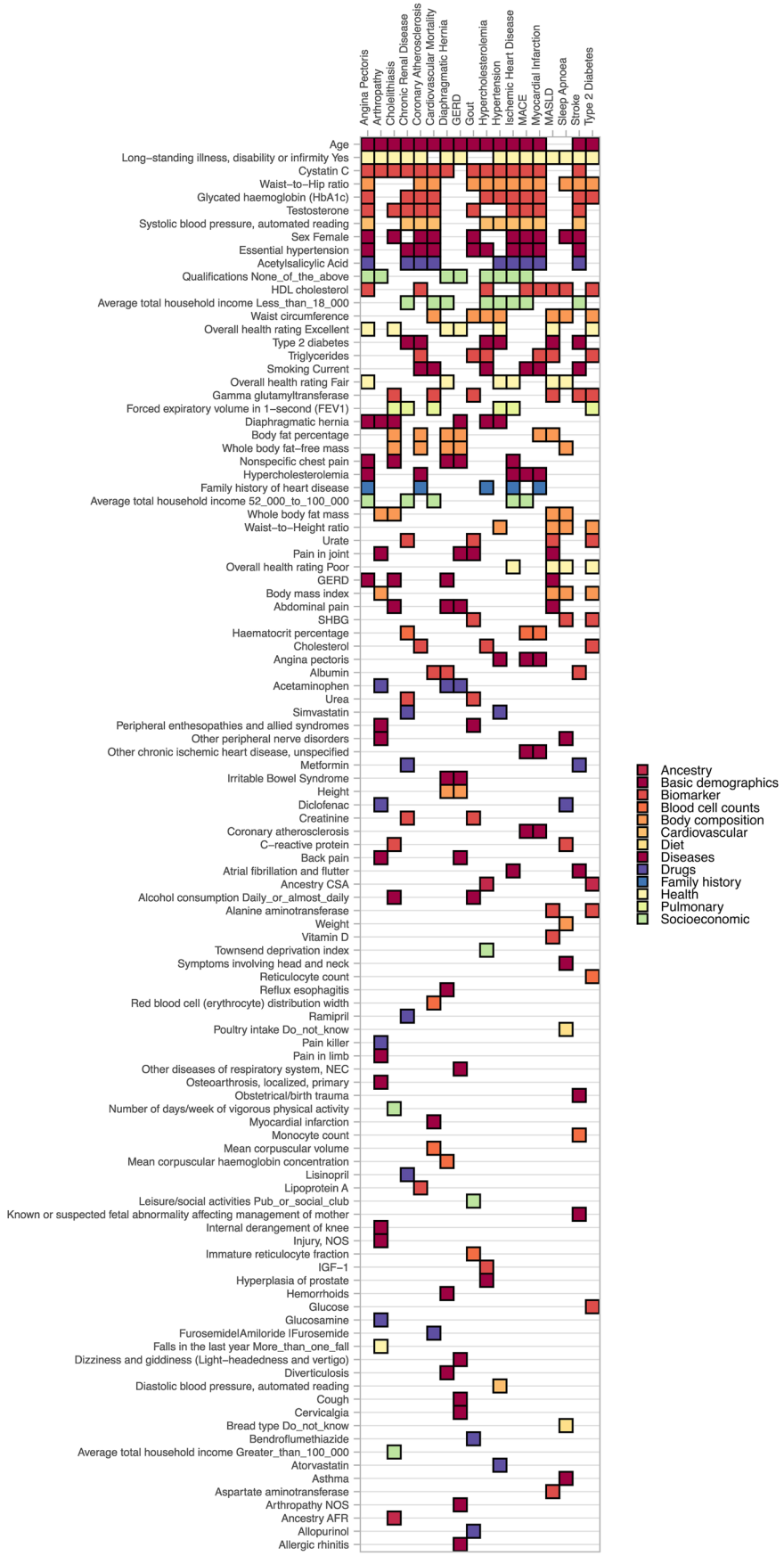
Reprints and permissions information is available at www.nature.com/reprints.

https://doi.org/10.1038/s41591-026-04353-2



Extended Data Fig. 1 | Performances of unimodal outcome-specific models. Outcome-specific models were built using the LASSO algorithm, selecting from up to 20 features from different categories of predictors individually, that is, indicators of general health and behavior, diseases and drug intake, clinical blood biomarkers, cardiopulmonary parameters, body composition, plasma metabolites, and polygenic scores. The Concordance-index (C-index) with 95% confidence intervals computed using 1,000 bootstrap resamples is plotted for each outcome. Vertical dashed line at C-index = 0.5 denotes a random predictor. Supplementary Table 2 contains sample size information.

https://doi.org/10.1038/s41591-026-04353-2



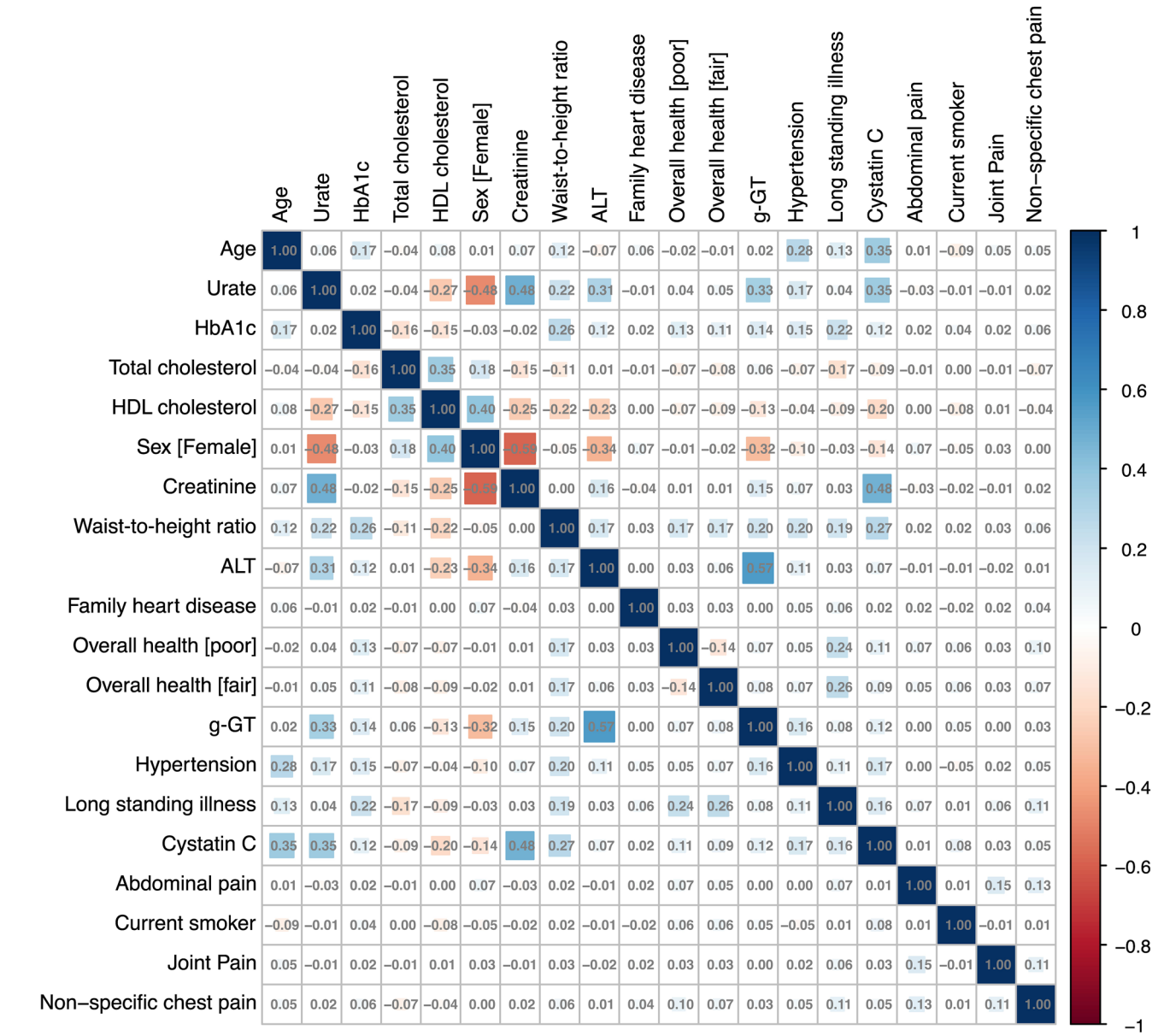
Extended Data Fig. 2 | Features from clinical models. Top 20 features selected in each multimodal outcome-specific model are depicted, filled by category of the feature. The pool of features includes general health & behaviour, diseases and medication, clinical blood biomarkers, cardiopulmonary parameters and body composition.

https://doi.org/10.1038/s41591-026-04353-2



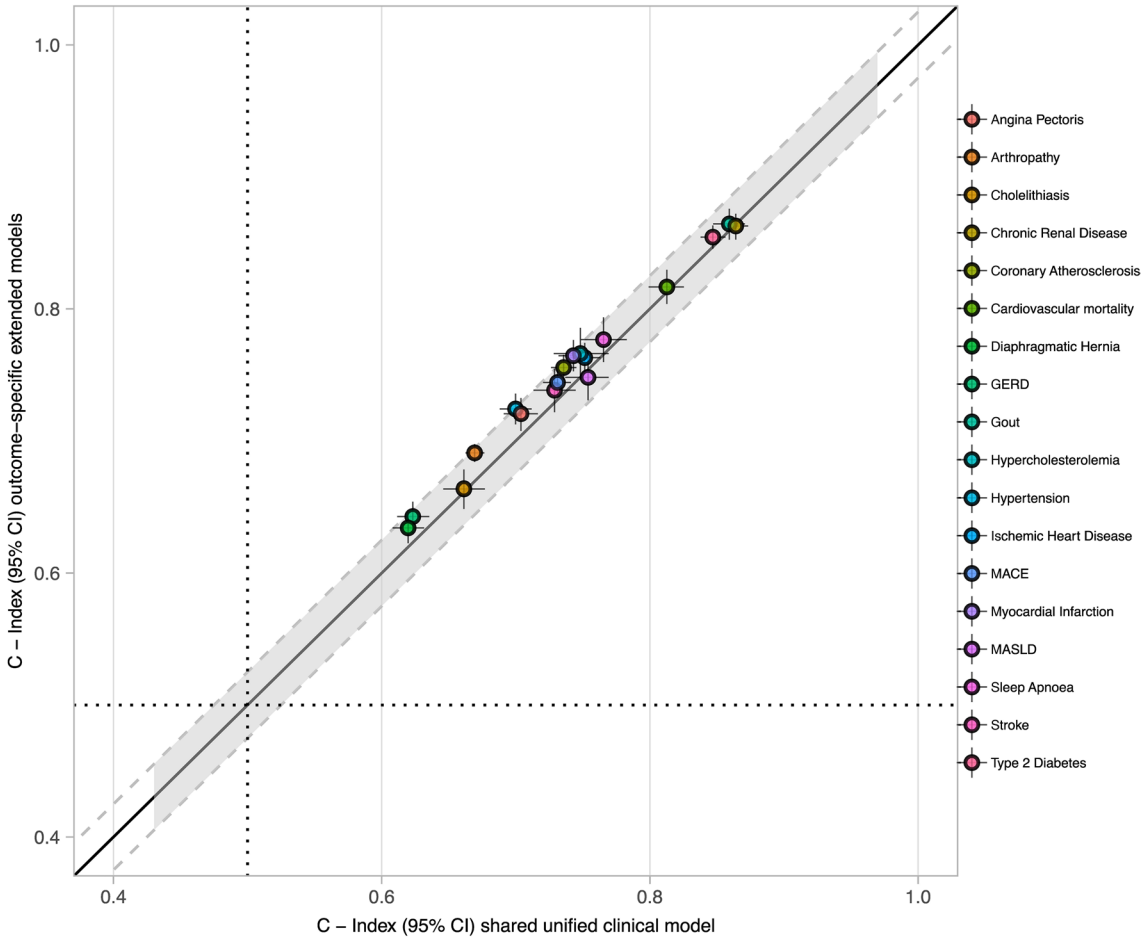
Extended Data Fig. 3 | Features from extended models. Top 20 features selected in each multimodal outcome-specific model are depicted, filled by category of the feature. The pool of features includes general health & behavior, diseases and medication, clinical blood biomarkers, cardiopulmonary parameters and body composition, plasma metabolites, and polygenic scores.

https://doi.org/10.1038/s41591-026-04353-2



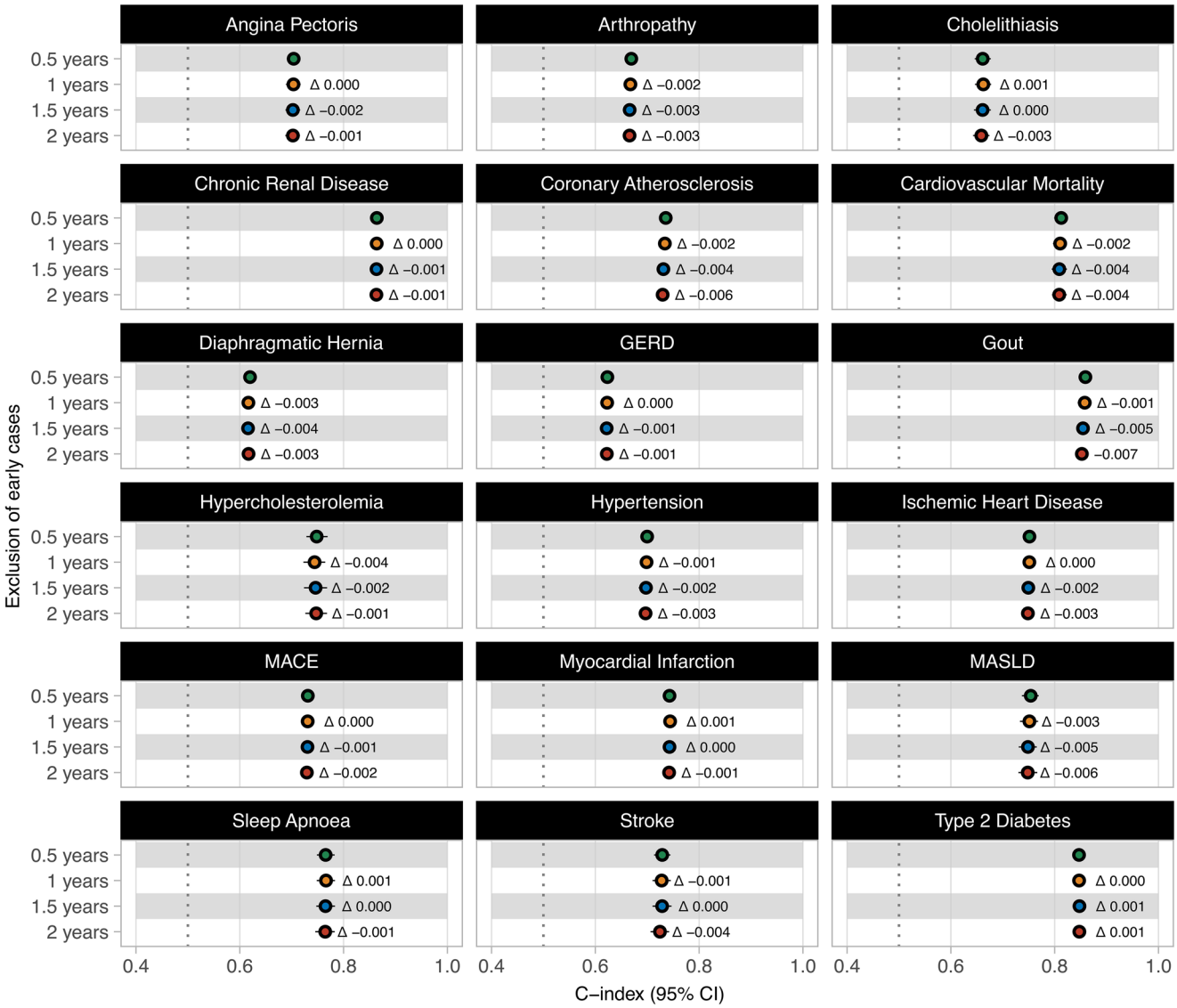
Extended Data Fig. 4 | Correlation matrix of features in OBSCORE. Two-sided Pearsons' correlation test was used to compute correlation coefficients.

https://doi.org/10.1038/s41591-026-04353-2



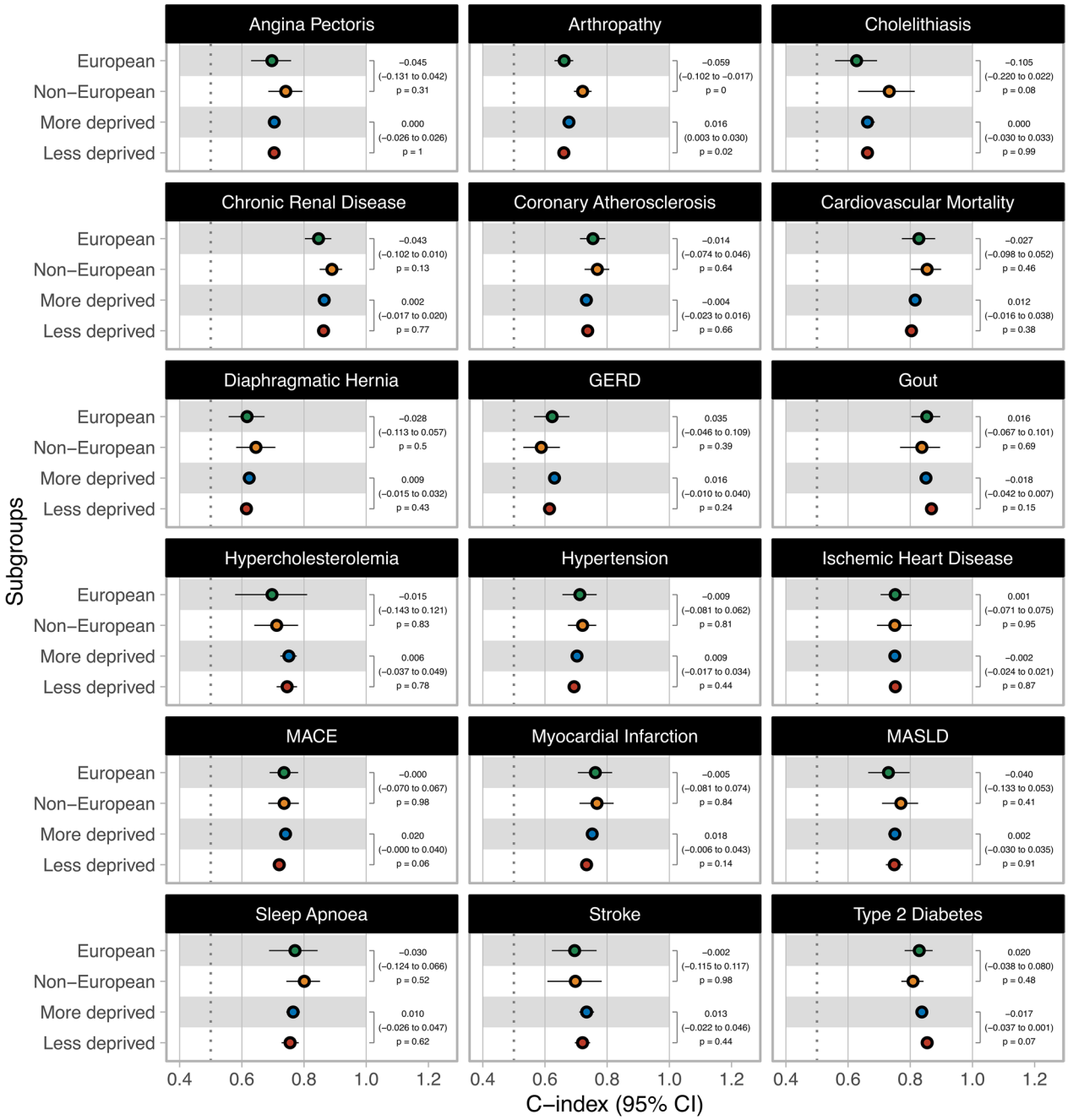
Extended Data Fig. 5 | OBSCORE versus extended outcome-specific models. The performances (expressed as Concordance index with 95% confidence intervals computed using 1000 bootstrap resamples) of the shared model (X-axis) are compared with performances of extended outcome-specific models (Y-axis). Dashed lines represent 2.5% deviation intervals. Dotted lines represent a random predictor (C-index = 0.5).

https://doi.org/10.1038/s41591-026-04353-2



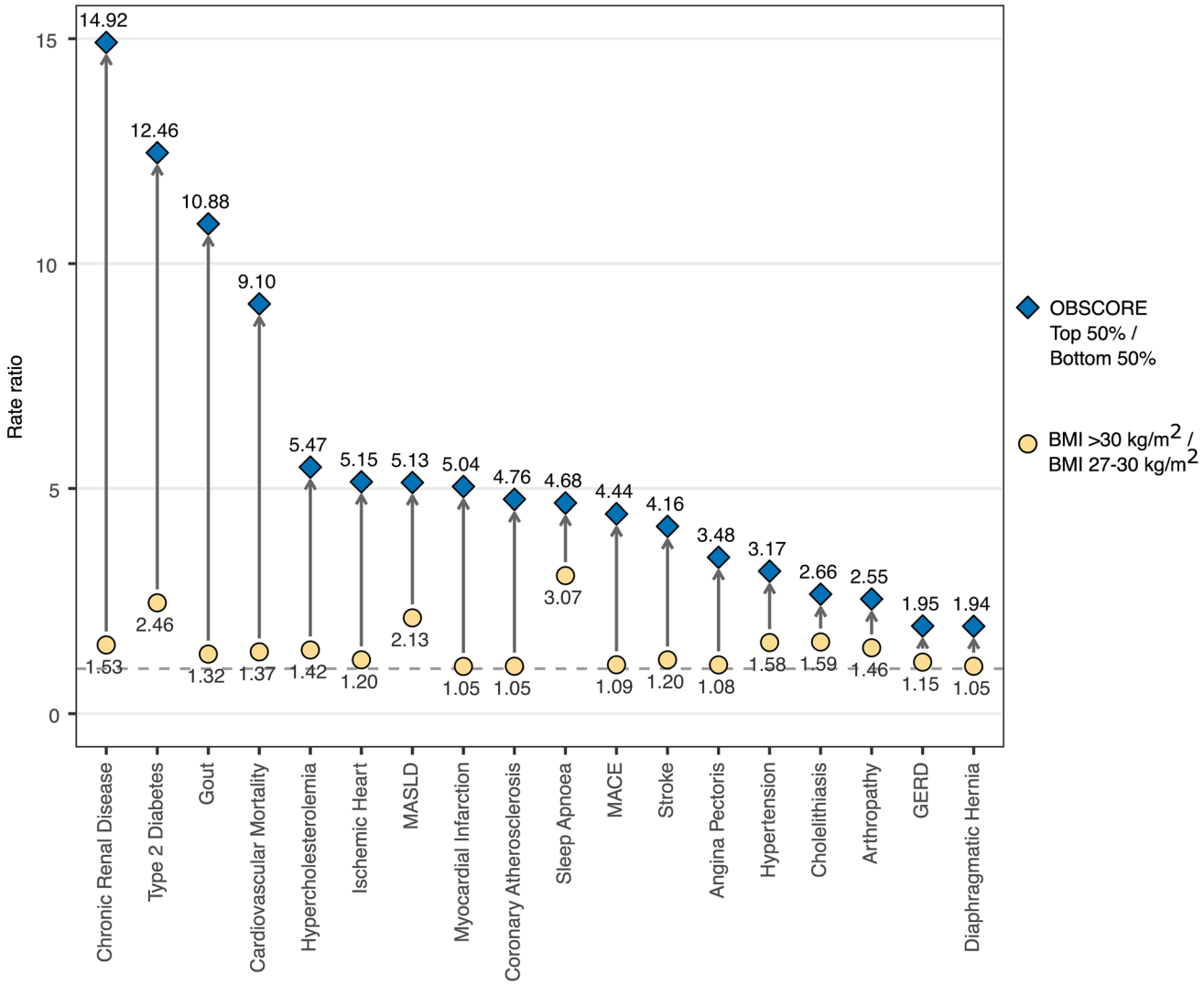
Extended Data Fig. 6 | Sensitivity analyses with extended exclusion of early incident cases. In addition to the main analyses excluding incident cases occurring in the first 6 months, performances (expressed as Concordance index (C-index) with 95% confidence intervals estimated using 1,000 bootstrap resamples) of OBSCORE were tested when excluding cases occurring within 1, 1.5, and 2 years after baseline.

https://doi.org/10.1038/s41591-026-04353-2



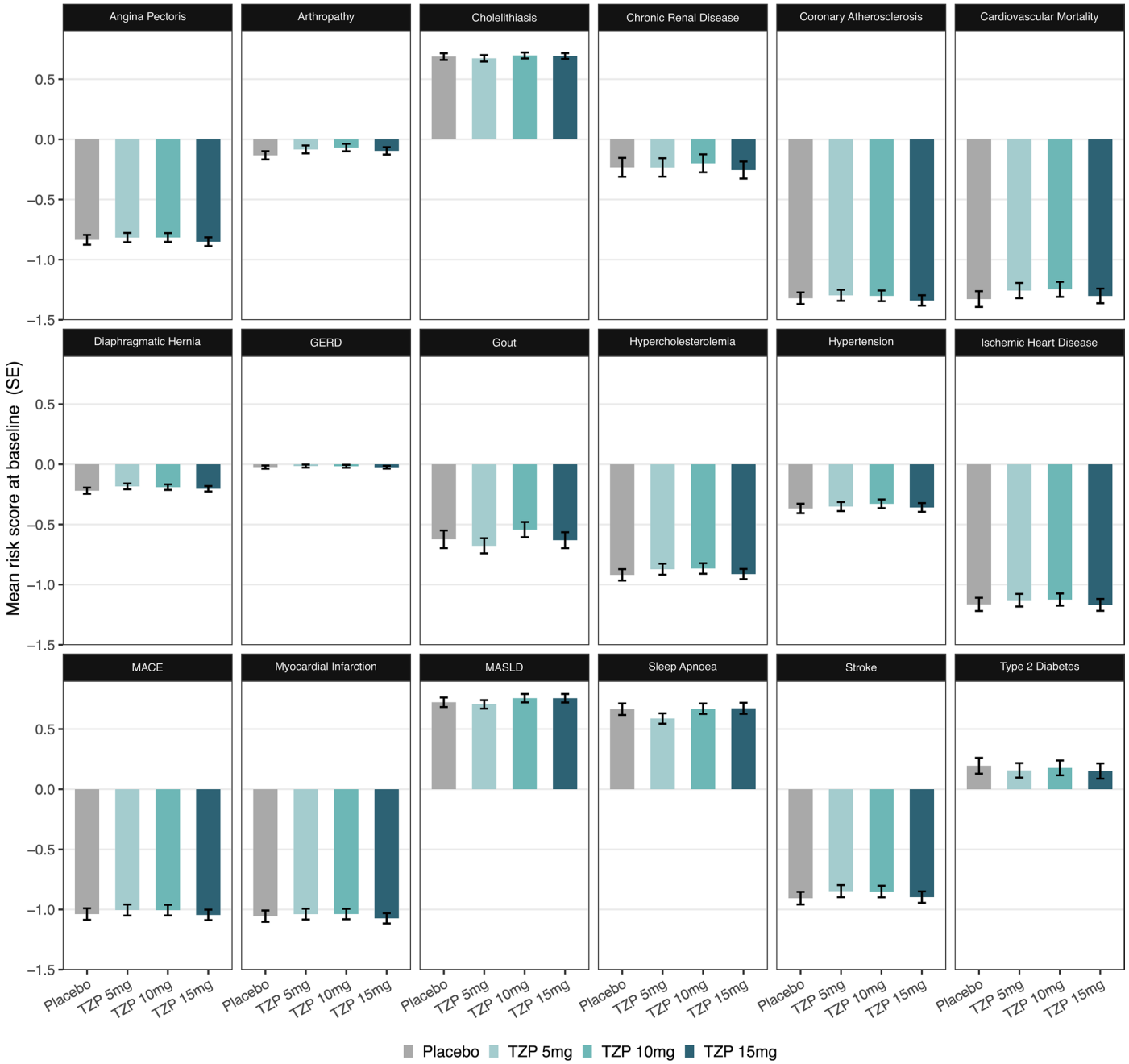
Extended Data Fig. 7 | Subgroup analyses in the UK Biobank. Performances of OBSCORE in individuals of European and non-European ancestry and lower and higher deprivation subgroups (cut-off cohort median of Townsend deprivation index, corresponding to −2.05). Highlighted are mean C-indices and differences between subgroups with differences 95% confidence intervals and p-values estimated from comparing 1000 bootstrap samples.

https://doi.org/10.1038/s41591-026-04353-2



Extended Data Fig. 8 | Risk stratification comparing OBSCORE and obesity/overweight classification. Ratios of number of observed events in individuals with 50% highest predicted OBSCORE vs. lowest 50% predicted OBSCORE risks and ratios of number of observed events in individuals living with obesity vs individuals with overweight, across outcomes.

https://doi.org/10.1038/s41591-026-04353-2



Extended Data Fig. 9 | Baseline OBSCORE (12 features) risk scores across treatment arms and outcomes in the SURMOUNT-1 trial. Mean risk scores at baseline with standard errors of the mean are shown. $n_{placebo} = 388$, $n_{TZP\ 5mg} = 462$, $n_{TZP\ 10mg} = 479$, $n_{TZP\ 15mg} = 475$.

https://doi.org/10.1038/s41591-026-04353-2

Extended Data Table 1 | Study participant characteristics

	Overall N = 197,264	**BMI 27-30** kg/m^2 N = 97,421	**BMI >30** kg/m^2 N = 99,843
Age (years)	58 (12)	59 (13)	58 (12)
Sex			
Female	95,034 (48%)	43,328 (44%)	51,706 (52%)
Male	102,230 (52%)	54,093 (56%)	48,137 (48%)
BMI category			
Overweight	97,421 (49%)	97,421 (100%)	0 (0%)
Obesity Class I	72,271 (37%)	0 (0%)	72,271 (72%)
Obesity Class II	20,046 (10%)	0 (0%)	20,046 (20%)
Obesity Class III	7,526 (3.8%)	0 (0%)	7,526 (7.5%)
Type 2 diabetes	18,284 (9.3%)	5,530 (5.7%)	12,754 (13%)
HbA1c (mmol/mol)	36 (6)	35 (5)	37 (6)
Systolic blood pressure (mm Hg)	142 (25)	141 (25)	143 (25)
Diastolic blood pressure (mm Hg)	85 (14)	83 (13)	86 (14)
Ancestry			
African	4,039 (2.0%)	1,567 (1.6%)	2,472 (2.5%)
Admixed American	406 (0.2%)	196 (0.2%)	210 (0.2%)
Central/South Asian	3,749 (1.9%)	1,987 (2.0%)	1,762 (1.8%)
East Asian	628 (0.3%)	417 (0.4%)	211 (0.2%)
European	187,596 (95%)	92,877 (95%)	94,719 (95%)
Middle Eastern	846 (0.4%)	377 (0.4%)	469 (0.5%)
Townsend deprivation index	-2.05 (4.23)	-2.29 (3.89)	-1.79 (4.54)
Smoking status			
Never	102,472 (52%)	51,165 (53%)	51,307 (51%)
Previous	76,071 (39%)	36,776 (38%)	39,295 (39%)
Current	18,622 (9.4%)	9,443 (9.7%)	9,179 (9.2%)
Alcohol consumption			
Never	15,609 (7.9%)	6,570 (6.7%)	9,039 (9.1%)
Special occasions only	24,700 (13%)	9,798 (10%)	14,902 (15%)
One to three times a month	23,765 (12%)	10,298 (11%)	13,467 (13%)
Once or twice a week	52,651 (27%)	25,907 (27%)	26,744 (27%)
Three or four times a week	44,090 (22%)	24,158 (25%)	19,932 (20%)
Daily or almost daily	36,350 (18%)	20,653 (21%)	15,697 (16%)

Median (IQR); n (%)

nature portfolio

Corresponding author(s): Claudia Langenberg

Last updated by author(s): 25.02.2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Data analyses were performed in R (version 4.3.1.), using following packages arrow (16.1.0), caret (6.0-94), corrplot (0.95), data.table (1.17.0), doMC (1.3.8), doParallel (1.0.17), dplyr (1.1.4), fastDummies (1.7.5), flextable (0.9.6), foreach (1.5.2), ggplot2 (3.5.1), ggradar (0.2), ggrepel (0.9.5), ggridges (0.5.6), ggsci (3.2.0), glmnet (4.1-8), gtsummary (2.0.3), igraph (2.1.4), influential (2.2.9), magrittr (2.0.3), miceRanger (1.5.0), patchwork (1.3.0), Rfast (2.1.0), ROSE (0.0-4), shiny (1.9.1), shinyBS (0.61.1), shiny themes (1.2.0), sna (2.8), survival (3.8-3), survminer (0.4.9), tidyr (1.3.1), tidyverse (2.0.0). Analysis code is available on https://github.com/comp-med/obesity-prognosis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All individual-level data is publicly available on application to bona fide researchers from the UKB (https://www.ukbiobank.ac.uk/), EPIC-Norfolk (https://www.epic-norfolk.org.uk/for-researchers/data-sharing/data-requests/), and Genes & Health (https://www.genesandhealth.org/). For UKB, this research has been conducted under the applications 44448 and 30418. SURMOUNT-1: Lilly provides access to all individual participant data collected during the trial, after anonymization, with the exception of pharmacokinetic or genetic data. Data are available on request 6 months after the indication studied has been approved in the US and EU and after primary publication acceptance, whichever is later. No expiration date of data requests is currently set once data are made available. Access is provided after a proposal has been approved by an independent review committee identified for this purpose and after receipt of a signed data sharing agreement. Data and documents, including the study protocol, statistical analysis plan, clinical study report, blank or annotated case report forms, will be provided in a secure data sharing environment. For details on submitting a request, see the instructions provided at www.vivli.org.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	In UKB We defined 'female' and 'male' sex including participants where the recorded sex and sex chromosomes aligned (XX for females and XY for males). The recorded sex was self-reported, and it was not possible to distinguish sex from gender. We acknowledge the importance of distinguishing between sex and gender in research and that chromosomal make-up does not always align with self-identified gender.
Reporting on race, ethnicity, or other socially relevant groupings	We used previously published ancestral assignments by the pan-UKB consortium to assign individuals to ancestral groups, and made a further effort to assign unclassified individuals to their respective ancestries based on a k-nearest neighbour approach using genetic principal components.
Population characteristics	UK Biobank is a prospective cohort study from the UK that contains more than 500,000 volunteers between 40 and 69 years of age at inclusion. The cohort has been extensively described elsewhere (www.ukbiobank.ac.uk). We included 197,264 individuals based on criteria of recent weight-loss medication trials, e.g. excluding BMI <27 kg/m^2, pregnancy, and others.
Recruitment	All individuals between the age of 40-69 (men and women) who were registered with the National Health Service and living within a 25-mile radius from one of 22 recruitment centers spread across the United Kingdom were invited to participate in 2006-2010. Overall, about 9.2M individuals were invited to recruit around 0.5M individuals.
Ethics oversight	UKB was granted ethical approval from the North West Centre for Research Ethics Committee (11/NW/0382), in the form of a Research Tissue Bank (RTB) approval. All participants provided informed consent, and study procedures were performed according to the Helsinki declaration. This study was conducted with data accessed under the application numbers 44448 and 30418. The EPIC-Norfolk study has been approved by the Norfolk Research Ethics Committee, under the reference number 05/Q0101/191. G&H was approved by the London Southeast NRES Committee of the Health Research Authority (14/LO/1240). SURMOUNT-1 was conducted according to the principles of the Declaration of Helsinki and Good Clinical Practice guidelines and was approved by independent ethics committees or institutional review boards at each of the 119 sites across nine countries.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	UKB: We included 197,264 individuals with BMI>=27 kg/m^2 EPIC-Norfolk: We included 2,112 individuals with complete data on available OBSCORE features. Genes & Health: We included 1,740 individuals with complete data on available OBSCORE features. SURMOUNT-1: We included up to 1,804 individuals (placebo= 388, TZP 5mg = 462, TZP 10mg = 479, TZP 15mg = 475). This was an exploratory, non-prespecified analysis of individual level data in SURMOUNT-1.
Data exclusions	UKB: We excluded individuals based on criteria of recent weight-loss medication trials, e.g. BMI <27 kg/m^2, pregnancy, and others.

Data exclusions	Moreover, we excluded individuals where we did not have access to matching genotyping (array or sequencing-based) and polygenic score data. Furthermore, individual samples failing standard genotyping quality control or not assigned to one of the three major ancestral groups were excluded. These decisions were made before performing any statistical analysis. Participants were also excluded based on BMI<27 kg/m^2 in EPIC-Norfolk and Genes & Health.
Replication	UKB: We divided the dataset into test, optimisation and validation sets to derive the model and test (replicate) the models in internal validation. External validation: Additionally, to display generalisability of the OBSCORE model, we performed external validation in a non-overlapping external study, European Investigation into Cancer - Norfolk (EPIC-Norfolk) study. We were able to validate performance and generalisability for 14 out of 18 outcomes utilising 18 of the 20 features included in OBSCORE. The validation study displayed high generalisability and performance of the model in an external study. We were able to test performance and generalisability of OBSCORE for incident T2D additionally in the Genes & Health study, which is exclusively enrolling individuals with self-reported Bangladeshi or Pakistani backgrounds.
Randomization	UKB/EPIC-Norfolk/Genes & Health: N/A. SURMOUNT-1: Participants were randomly assigned in a 1:1:1:1 ratio to receive tirzepatide at a dose of 5 mg, 10 mg, 15 mg, or placebo.
Blinding	UKB/EPIC-Norfolk/Genes & Health: N/A. SURMOUNT-1: Double blinded in the original study, no blinding during these analyses.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	NCT04184622
Study protocol	See protocol for: Jastreboff AM, Aronne LJ, Ahmad NN, et al. Tirzepatide once weekly for the treatment of obesity. N Engl J Med 2022;387:205-16. (https://www.nejm.org/doi/suppl/10.1056/NEJMoa2206038/suppl_file/nejmoa2206038_protocol.pdf)
Data collection	Described in Jastreboff AM, Aronne LJ, Ahmad NN, et al. Tirzepatide once weekly for the treatment of obesity. N Engl J Med 2022;387:205-16.
Outcomes	We performed a non-prespecified exploratory analysis of the SURMOUNT-1 trial.

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*