Research Square

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

Comorbidity Patterns and Temporal Pathways in Metabolic Multimorbidity: A Five-Year Retrospective Cohort Study of Dyslipidemia, Hyperuricemia, Obesity, and Diabetes

Lele Deng
Beijing Chao-Yang Hospital, Capital Medical University
Xiaoyi Jiang
Beijing Chao-Yang Hospital, Capital Medical University
Ruijun Li
Beijing Chao-Yang Hospital, Capital Medical University
Hang Zhao
Beijing Chao-Yang Hospital, Capital Medical University
Mengyu Yuan
Beijing Chao-Yang Hospital, Capital Medical University
Chenxu Wang
Beijing Chao-Yang Hospital, Capital Medical University
Rui Zhao
Beijing Chao-Yang Hospital, Capital Medical University
Yanwei Hu
Beijing Chao-Yang Hospital, Capital Medical University
Gaowei Fan
fangw890163.com

Beijing Chao-Yang Hospital, Capital Medical University

Research Article

Keywords: metabolic multimorbidity, comorbidity network, dyslipidemia, obesity, type 2 diabetes, hyperuricemia, trajectory modeling, multi-state Markov model, time-varying Cox regression, latent class analysis

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11049427/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background: Dyslipidemia, hyperuricemia, obesity, and type 2 diabetes frequently co-occur, yet the structural organization, temporal ordering, and directional predictive relationships among these conditions remain poorly characterized.

Methods: This retrospective cohort study included 8,383 adults undergoing annual health examinations at Beijing Chao-Yang Hospital (2020–2024), contributing to five observations each. Four metabolic diseases were defined using established diagnostic criteria. Several complementary methods were applied: comorbidity network analysis, generalized estimating equations (GEE), sequence analysis, group-based trajectory modeling, multi-state Markov modeling, time-varying covariate Cox regression, and latent class analysis. Multiple-comparison corrections were applied throughout.

Results: Among 8,383 participants (62.27% female; mean age 45.82 ± 13.50 years) with 41,915 person-observations over five years, dyslipidemia was the most prevalent condition (32.53%), followed by hyperuricemia (15.75%), obesity (13.59%), and diabetes (7.23%). Dyslipidemia occupied the central position in the comorbidity network (node strength = 0.59) and served as the most common metabolic gateway (one-year onset probability 12.28%). The dyslipidemia–hyperuricemia pair showed the strongest co-occurrence (odds ratios [OR] = 3.74, 95% confidence intervals [CI] 3.46–4.06), while the diabetes–hyperuricemia association was fully attenuated after adjustment (OR = 1.02; *p*-FDR = 0.782). Four distinct longitudinal trajectories were identified: persistently healthy (54.38%), stable single-disease (29.01%), stable dual-disease (11.15%), and persistently high-burden (5.46%). All four trajectories remained flat over five years, indicating that metabolic comorbidity patterns persist once established. Directional analyses revealed that obesity preceded diabetes in 83.87% of sequential onsets, with the obesity → diabetes pathway representing the strongest unidirectional effect (hazard ratios [HR] = 2.52, 95% CI 2.06–3.09). Diabetes-involving multimorbidity states showed high retention (probability 0.73) and obesity had the lowest remission probability (11.32%). Four comorbidity phenotypes were identified, including a dyslipidemia–obesity–diabetes cluster with low hyperuricemia probability, consistent with sex-stratified findings showing the diabetes–hyperuricemia association reversed direction between males and females.

Conclusions: Metabolic multimorbidity is a structurally ordered process in which dyslipidemia acts as the central hub and the most common gateway condition, obesity serves as the dominant unidirectional driver of diabetes, and established multi-disease states, particularly those involving diabetes, are largely irreversible. Early lipid screening and weight management may offer the most effective intervention opportunities before comorbidity patterns become self-sustaining.

1 Background

Type 2 diabetes, dyslipidemia, hyperuricemia, and obesity are among the most prevalent chronic conditions worldwide, and their population-level burden continues to escalate. This trend is especially pronounced across East Asia, where rapid shifts in diet and physical activity have compressed decades of epidemiological change into a remarkably short period [1–3]. What makes these conditions especially consequential is not their individual prevalence but their tendency to cluster within the same person. Clinical

experience and population surveys consistently demonstrate that metabolic diseases seldom travel alone: a patient found to have one abnormal metabolic parameter is far more likely than chance would predict to carry others [4, 5]. This clustering, commonly termed metabolic multimorbidity, carries cardiovascular and renal consequences that exceed the additive effects of individual diseases and imposes disproportionate costs on healthcare systems organized around single-disease pathways [6, 7].

The concept of metabolic syndrome has long served as a clinical shorthand for this co-occurrence, but its binary classification (present or absent) obscures the internal structure of how diseases relate to one another [8]. It does not reveal which disease sits at the center of the cluster connecting to all others, which pairs remain independently associated after accounting for shared confounders, or whether dyslipidemia precedes hyperuricemia or the reverse. These structural and temporal questions cannot be answered by simply counting risk factor thresholds. Addressing them requires analytical tools that go beyond traditional regression and are capable of capturing network structure, longitudinal trajectories, state transitions, and directional effects.

Network science offers one such tool. Originally developed in social and biological sciences, network analysis treats diseases as nodes and their pairwise associations as edges, enabling the identification of hub diseases, bridge diseases, and peripheral diseases within a comorbidity architecture [9, 10]. Several recent studies have applied network methods to multimorbidity patterns, revealing that metabolic conditions including lipid abnormalities and adiposity tend to occupy central positions within comorbidity architectures [11, 12]. However, most of these analyses have been cross-sectional, leaving critical gaps regarding how the network unfolds over time.

Longitudinal trajectory analysis provides a complementary perspective. Group-based trajectory modeling (GBTM) can identify latent subgroups of individuals following distinct temporal courses of disease accumulation, revealing whether multimorbidity progression is a continuous gradient or a set of discrete pathways [13, 14]. Multi-state Markov models further quantify the probabilities of transitioning between disease states at specific time horizons, illuminating which states are absorbing and which are transient [15]. Finally, time-varying covariate Cox regression enables directional prediction between disease pairs, testing, for instance, whether obesity predicts the future onset of diabetes more strongly than diabetes predicts the future onset of obesity [16].

No single method can capture all dimensions of multimorbidity simultaneously. The present study was therefore designed to integrate several complementary approaches, including network analysis, GEE-based association estimation, GBTM, multi-state Markov modeling, and time-varying covariate Cox regression, and apply them to a retrospective cohort. Each method addresses a distinct question, but together they converge on a coherent picture of how metabolic multimorbidity is organized and assembled. Our objective was to identify specific diseases, transitions, and time windows where preventive intervention is most likely to interrupt the assembly of multimorbidity before it becomes self-sustaining.

2 Methods

2.1 Study design and participants

Data for this retrospective cohort study were sourced from the Health Management Center of Beijing Chao-Yang Hospital between January 2020 and December 2024. The study protocol received ethical approval from the Institutional Review Board of Beijing Chao-Yang Hospital (approval number: 2026-ke-775). The study population comprised 8,383 adults who underwent annual comprehensive health examinations during the study period. Each participant contributed exactly five annual observations (one per calendar year, 2020–2024).

2.2 Definition of Metabolic Diseases

Four metabolic diseases were defined based on clinical measurements and established diagnostic criteria at each health examination visit. (1) Dyslipidemia was defined as meeting any of the following: total cholesterol (TC) ≥ 6.22 mmol/L, triglycerides (TG) ≥ 2.26 mmol/L, low-density lipoprotein cholesterol (LDL-C) ≥ 4.14 mmol/L, or high-density lipoprotein cholesterol (HDL-C) < 1.04 mmol/L, in accordance with the 2016 Chinese guideline for dyslipidemia management [17]. (2) Hyperuricemia was defined as serum uric acid > 420 μmol/L in males or > 360 μmol/L in females, consistent with Chinese expert consensus criteria [18]. (3) Obesity was defined as body mass index (BMI) ≥ 28.0 kg/m², following the Working Group on obesity in China classification [19]. (4) Type 2 diabetes mellitus was defined as fasting plasma glucose ≥ 7.0 mmol/L, or glycated hemoglobin (HbA1c) ≥ 6.5%, consistent with the diagnostic criteria of the American Diabetes Association and the Chinese Diabetes Society [20, 21].

In this study, metabolic multimorbidity was defined as the co-occurrence of two or more of the following four metabolic diseases within the same individual at the same time point: diabetes, dyslipidemia, hyperuricemia, and obesity. A composite multimorbidity count was calculated at each visit as the number of the four conditions simultaneously present (range 0–4). A categorical disease state variable was also derived: healthy (0 diseases), single-disease (1 disease), dual-disease (2 diseases), and multi-disease (3 or 4 diseases).

2.3 Cross-sectional comorbidity network analysis

The comorbidity network was constructed using a pooled annual cross-sectional approach. For each calendar year, each participant's last examination record within that year was selected, yielding one observation per person per year. All person-year observations were then pooled to form the analytical dataset for network construction, maximizing statistical power while preserving the within-year cross-sectional structure.

Nodes represented the four metabolic conditions. Edge weights were quantified using phi coefficients (the Pearson correlation coefficient between two binary variables), computed for all six disease pairs. Phi coefficients were preferred over odds ratios (ORs) for edge weighting because they are symmetric, bounded (− 1 to + 1), and directly interpretable on a standardized scale [22]. Crude OR with 95% confidence intervals (CIs) were computed in parallel using a 0.5 continuity correction to handle sparse cells; statistical significance was assessed by Fisher's exact test. Edges were retained in the network graph if the Fisher test *p*-value was < 0.05.

Centrality metrics were calculated for each node using the igraph package in R [23]: strength centrality (sum of absolute phi coefficients of incident edges), unweighted and weighted betweenness centrality, unweighted

and weighted closeness centrality, and expected influence (sum of signed phi coefficients). For weighted betweenness and closeness, edge distances were defined as the reciprocal of the absolute phi coefficient ($1/|\varphi|$), so that stronger associations corresponded to shorter distances. A node ranking consistently highest across multiple centrality metrics was interpreted as the hub disease of the comorbidity network.

Participant-level cluster bootstrap (1 000 iterations) was used to estimate 95% CIs for phi coefficients, ORs, and node strength, accounting for the correlation among repeated observations from the same individual [24]. In each bootstrap iteration, participants (not observations) were resampled with replacement, and all person-year records belonging to each sampled participant were included, preserving the within-person correlation structure.

Sex-stratified networks were constructed by applying the same procedures separately to male and female participants.

2.4 GEE-based adjusted pairwise associations

Because phi coefficients and crude ORs do not control for confounding, adjusted associations between each disease pair were estimated using GEE logistic regression models implemented with the geeglm function in the geepack package [25]. Each of the six pairwise models specified one disease as the binary outcome and the paired disease as the primary exposure variable. Three model specifications were fitted for each pair: Model 1 (crude), Model 2 (adjusted for age [continuous] and sex [binary]), and Model 3 (fully adjusted for age, sex, and the remaining two metabolic conditions). An exchangeable working correlation structure was specified, and robust (sandwich) standard errors were used for inference, ensuring valid inference even if the working correlation was misspecified [26].

As a sensitivity analysis, all six Model 2 specifications were re-estimated under a first-order autoregressive (AR1) working correlation structure. Concordance of effect direction, magnitude, and statistical significance between exchangeable and AR1 results was assessed. Sex-stratified GEE models were also fitted for each disease pair, adjusting for age only (sex adjustment being inapplicable within strata). All p-values from the six pairwise GEE tests within each model specification were corrected for multiplicity using the Benjamini–Hochberg false discovery rate (FDR) procedure at the 5% threshold [27].

2.5 Sequence analysis and clustering

Longitudinal disease-state sequences were constructed for participants with five visits using the TraMineR package [28]. Each participant's sequence consisted of ordered disease states (healthy, single-disease, dual-disease, multi-disease ≥ 3) at consecutive annual visits. Optimal matching (OM) distances between sequence pairs were computed using transition-rate-derived substitution costs (TRATE method) and an insertion-deletion cost of 1.0. Ward's minimum variance method (ward.D2) was applied to the resulting distance matrix for hierarchical agglomerative clustering. The optimal number of clusters was selected by maximizing the mean silhouette width across two to six candidate solutions. State distribution plots, frequency plots (top 20 most common sequences), modal state plots, and mean time-in-state plots were generated to characterize the identified trajectory types.

2.6 Group-based trajectory modeling (GBTM)

GBTM was implemented using latent class mixed models fitted with the lcmm package in R [29]. The outcome variable was the integer-valued multimorbidity count (0–4 diseases per visit), and time was parameterized as years from baseline (continuous). A linear link function was used uniformly across all models (link = “linear”), treating the observed count as a continuous outcome mapped through a linear transformation to the latent process. This specification was adopted because models with a threshold link function failed to converge adequately, likely due to the limited range of the outcome variable (0–4) and sparse cell counts at certain scale points; by contrast, the linear link function provided stable parameter estimates and acceptable model fit.

Both linear and quadratic polynomial time specifications were evaluated. For each specification, models with one to five latent trajectory groups were fitted. A random intercept (random = ~ 1) was included to account for within-individual heterogeneity. To mitigate sensitivity to starting values, each multi-group model was fitted using 30 sets of random initial values (generated via the random() function from the converged one-group model), with a maximum of 1 500 iterations per run. The solution with the highest log-likelihood among converged runs was retained for each group number. If no model converged with random intercepts, a fixed-effects-only specification (no random term) was attempted as a fallback.

Model selection followed a sequential strategy: (i) the Bayesian information criterion (BIC) was used to rank competing models; (ii) among models with similar BIC, the following diagnostic criteria were applied—relative entropy > 0.70, average posterior probability of group assignment (AvePP) > 0.70 for all groups, and minimum group proportion > 5% [13, 14, 30]. Models failing multiple diagnostic thresholds were excluded unless no alternative converged. Sex-stratified GBTM was performed as a sensitivity analysis to assess whether the optimal number of trajectory groups differed between males and females.

2.7 Multi-state Markov model

A continuous-time Markov model was fitted to characterize transitions among seven mutually exclusive health states using the msm package in R [15, 31]. The seven states were defined as: (1) healthy (no metabolic disease), (2) dyslipidemia only, (3) hyperuricemia only, (4) obesity only, (5) diabetes only, (6) multi-disease without diabetes, and (7) multi-disease with diabetes. This state definition was chosen to distinguish between single-disease entry points (states 2–5) and to separate multi-disease states by the presence or absence of diabetes, given the clinical prognostic importance of this distinction.

The initial transition intensity matrix Q was informed by observed transition frequencies: for state pairs with ≥ 5 observed transitions, the initial intensity was set proportional to the observed transition count divided by the total exits from the origin state; for state pairs with 1–4 observed transitions, a small positive value (0.005) was assigned to permit estimation; and for state pairs with zero observed transitions, the intensity was fixed at zero. Model estimation employed the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm with the following control parameters: function scaling factor (fnscale) = 60 000, maximum iterations = 5 000, and relative tolerance = 1×10^{-8}.

One-year, three-year, and five-year transition probability matrices were derived from the estimated intensity matrix via matrix exponentiation. Mean sojourn times in each state were computed as the negative reciprocal of the diagonal elements of Q. A covariate-adjusted model incorporating sex (binary) and age (centered at

the sample mean) was fitted to estimate hazard ratios (HRs) for the effect of these covariates on transition intensities.

Goodness of fit was evaluated using the Pearson-type chi-squared test implemented in pearson.msm(), which compares observed and expected transition counts under the fitted model with two time groups [31]. To assess the time-homogeneity assumption, a piecewise-constant intensities model was fitted with a single change point at the median follow-up time, and a likelihood ratio test (LRT) was used to compare it against the time-homogeneous baseline model.

2.8 Interval transition probability analysis

To complement the Markov model, disease-specific onset and remission probabilities were estimated using an interval-based approach that did not require the Markov assumption. For each disease, consecutive visit pairs were classified into four transition types: onset (0→1), remission (1→0), persistent disease (1→1), and persistent health (0→0). Person-time at risk was calculated from the interval length between consecutive visits. Incidence rates (events per person-year) were computed for onset (among disease-free intervals) and remission (among diseased intervals). One-year transition probabilities were derived from incidence rates using the exponential formula: P(1 year) = 1 − exp(− rate). Cluster-robust bootstrap CIs (1 000 iterations, resampling at the participant level) were computed to account for within-individual correlation.

2.9 Time-varying covariate Cox regression

To estimate the directional predictive effect of each metabolic condition on the subsequent incidence of each other condition, time-varying covariate Cox proportional hazards models were fitted using the survival package in R [32]. For each of the 12 ordered disease pairs (4 predictor diseases × 3 possible target diseases), the data were structured in counting-process format (tstart, tstop, event) [16, 33]. Participants who already had the target disease at baseline were excluded from that specific model. Follow-up was censored at the first occurrence of the target disease (i.e., each participant contributed at-risk time only until first event or end of observation).

The primary exposure was the time-varying status of the predictor disease, updated at each annual visit. To avoid simultaneity bias, the predictor disease status was lagged by one visit—that is, the model used the predictor status recorded at the visit immediately preceding each risk interval. All models were adjusted for baseline age (continuous), sex (binary), and the time-varying status of the remaining two metabolic conditions (also lagged by one visit). The proportional hazards assumption was evaluated globally and for each covariate using scaled Schoenfeld residuals (cox.zph test) [32].

HRs with 95% CIs were reported for all 12 directional paths. *P*-values were corrected for multiplicity across the 12 tests using the Benjamini–Hochberg FDR procedure. A pair of directional paths (A→B versus B→A) was classified as “asymmetric” if one direction reached FDR-adjusted significance while the reverse did not.

2.10 Latent class analysis (LCA)

LCA was applied to characterize cross-sectional comorbidity phenotypes using the binary indicators of all four metabolic conditions measured at each participant's most recent visit [34]. Models with two to six latent classes were fitted using the poLCA package in R, with 30 sets of random starting values and a maximum of

5 000 iterations per run to ensure global likelihood maximization [35]. The optimal number of classes was selected based on BIC, akaike information criterion (AIC), entropy (target > 0.60), and the requirement that each class contain at least 5% of the sample. Each class was characterized by the conditional probability of each disease (probability of disease = 1 given class membership) and the demographic profile of its members (age, sex, BMI, metabolic biomarkers).

2.11 Sensitivity analyses

Several sensitivity analyses were conducted. First, disease-definition thresholds were shifted by ± 5% from the standard cutoffs to assess the stability of phi coefficients, ORs, and disease prevalence estimates. Second, GEE models were re-fitted under the AR1 working correlation structure, and concordance of adjusted ORs with exchangeable-correlation results was quantified. Third, sex-stratified GBTM models were estimated to verify that the optimal number of trajectory groups was consistent with the pooled analysis. Fourth, a comprehensive summary of all hypothesis tests subject to FDR correction was compiled by analytical module (network GEE, time-varying Cox, logistic risk factor), and the concordance rate between unadjusted and FDR-adjusted significance was reported.

2.12 Statistical software

All analyses were performed in R version 4.6.0 (R Foundation for Statistical Computing, Vienna, Austria). Network analysis used igraph and ggraph; GEE used geepack; sequence analysis used TraMineR; GBTM used lcmm; Markov modeling used msm; Cox regression used survival; LCA used poLCA; and table construction used tableone. Two-sided p-values < 0.05 (or FDR-adjusted q-values < 0.05 where applicable) were considered statistically significant.

3 Results

3.1 Study Population and Prevalence of Metabolic Diseases

The study included 8,383 participants observed annually from 2020 to 2024, yielding a total of 41,915 person-observations. Among all participants, 5,220 (62.27%) were female and 3,163 (37.73%) were male, with an overall mean age of 45.82 ± 13.50 years (Table 1). Dyslipidemia was by far the most prevalent single disease (32.53%), followed by hyperuricemia (15.75%), obesity (13.59%), and diabetes (7.23%). Males had significantly higher prevalence rates for all four conditions compared with females (all **p** < 0.001).

Dyslipidemia was the most prevalent metabolic disease across all years, with prevalence ranging from 32.53% to 31.42% from 2020 to 2024 (Fig. 1A). Hyperuricemia prevalence fluctuated between 13.96% and 17.57%, peaking in 2021 (17.57%) and declining thereafter. Obesity prevalence remained relatively stable at approximately 13.30%–14.28%. Diabetes prevalence rose from 7.23% in 2020 to 9.51% in 2024.

Regarding multimorbidity burden (Fig. 1B), 53.54%–56.28% of participants had no metabolic diseases at any given time point. The proportion with single morbidity was 26.68%–27.83%, while those with two concurrent diseases comprised 12.22%–13.16%. Multi-diseases (≥ 3 diseases) affected 4.70%–5.63% of the study population, with 0.29%–0.54% carrying all four metabolic conditions simultaneously.

Over the five-year follow-up, dyslipidemia exhibited the highest cumulative incidence, far exceeding hyperuricemia, diabetes, and obesity, with its curve rising steeply from early follow-up, confirming dyslipidemia as the earliest and most frequently occurring metabolic disease among the four conditions (Fig. 1C).

Table 1
Baseline characteristics of study participants stratified by sex.

Variable	Overall (n = 8383)	Female (n = 5220)	Male (n = 3163)	*p* value
Age, years	45.82 ± 13.50	45.14 ± 13.64	46.94 ± 13.20	< 0.001
BMI, kg/m²	24.23 ± 3.62	23.21 ± 3.42	25.92 ± 3.29	< 0.001
Fasting glucose, mmol/L	4.91 ± 1.15	4.79 ± 0.98	5.12 ± 1.37	< 0.001
HbA1c, %	5.50 [5.32, 5.80]	5.50 [5.30, 5.76]	5.60 [5.40, 5.90]	< 0.001
TC mmol/L	4.94 ± 0.90	4.95 ± 0.90	4.91 ± 0.89	0.039
TG, mmol/L	1.16 [0.82, 1.69]	1.01 [0.75, 1.45]	1.42 [1.00, 2.04]	< 0.001
LDL-C, mmol/L	3.10 ± 0.87	3.02 ± 0.84	3.22 ± 0.89	< 0.001
HDL-C, mmol/L	1.38 ± 0.36	1.50 ± 0.36	1.19 ± 0.28	< 0.001
Uric acid, μmol/L	321.00 [267.00, 385.00]	285.00 [248.00, 330.00]	389.00 [339.00, 445.00]	< 0.001
Diabetes, n (%)	606 (7.23)	263 (5.04)	343 (10.84)	< 0.001
Dyslipidemia, n (%)	2727 (32.53)	1173 (22.47)	1554 (49.13)	< 0.001
Hyperuricemia, n (%)	1320 (15.75)	219 (4.20)	1101 (34.81)	< 0.001
Obesity, n (%)	1139 (13.59)	444 (8.51)	695 (21.97)	< 0.001

Notes: Values are presented as mean ± SD for normally distributed continuous variables, median [interquartile range] for non-normally distributed variables (triglycerides, HbA1c, uric acid), or n (%) for categorical variables. Non-normally distributed variables were compared using the Mann–Whitney U test; normally distributed continuous variables were compared using the independent samples t-test; and categorical variables were compared using the chi-squared test.

3.2 Cross-Sectional Network Structure

In the comorbidity network, the dyslipidemia-hyperuricemia pair exhibited the strongest association (phi = 0.241; crude OR = 3.74; 95% CI 3.46–4.06), followed by obesity-hyperuricemia (phi = 0.207; crude OR = 3.64) and obesity-dyslipidemia (phi = 0.199; crude OR = 3.11). The weakest pairwise association was between diabetes and hyperuricemia (phi = 0.041; crude OR = 1.45) (Fig. 2A). Dyslipidemia ranked highest on most centrality metrics, with a node strength of 0.59, compared with 0.53 for obesity, 0.49 for hyperuricemia, and 0.31 for diabetes, identifying it as the most densely connected node in the network (Fig. 2B). Year-specific global network metrics, including density, transitivity, mean phi coefficient, and mean OR, remained stable

across 2020–2024 (Supplementary Table S1). Bootstrap 95% CIs confirmed that these strength centrality estimates were robust and non-overlapping between dyslipidemia and diabetes (Supplementary Table S2).

GEE models progressively adjusted for confounders modified these associations in important ways (Table 2; Fig. 2C). In Model 2 (adjusted for age and sex), five of the six disease pairs retained statistically significant positive associations after FDR correction. The dyslipidemia–hyperuricemia pair remained the strongest (adjusted OR = 1.72; 95% CI 1.61–1.85; *p*-FDR < 0.001), followed by diabetes–dyslipidemia (OR = 1.66; 95% CI 1.51–1.84), diabetes–obesity (OR = 1.41; 95% CI 1.27–1.55), dyslipidemia–obesity (OR = 1.28; 95% CI 1.22–1.35), and hyperuricemia–obesity (OR = 1.15; 95% CI 1.08–1.22; all *p*-FDR < 0.001). The diabetes-hyperuricemia pair was an exception: its adjusted OR fell to 1.02 (95% CI 0.89–1.17; *p*-FDR = 0.782), indicating that the crude co-occurrence of these two conditions was largely explained by shared confounding through age and sex.

In Model 3 (fully adjusted for age, sex, and the remaining two metabolic conditions), associations were modestly attenuated but remained statistically significant for the same five pairs. The dyslipidemia–hyperuricemia pair retained the strongest independent association (OR = 1.67; 95% CI 1.56–1.79; *p*-FDR < 0.001), followed by diabetes–dyslipidemia (OR = 1.61; 95% CI 1.45–1.77), diabetes–obesity (OR = 1.38; 95% CI 1.25–1.53), dyslipidemia–obesity (OR = 1.27; 95% CI 1.20–1.33), and hyperuricemia–obesity (OR = 1.13; 95% CI 1.07–1.20; all *p*-FDR < 0.001). In the fully adjusted model, the diabetes-hyperuricemia pair shifted toward an inverse direction (OR = 0.89; 95% CI 0.78–1.03; *p*-FDR = 0.127), suggesting that once shared associations with dyslipidemia and obesity are accounted for, diabetes and hyperuricemia may be negatively coupled. This finding is explored further in the directional Cox analyses (Section 3.5). GEE results were essentially unchanged under the AR1 working correlation structure (Supplementary Table S5), with all six disease pairs yielding directionally identical adjusted ORs and identical significance classifications under both Model 2 and Model 3.

Sex-stratified analyses revealed notable differences in the diabetes-hyperuricemia association. At the descriptive level, males exhibited a negative phi coefficient (– 0.051) with a crude OR of 0.71 (95% CI 0.58–0.88), whereas females exhibited a positive phi coefficient (0.098) with a crude OR of 2.58 (95% CI 1.75–3.81) (Fig. 2A). After accounting for within-person correlation and adjusting for age (GEE Model 2), the sex divergence persisted: the age-adjusted OR among males was 0.85 (95% CI 0.73–0.99), and among females was 1.86 (95% CI 1.42–2.43) (Fig. 2C).

Table 2
Pairwise associations between four metabolic diseases estimated from pooled annual cross-sectional data.

Disease Pair	Phi	Crude OR (Bootstrap 95% CI) [a]	Model 1 OR (95% CI) [b]	Model 2 OR (95% CI) [c]	Model 3 OR (95% CI) [d]	*p*-FDR (Model 2) [e]	*p*-FDR (Model 3) [e]
Dyslipidemia vs Hyperuricemia	0.241	3.74 (3.46–4.06)	1.80 (1.69–1.92)	1.72 (1.61–1.85)	1.67 (1.56–1.79)	< 0.001	< 0.001
Hyperuricemia vs Obesity	0.207	3.64 (3.29–4.06)	1.21 (1.13–1.30)	1.15 (1.08–1.22)	1.13 (1.07–1.20)	< 0.001	< 0.001
Dyslipidemia vs Obesity	0.199	3.11 (2.82–3.44)	1.32 (1.25–1.38)	1.28 (1.22–1.35)	1.27 (1.20–1.33)	< 0.001	< 0.001
Diabetes vs Dyslipidemia	0.154	2.99 (2.67–3.36)	1.86 (1.69–2.04)	1.66 (1.51–1.84)	1.61 (1.45–1.77)	< 0.001	< 0.001
Diabetes vs Obesity	0.119	2.64 (2.28–3.05)	1.46 (1.32–1.62)	1.41 (1.27–1.55)	1.38 (1.25–1.53)	< 0.001	< 0.001
Diabetes vs Hyperuricemia	0.041	1.45 (1.27–1.67)	1.12 (0.98–1.28)	1.02 (0.89–1.17)	0.89 (0.78–1.03)	0.782	0.127

Notes: [a] Crude ORs with Haldane correction (+ 0.5 added to each cell of the 2×2 table); 95% CIs were obtained from 1,000-iteration participant-level cluster bootstrap resampling, in which participants (not observations) were resampled with replacement to preserve within-person correlation. [b] Model 1: unadjusted (crude) population-averaged OR estimated from GEE with logit link and exchangeable working correlation structure; robust (sandwich) standard errors were used for inference. [c] Model 2: GEE adjusted for age (continuous) and sex (binary). [d] Model 3: GEE adjusted for age, sex, and the remaining two metabolic conditions. This model estimates the independent association between each disease pair after accounting for shared co-occurrence with the other metabolic conditions. [e] Benjamini–Hochberg FDR correction applied separately within each model specification (6 tests per model). Bold values indicate statistical significance after FDR correction (*p*-FDR < 0.05). The difference in magnitude between bootstrap-derived crude ORs and GEE Model 1 ORs reflects the treatment of repeated measures: crude ORs were computed from pooled 2×2 contingency tables in which all person-year observations were treated as independent units, thereby inflating concordant cells due to within-person disease persistence across years; GEE Model 1 ORs, while also unadjusted for covariates, account for within-person correlation through the exchangeable working correlation structure, yielding appropriately attenuated population-averaged marginal estimates.

3.3 Longitudinal Trajectories

GBTM identified four trajectory groups as optimal based on BIC, average posterior probabilities, and adequate group sizes (Fig. 3A–B; Table 3). Full model-selection diagnostics—including BIC, AIC, entropy, minimum average posterior probability (0.86), and minimum group proportion (5.46%)—are detailed in Supplementary Table S3. From lowest to highest comorbidity burden, these were as follows. Trajectory 1 (persistently healthy, 54.38%) comprised individuals who maintained a near-zero multimorbidity count across all five visits; they were younger (mean age 43.39 years), predominantly female (80.32%), and had the most favorable baseline metabolic profiles. Trajectory 2 (stable single-disease, 29.01%) consisted of individuals

whose disease count hovered around one throughout follow-up, with dyslipidemia (53.45%) being the most common condition. Trajectory 3 (stable dual-disease, 11.15%) included individuals carrying approximately two concurrent conditions at every visit, most commonly dyslipidemia combined with hyperuricemia or obesity. Trajectory 4 (persistently high-burden, 5.46%) was the smallest group, maintaining a count near three diseases across all years and composed disproportionately of males (83.19%).

All four trajectories were essentially flat, with no group showing meaningful upward or downward trends over the five-year period. This stability indicates that metabolic comorbidity patterns, once established, tend to persist in the absence of deliberate intervention.

Table 3
Baseline characteristics by GBTM trajectory group

Variable	Trajectory 1 (n = 4,559, 54.38%)	Trajectory 2 (n = 2,432, 29.01%)	Trajectory 3 (n = 934, 11.15%)	Trajectory 4 (n = 458, 5.46%)	*p* value
Demographics					
Age (years), mean ± SD	43.39 ± 12.91	48.67 ± 13.62	49.57 ± 13.59	47.23 ± 13.59	< 0.001
Male, n (%)	897 (19.68%)	1215 (49.96%)	670 (71.73%)	381 (83.19%)	< 0.001
Anthropometrics					
BMI (kg/m²), mean ± SD	22.49 ± 2.53	25.14 ± 3.08	27.40 ± 3.31	30.33 ± 3.32	< 0.001
Glycemic markers					
Fasting glucose (mmol/L)	4.60 ± 0.53	5.01 ± 1.11	5.61 ± 1.84	6.03 ± 2.19	< 0.001
HbA1c (%)	5.48 ± 0.31	5.75 ± 0.70	6.09 ± 1.04	6.38 ± 1.25	< 0.001
Lipid profile					
TC (mmol/L)	4.76 ± 0.74	5.16 ± 1.00	5.12 ± 1.01	5.11 ± 1.07	< 0.001
TG (mmol/L)	1.02 ± 0.56	1.64 ± 1.04	2.25 ± 1.90	2.55 ± 1.69	< 0.001
LDL-C (mmol/L)	2.85 ± 0.70	3.39 ± 0.92	3.37 ± 0.96	3.40 ± 1.07	< 0.001
HDL-C (mmol/L)	1.53 ± 0.33	1.27 ± 0.33	1.11 ± 0.24	1.04 ± 0.21	< 0.001
Other biomarkers					
Uric acid (μmol/L)	291.23 ± 62.33	353.83 ± 80.66	406.97 ± 86.81	453.56 ± 84.70	< 0.001
Baseline diseases, n (%)					
Diabetes	18 (0.39%)	208 (8.55%)	215 (23.02%)	165 (36.03%)	< 0.001
Dyslipidemia	346 (7.59%)	1300 (53.45%)	685 (73.34%)	396 (86.46%)	< 0.001
Hyperuricemia	106 (2.33%)	483 (19.86%)	409 (43.79%)	322 (70.31%)	< 0.001
Obesity	26 (0.57%)	365 (15.01%)	378 (40.47%)	370 (80.79%)	< 0.001
Comorbidity burden					

Notes: The four trajectory groups were identified using GBTM with a linear specification and link = “linear”. Continuous variables are presented as mean ± SD and compared using the Kruskal–Wallis test; categorical variables are presented as n (%) and compared using the chi-squared test.

Variable	Trajectory 1 (n = 4,559, 54.38%)	Trajectory 2 (n = 2,432, 29.01%)	Trajectory 3 (n = 934, 11.15%)	Trajectory 4 (n = 458, 5.46%)	*p* value
Demographics					
No. of diseases, mean ± SD	0.11 ± 0.32	0.97 ± 0.60	1.81 ± 0.64	2.74 ± 0.65	< 0.001

Notes: The four trajectory groups were identified using GBTM with a linear specification and link = "linear". Continuous variables are presented as mean ± SD and compared using the Kruskal–Wallis test; categorical variables are presented as n (%) and compared using the chi-squared test.

3.4 State Transitions and Disease Temporal Sequencing

The multi-state Markov model estimated one-year transition probabilities among seven metabolic states (Fig. 4A). Among disease-free participants, the probability of remaining healthy one year later was 0.84, with a total probability of 0.16 for transitioning to any disease state. The most frequent initial transition was to dyslipidemia only (0.09), followed by hyperuricemia only (0.03), with obesity only and diabetes only each at 0.01. Disease-specific interval analyses confirmed this hierarchy (Table 4): the one-year onset probability was highest for dyslipidemia (12.28%, 95% CI: 11.82–12.72%), followed by hyperuricemia (6.02%, 5.72–6.35%), diabetes (2.21%, 2.04–2.37%), and obesity (2.09%, 1.92–2.27%), positioning dyslipidemia as the most common metabolic gateway. Extended transition matrices showed that the probability of remaining healthy declined from 0.84 at one year to 0.70 at three years and 0.63 at five years, while the cumulative probability of entering a multimorbidity state rose correspondingly (Supplementary Table S4)

Once acquired, diseases differed markedly in persistence (Table 4). Obesity exhibited the lowest one-year remission probability (11.32%, 95% CI: 10.36–12.22%), followed by diabetes (16.79%, 15.36–18.17%), dyslipidemia (25.68%, 24.89–26.42%), and hyperuricemia (30.39%, 29.37–31.54%), yielding corresponding one-year retention probabilities of 88.7%, 83.2%, 74.3%, and 69.6%, respectively.

In the multi-state model, within-state retention for single-disease states was highest for diabetes only (0.60) and obesity only (0.59), compared with dyslipidemia only (0.53) and hyperuricemia only (0.42). These lower within-state retention probabilities partly reflected progression toward multimorbidity: within one year, 10% of those with obesity only, 10% of those with diabetes only, 7% of those with hyperuricemia only, and 15% of those with dyslipidemia only transitioned to a multimorbidity state. Multimorbidity states were themselves highly stable. The probability of remaining in the multimorbidity-with-diabetes state was 0.73. Similarly, the retention probability for the multimorbidity-without-diabetes state was 0.66. These patterns reinforced the trajectory stability observed in the GBTM analysis and underscored the difficulty of reversing established multimorbidity.

Temporal sequencing analysis further characterized the order of disease emergence (Fig. 4B–C). Across all disease pairs, simultaneous diagnosis was the most common pattern, accounting for 47.72%–58.51% of co-occurrences. Among the 832 participants who developed both dyslipidemia and diabetes, dyslipidemia was diagnosed first in 35.22%, diabetes first in 17.06%, and both simultaneously in 47.72%; excluding simultaneous cases, dyslipidemia preceded diabetes in 67.36% of sequential onsets (293/435; binomial test

$p < 0.001$). A similar temporal precedence of dyslipidemia was observed for the dyslipidemia–hyperuricemia pair (n = 1,792; dyslipidemia first in 31.03% vs. hyperuricemia first in 18.64%; among sequential onsets, 62.47% dyslipidemia-first, $p < 0.001$). In contrast, for the dyslipidemia–obesity pair (n = 1,157), obesity was diagnosed before dyslipidemia more often than the reverse (24.38% vs. 17.11%; among sequential onsets, 58.75% obesity-first, $p < 0.001$). Notably, obesity also preceded diabetes (41.16% vs. 7.92%, n = 379) and hyperuricemia (30.88% vs. 16.91%, n = 816) in the majority of sequential cases. These sequencing data, combined with dyslipidemia's hub position in the cross-sectional network and its highest incidence rate, suggest that dyslipidemia is the most common gateway metabolic disease on the path toward multimorbidity for lipid–glycemic and lipid–urate trajectories, whereas obesity tends to emerge as an independent early precursor across multiple comorbidity pathways.

Table 4
Interval transition probabilities for each metabolic disease

Disease	New Onset (Healthy → Disease)				Remission (Disease → Healthy)			
	1-Year Probability, % (95% CI)[a]	Events	Person-Years	Rate (/PY)	1-Year Probability, % (95% CI)[a]	Events	Person-Years	Rate (/PY)
Dyslipidemia	12.28 (11.82–12.72)	2 884	22 011.3	0.131	25.68 (24.89–26.42)	2 977	10 029.5	0.297
Hyperuricemia	6.02 (5.72–6.35)	1 677	26 998.3	0.062	30.39 (29.37–31.54)	1 827	5 042.5	0.362
Diabetes	2.21 (2.04–2.37)	658	29 499.8	0.022	16.79 (15.36–18.17)	467	2 541.0	0.184
Obesity	2.09 (1.92–2.27)	585	27 652.2	0.021	11.32 (10.36–12.22)	527	4 388.6	0.120

Notes: [a] One-year transition probabilities were estimated as 1 − exp(− rate), where rate = events / person-years; 95% CIs were derived from 1,000-iteration participant-level cluster bootstrap resampling. Median observation interval = 0.98 years for all diseases. New onset was defined as transition from disease-free to disease-present status between consecutive visits; remission was defined as disease-present to disease-free. Cohort restricted to participants with 5 visits (n = 8,383); each disease-specific analysis included only individuals free of (for onset) or diagnosed with (for remission) the target disease at the preceding visit.

3.5 Directional Predictions Between Metabolic Diseases

Time-varying covariate Cox models estimated 12 directional HRs (Fig. 5A; Table 5). After FDR correction, 10 of the 12 paths reached statistical significance.

Obesity was the single strongest predictor of incident diabetes (HR = 2.52; 95% CI 2.06–3.09; p-FDR < 0.001), representing the most asymmetric directional path observed. The reverse path, diabetes predicting incident

obesity, was non-significant (HR = 1.05; 95% CI 0.72–1.54; *p*-FDR = 0.804). Hyperuricemia and obesity showed a bidirectional association. Hyperuricemia predicted incident obesity (HR = 1.93; 95% CI 1.51–2.47; *p*-FDR < 0.001), and obesity predicted incident hyperuricemia (HR = 1.77; 95% CI 1.52–2.05; *p*-FDR < 0.001).

Dyslipidemia significantly predicted the incidence of all three other diseases: diabetes (HR = 1.63; 95% CI 1.35–1.96), hyperuricemia (HR = 1.60; 95% CI 1.41–1.82), and obesity (HR = 1.86; 95% CI 1.51–2.30; all *p*-FDR < 0.001). However, the reverse paths were also significant and of comparable magnitude: diabetes → dyslipidemia (HR = 1.55; 95% CI 1.30–1.85), obesity → dyslipidemia (HR = 1.71; 95% CI 1.50–1.96), and hyperuricemia → dyslipidemia (HR = 1.86; 95% CI 1.63–2.14). The last of these exceeded the corresponding forward effect.

The path from diabetes to incident hyperuricemia yielded a point estimate below unity (HR = 0.81; 95% CI 0.66–0.999; raw p = 0.049), though it did not survive FDR correction (*p*-FDR = 0.053).

Table 5
Time-varying covariate cox proportional hazards model: longitudinal predictive effects between metabolic diseases

Outcome (New-Onset)	Predictor (Time-Varying)	Direction	HR (95% CI)	*p* value	*p-FDR*[a]
Diabetes	Obesity	Obesity → Diabetes	2.52 (2.06–3.09)	< 0.001	< 0.001
	Dyslipidemia	Dyslipidemia → Diabetes	1.63 (1.35–1.96)	< 0.001	< 0.001
	Hyperuricemia	Hyperuricemia → Diabetes	1.38 (1.10–1.74)	0.006	0.007
Dyslipidemia	Hyperuricemia	Hyperuricemia → Dyslipidemia	1.86 (1.63–2.14)	< 0.001	< 0.001
	Obesity	Obesity → Dyslipidemia	1.71 (1.50–1.96)	< 0.001	< 0.001
	Diabetes	Diabetes → Dyslipidemia	1.55 (1.30–1.85)	< 0.001	< 0.001
Hyperuricemia	Obesity	Obesity → Hyperuricemia	1.77 (1.52–2.05)	< 0.001	< 0.001
	Dyslipidemia	Dyslipidemia → Hyperuricemia	1.60 (1.41–1.82)	< 0.001	< 0.001
	Diabetes	Diabetes → Hyperuricemia	0.81 (0.66–0.999)	0.049	0.053[b]
Obesity	Hyperuricemia	Hyperuricemia → Obesity	1.93 (1.51–2.47)	< 0.001	< 0.001
	Dyslipidemia	Dyslipidemia → Obesity	1.86 (1.51–2.30)	< 0.001	< 0.001

Notes: All models were fitted in the counting-process (start, stop) format, with each metabolic disease serving alternately as the outcome (new onset) and as a time-varying predictor. Models were adjusted for age, sex, and all other metabolic diseases as time-varying covariates. Analysis was restricted to participants free of the outcome disease at baseline. Bold HR values indicate statistical significance after FDR correction (*p*-FDR < 0.05). a Benjamini–Hochberg FDR correction applied within the Time-Varying Cox module (12 tests). b Nominally significant (raw p = 0.049) but no longer significant after FDR correction (*p*-FDR = 0.053).

Outcome (New-Onset)	Predictor (Time-Varying)	Direction	HR (95% CI)	*p* value	*p-FDR*[a]
	Diabetes	Diabetes → Obesity	1.05 (0.72–1.54)	0.804	0.804

Notes: All models were fitted in the counting-process (start, stop) format, with each metabolic disease serving alternately as the outcome (new onset) and as a time-varying predictor. Models were adjusted for age, sex, and all other metabolic diseases as time-varying covariates. Analysis was restricted to participants free of the outcome disease at baseline. Bold HR values indicate statistical significance after FDR correction (*p*-FDR < 0.05). a Benjamini–Hochberg FDR correction applied within the Time-Varying Cox module (12 tests). b Nominally significant (raw p = 0.049) but no longer significant after FDR correction (*p*-FDR = 0.053).

3.6 Latent Class Analysis: Comorbidity Phenotypes

LCA identified four latent classes as optimal (Fig. 5B-C). Class 1 (dyslipidemia-dominant) was characterized by a very high probability of dyslipidemia (0.88) with moderate hyperuricemia (0.28) and low probabilities of diabetes (0.18) and obesity (0.02). Class 2 (metabolically healthy) showed low probabilities for all four diseases (all ≤ 0.12). Class 3 (dyslipidemia–obesity–hyperuricemia triad) exhibited co-elevation of dyslipidemia (0.71), obesity (0.65), and hyperuricemia (0.58), with a notably low probability of diabetes (0.01). Class 4 (dyslipidemia-obesity-diabetes cluster) was marked by elevated probabilities of dyslipidemia (0.65), obesity (0.64), and diabetes (0.43), but a relatively low probability of hyperuricemia (0.17), a pattern consistent with the inverse diabetes-hyperuricemia association observed in the Cox models.

3.7 Sensitivity Analyses

The main findings proved robust across all sensitivity checks (Supplementary Tables S5–S7). Shifting disease-definition thresholds by ± 5% altered phi coefficients by less than 0.05 for five of the six disease pairs; the exception was diabetes–hyperuricemia (phi change of 0.062 under the relaxed diabetes threshold), though the qualitative conclusion—diabetes–hyperuricemia as the weakest network edge—was unchanged. GEE results under the AR1 correlation structure agreed with exchangeable-structure results in both direction and significance for all six pairs (concordance 6/6). GBTM in sex-stratified models retained four groups as optimal for both males and females, with trajectory shapes mirroring the pooled analysis. Across 26 hypothesis tests subjected to FDR correction, only one changed its significance classification (96.2% concordance), and this was the borderline diabetes → hyperuricemia Cox path already flagged in the main analysis.

4 Discussion

Five analytical perspectives (network structure, adjusted associations, longitudinal trajectories, state transitions, and directional prediction) converge on a coherent picture of how metabolic multimorbidity is organized and assembled. Three principal insights emerge from this convergence, each with direct implications for clinical practice and cardiometabolic disease prevention.

4.1 Dyslipidemia as the Structural Keystone

Across every analysis in which dyslipidemia featured, it occupied a privileged position. It ranked highest on most centrality metrics in the cross-sectional network, maintained the strongest adjusted association with two of the three other conditions, exhibited the highest one-year incidence among disease-free individuals (12.28%), was most frequently the first condition diagnosed in temporal sequencing analyses involving lipid–glycemic and lipid–urate disease pairs, and significantly predicted the future onset of all three remaining conditions in time-varying Cox models. No other condition matched this breadth of structural influence.

In the metabolic comorbidity network, dyslipidemia plays an analogous structural role: it connects to every other node, it appears earliest and most frequently, and its presence forecasts the arrival of further conditions. This finding is mechanistically plausible. Atherogenic dyslipidemia, particularly hypertriglyceridemia and reduced HDL-C, is tightly linked to both hepatic and peripheral insulin resistance and to chronic low-grade inflammation, two pathways through which lipid derangement may promote downstream metabolic dysfunction [39, 40]. Recent evidence from Mendelian randomization studies further supports a causal pathway from triglyceride-rich lipoproteins to insulin resistance [41].

An important nuance, however, emerged from the directional Cox models. Although dyslipidemia significantly predicted the onset of hyperuricemia, obesity, and diabetes, the reverse paths were also statistically significant and of comparable magnitude. Most notably, hyperuricemia predicted incident dyslipidemia (HR = 1.86) with an effect size exceeding that of the forward path (HR = 1.60). This bidirectionality indicates that the hub role of dyslipidemia in the comorbidity network reflects mutual metabolic reinforcement, likely mediated by shared upstream determinants such as insulin resistance and visceral adiposity, rather than a strictly unidirectional causal cascade. Clinically, this means that while lipid abnormalities are the most common point of entry, intervening on any single node in the network is unlikely to fully prevent the activation of others once shared metabolic drivers are in place.

From a clinical standpoint, this insight argues for repositioning lipid screening as more than a cardiovascular risk assessment tool. A newly detected abnormal lipid profile in an otherwise healthy adult could serve as the sentinel event that triggers comprehensive metabolic evaluation, including assessment of uric acid, fasting glucose, and body composition, along with early lifestyle counseling. Several recent trials have demonstrated that intensive behavioral intervention initiated at the dyslipidemia stage can delay the progression to impaired glucose tolerance and hyperuricemia [42, 43], yet these interventions are not yet systematically embedded in clinical pathways.

4.2 Obesity: Unidirectional Driver of Diabetes and Parallel Metabolic Gateway

The time-varying Cox analysis revealed the single strongest and most asymmetric directional effect in the entire network: obesity more than doubled the hazard of incident diabetes (HR = 2.52; 95% CI 2.06–3.09), while diabetes carried no reciprocal predictive effect on obesity onset (HR = 1.05; 95% CI 0.72–1.54). This asymmetry represents longitudinal, time-varying evidence consistent with excess adiposity acting as an upstream contributor to type 2 diabetes in this population.

The magnitude and direction of this effect are consistent with pooled estimates from large meta-analyses, where each 5 kg/m² increment in BMI roughly doubles the risk of type 2 diabetes [44], and with intervention

evidence from the Diabetes Prevention Program and the Da Qing study, in which modest weight reduction achieved 40–58% reductions in diabetes incidence [45, 46]. What our analysis adds is the explicitly directional contrast: by simultaneously estimating both the obesity-to-diabetes and diabetes-to-obesity paths within the same cohort using identical time-varying methodology, we confirmed that the former is strong and the latter is null. Prior studies have typically examined only one direction, making it impossible to distinguish a truly unidirectional effect from incomplete evidence.

Beyond its role as a driver of diabetes, temporal sequencing analysis revealed that obesity also functions as a parallel metabolic gateway—distinct from, and complementary to, dyslipidemia. When dyslipidemia and obesity co-occurred in the same individual, obesity was diagnosed first in 58.75% of sequential onsets, reversing the temporal priority that dyslipidemia held over all other disease pairs. Similarly, obesity preceded diabetes in 83.87% and hyperuricemia in 64.61% of sequential onsets involving those pairs. Taken together with the Cox models, which showed that obesity predicted incident hyperuricemia (HR = 1.93) and dyslipidemia (HR = 1.71) in addition to diabetes, these findings suggest that obesity may function as a broad upstream contributor that tends to appear early when it participates in multimorbidity assembly. The LCA results echoed this pattern: both the dyslipidemia–hyperuricemia–obesity triad (Class 3) and the dyslipidemia–obesity–diabetes cluster (Class 4) featured high conditional probabilities of obesity (0.65 and 0.64, respectively), underscoring its centrality in multi-disease phenotypes.

A bidirectional feedback loop between hyperuricemia and obesity further complicates the picture. Hyperuricemia predicted incident obesity (HR = 1.93) and obesity predicted incident hyperuricemia (HR = 1.77), suggesting a potentially self-reinforcing cycle that may contribute to the persistence of both conditions once established. This bidirectional relationship is biologically plausible: hyperuricemia promotes adipogenesis and inhibits fatty acid oxidation, while excess adipose tissue increases xanthine oxidase activity and reduces renal uric acid clearance [36, 37].

For cardiometabolic prevention, the implication is twofold. First, these findings support the hypothesis that weight management may be more effectively positioned as a primary prevention strategy for diabetes in obese individuals who have not yet developed glucose dysregulation, rather than solely as an adjunctive recommendation after diabetes diagnosis. Second, the identification of obesity as a parallel gateway suggests that BMI screening deserves the same sentinel status as lipid screening—each can serve as an early warning signal that should trigger comprehensive metabolic evaluation. The flat GBTM trajectories further underscore the urgency of acting early—once diabetes is established as part of a multimorbid constellation, the probability of spontaneous reversion is low, and the system locks into a self-sustaining state.

4.3 Trajectory Stability and Its Consequences

A striking finding was the flatness of all trajectory groups over five years. The 54% who were healthy at baseline remained healthy; the approximately 5% who carried three conditions at baseline still carried three conditions five years later. The Markov model reinforced this picture by showing high self-maintenance probabilities for disease states, particularly multi-disease states involving diabetes (retention probability 0.73), and low probabilities of reversion to a disease-free state (0.01 from multimorbidity with diabetes).

This stability should not be mistaken for benignity. It means that the window for effective clinical intervention narrows as comorbidity accumulates. The data suggest two critical intervention junctures: the transition from zero to one disease, where annual conversion rates are modest (approximately 16% per year) and the first disease is most often dyslipidemia; and the transition from one to two or more diseases, where the addition of a second condition dramatically increases the probability of remaining in a multi-disease state. Once a patient crosses into multi-disease territory, especially if diabetes is among the conditions present, reversal becomes improbable within the time frame of routine clinical follow-up.

This persistence of metabolic multimorbidity echoes findings from other chronic disease contexts. Whitty and colleagues [47] observed similar trajectory stability in a large UK cohort examining general multimorbidity, and recent studies in Chinese populations have demonstrated that cardiometabolic multimorbidity trajectories in midlife strongly predict cardiovascular events and mortality [48, 49]. Our contribution extends this literature by demonstrating trajectory stability specifically within a four-disease metabolic domain, by linking it to network architecture and state-transition dynamics, and by doing so in an occupational cohort with regular access to health screening. Despite annual checkups, this population evidently lacked sufficient systematic intervention to alter the course of established multimorbidity.

4.4 Sex Differences in Comorbidity Architecture

The reversal of the diabetes–hyperuricemia association by sex warrants particular attention. In the pooled analysis, the GEE-adjusted OR for this pair was 1.02 (95% CI 0.89–1.17; p-FDR = 0.782), indicating apparent independence after controlling for age and sex. Sex-stratified analysis, however, revealed that this null finding masked a qualitative reversal: in males, diabetes was negatively associated with hyperuricemia (crude OR = 0.71), whereas in females, the association was strongly positive (crude OR = 2.58). The pooled null result therefore represents not the absence of association but the algebraic cancellation of two opposing sex-specific effects.

The negative association in males is biologically plausible and consistent with the borderline negative diabetes → hyperuricemia path observed in the Cox models (HR = 0.81; p-FDR = 0.053). Hyperglycemia-induced glycosuria competes with uric acid for reabsorption in the proximal renal tubule, promoting uricosuria and lowering serum uric acid levels; this effect has been documented in cross-sectional studies linking higher HbA1c to lower uric acid [38]. In an occupational cohort with a high baseline prevalence of hyperuricemia among males, this glycosuric mechanism may be sufficiently strong to produce a net negative cross-sectional association. In females, where both diabetes and hyperuricemia are less prevalent, co-occurrence more likely reflects shared underlying metabolic vulnerability—insulin resistance, visceral adiposity, or genetic predisposition—that is unmasked only after the protective uricosuric effect of estrogen wanes with menopause [50].

This finding has practical implications for risk stratification. Applying a single pooled-sample comorbidity network to both sexes would misrepresent the expected disease co-occurrence patterns in each group. Sex-specific screening protocols and risk calculators may be warranted, particularly for uric acid–glucose interactions.

4.5 Toward a Structural Model of Metabolic Multimorbidity

Synthesizing all analytical perspectives—network structure, GEE-adjusted associations, temporal sequencing, trajectory modeling, Markov transitions, directional Cox prediction, and latent class phenotyping—we propose the following structural model of metabolic multimorbidity assembly.

Two diseases serve as parallel gateways into metabolic multimorbidity, each through a distinct mechanism. Dyslipidemia is the most incident metabolic condition and the network's central hub; it arises earliest and most frequently in lipid–glycemic and lipid–urate trajectories and connects bidirectionally to all other nodes. Obesity, though less prevalent overall, tends to precede other metabolic diseases when it participates in co-morbid pairs and exerts the strongest unidirectional predictive effect in the entire network—more than doubling the hazard of incident diabetes. These two gateways are not mutually exclusive: a substantial proportion of individuals develop both, and LCA Class 3 (dyslipidemia–obesity–hyperuricemia triad) and Class 4 (dyslipidemia–obesity–diabetes cluster) illustrate how the two entry points converge into distinct multi-disease phenotypes.

Once the gateway condition is established, the network appears to expand through bidirectional reinforcement, as suggested by the reciprocal predictive associations observed in the Cox models. Dyslipidemia and hyperuricemia amplify each other (Cox HRs of 1.60 and 1.86, respectively), as do hyperuricemia and obesity (HRs of 1.77 and 1.93). These bidirectional loops, hypothetically mediated by shared mechanisms of insulin resistance and inflammatory signaling, may contribute to stabilizing the multi-disease state and reducing the likelihood of spontaneous reversal [39, 51]—as confirmed by the flat GBTM trajectories and the high Markov retention probabilities.

Diabetes typically appears last in the temporal sequence. Its onset is strongly predicted by pre-existing obesity but does not, in turn, predict additional disease accumulation. Instead, it appears to function as a locking event: once diabetes joins the comorbidity constellation, the system enters a self-sustaining, high-burden state from which the Markov model estimates a reversion probability of only 1% per year. The LCA Class 4 phenotype, characterized by co-elevation of dyslipidemia (0.65), obesity (0.64), and diabetes (0.43) with low hyperuricemia (0.17), is the cross-sectional manifestation of this locked state. The notably low hyperuricemia probability in this class is consistent with the glycosuric uricosuria mechanism discussed above, providing convergent evidence for the negative diabetes → hyperuricemia pathway.

This model identifies three actionable leverage points for clinical prevention. First, lipid management at the earliest detectable stage—before further network activation—may represent an opportunity for primary prevention, pending confirmation from interventional studies, with lipid screening reconceived as a gateway to comprehensive metabolic evaluation. Second, weight control in individuals who already exhibit dyslipidemia or hyperuricemia may help attenuate the obesity-associated risk of subsequent diabetes development before glucose dysregulation becomes established. Third, early, aggressive multi-target intervention before diabetes onset could keep patients within reversible trajectory zones, given that post-diabetes reversion is rare. Together, these observational findings provide a rationale for exploring a shift from reactive, single-disease management toward proactive, structure-informed cardiometabolic prevention, although interventional evidence is needed to validate this approach.

4.6 Strengths and Limitations

This study has several strengths. It applies multiple complementary analytical methods to the same cohort, enabling triangulation across approaches with different assumptions. The time-varying covariate Cox design, which updates exposure status at each annual visit, overcomes the immortal time bias and exposure misclassification inherent in baseline-only analyses. The directional contrast design, which simultaneously estimates both the A → B and B → A paths within the same cohort using identical methodology, strengthens the evidence for asymmetric effects that single-direction analyses cannot establish.

Several limitations must be acknowledged. First, disease status was determined by laboratory values and anthropometric measurements at annual visits. Participants receiving pharmacological treatment (statins, urate-lowering agents, antidiabetic medications) whose biomarkers were normalized at the time of assessment may have been misclassified as disease-free, leading to underestimation of true prevalence and overestimation of remission rates. Access to prescription records would have mitigated this issue but was unavailable. Second, the cohort comprised employees of a single occupational sector undergoing mandated health examinations, yielding a sample that was predominantly female, of working age, and with stable healthcare access. Generalization to elderly populations, rural communities, or uninsured groups should be made cautiously. Third, we lacked data on behavioral risk factors (diet composition, physical activity, alcohol consumption, smoking status) and pharmacotherapy, precluding assessment of whether observed directional predictions operated through behavioral, pharmacological, or purely biological pathways and leaving unmeasured confounding as a potential concern in the GEE and Cox models. Fourth, while time-varying Cox regression provides stronger evidence for directionality than cross-sectional association or baseline-only exposure, it does not establish causation. Formal causal inference would require randomized intervention or target trial emulation methods [52]. Fifth, temporal sequencing was determined by the order of diagnosis at annual health examinations; a condition detected at a later visit may have developed between visits, and diseases with more stable biomarker expression (such as BMI-defined obesity) may appear to precede conditions with more variable laboratory values (such as hyperuricemia) partly because of differential detection sensitivity at discrete annual measurement points. Sixth, the four-node network is small by complex network standards, and centrality metrics should be interpreted as structured summaries of pairwise associations rather than as emergent properties of a high-dimensional network [53]. Seventh, the one-year interval between examinations may have missed short-duration episodes of disease that arose and remitted between visits, potentially underestimating transition rates and biasing temporal sequencing toward more persistent conditions. Finally, the dyslipidemia definition used in the threshold sensitivity analysis (TC ≥ 5.2 mmol/L) was a simplification of the composite definition used in the main analysis; the actual composite definition incorporating TG, LDL-C, and HDL-C was more clinically appropriate.

5 Conclusions

Metabolic multimorbidity is not a random accumulation of diseases but a structurally organized, temporally ordered, and directionally predictable process. Dyslipidemia functions as the comorbidity network's central hub and the most frequent first metabolic condition to appear, while obesity serves as a parallel gateway that tends to precede other metabolic diseases when co-occurring and acts as the dominant upstream driver of diabetes through a strikingly unidirectional pathway. Once a multi-disease state is established, particularly one involving diabetes, it tends to self-perpetuate, with a reversion probability of only 1% per year. These

findings from a single-center retrospective cohort suggest that cardiometabolic disease prevention may benefit from a structural perspective: lipid screening could be reconsidered as a gateway to comprehensive metabolic evaluation, BMI assessment may warrant equal sentinel status for triggering metabolic workup, weight management deserves consideration as a key lever for diabetes prevention in at-risk individuals, and the window for effective intervention may close earlier than most clinical pathways assume. Prospective multicenter studies and interventional trials are needed to confirm these observational findings and translate them into clinical recommendations.

Abbreviations

AIC
Akaike information criterion
BFGS
Broyden–Fletcher–Goldfarb–Shanno (optimization algorithm)
BIC
Bayesian information criterion
BMI
Body mass index
CI
Confidence interval
FDR
False discovery rate (Benjamini–Hochberg procedure)
GBTM
Group-based trajectory modeling
GEE
Generalized estimating equations
HDL-C
High-density lipoprotein cholesterol
HR
Hazard ratio
LCA
Latent class analysis
LDL-C
Low-density lipoprotein cholesterol
LOESS
Locally estimated scatterplot smoothing
LRT
Likelihood ratio test
OM
Optimal matching (sequence analysis)
OR
Odds ratio

SD
Standard deviation
TC
Total cholesterol
TG
Triglycerides

Declarations

Ethics approval and consent to participate: This study was approved by the institutional review board of Beijing Chao-Yang Hospital (approval number: 2026-ke-775). Individual informed consent was waived because data were de-identified and retrospectively collected.

Consent for publication: The author(s) declare(s) that they have no competing interests.

Availability of data and materials: The health checkup cohort data are not publicly available due to institutional privacy restrictions but are available from the corresponding author on reasonable request.

Competing interests: The authors declare that they have no competing interests.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 82372304), the Beijing Natural Science Foundation (Grant No. L256023), and the Beijing Natural Science Foundation (Grant No. L254078).

Authors' contributions: L.L. Deng contributed to manuscript writing, figure preparation and statistical analysis of the data. X.Y. Jiang and R.J. Li was responsible for data curation. H. Zhao contributed to figure preparation. M.Y. Yuan contributed to data curation. C.X. Wang contributed to manuscript writing. R. Zhao contributed to data curation. Y.W. Hu was responsible for reviewed and edited the manuscript. G.W. Fan designed the experiments, supervised the research, and reviewed and edited the manuscript.

Acknowledgements: Not applicable.

References

1. Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract. 2022;183:109119.
2. Pan XF, Wang L, Pan A. Epidemiology and determinants of obesity in China. Lancet Diabetes Endocrinol. 2021;9(6):373–92.
3. Zhang M, Zhu X, Wu J, Huang Z, Zhao Z, Zhang X, et al. Prevalence of Hyperuricemia Among Chinese Adults: Findings From Two Nationally Representative Cross-Sectional Surveys in 2015-16 and 2018-19. Front Immunol. 2022;12:791983.
4. Mottillo S, Filion KB, Genest J, Joseph L, Pilote L, Poirier P, et al. The metabolic syndrome and cardiovascular risk a systematic review and meta-analysis. J Am Coll Cardiol. 2010;56(14):1113–32.

5. Chowdhury SR, Chandra Das D, Sunna TC, Beyene J, Hossain A. Global and regional prevalence of multimorbidity in the adult population in community settings: a systematic review and meta-analysis. EClinicalMedicine. 2023;57:101860.
6. Tran J, Norton R, Conrad N, Rahimian F, Canoy D, Nazarzadeh M, et al. Patterns and temporal trends of comorbidity among adult patients with incident cardiovascular disease in the UK between 2000 and 2014: a population-based cohort study. PLoS Med. 2018;15(3):e1002513.
7. Calderón-Larrañaga A, Vetrano DL, Ferrucci L, Mercer SW, Marengoni A, Onder G, et al. Multimorbidity and functional impairment: bidirectional interplay, synergistic effects and common pathways. J Intern Med. 2019;285(3):255–71.
8. Saklayen MG. The Global Epidemic of the Metabolic Syndrome. Curr Hypertens Rep. 2018;20(2):12.
9. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011;12(1):56–68.
10. Dong G, Feng J, Sun F, Chen J, Zhao XM. A global overview of genetically interpretable multimorbidity clusters among common diseases in the UK Biobank. Genome Med. 2021;13(1):110.
11. Hidalgo CA, Blumm N, Barabási AL, Christakis NA. A dynamic network approach for the study of human phenotypes. PLoS Comput Biol. 2009;5(4):e1000353.
12. Lee DS, Park J, Kay KA, Christakis NA, Oltvai ZN, Barabási AL. The implications of human metabolic network topology for disease comorbidity. Proc Natl Acad Sci U S A. 2008;105(29):9880–5.
13. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol. 2010;6:109–38.
14. Nagin DS, Jones BL, Passos VL, Tremblay RE. Group-based multi-trajectory modeling. Stat Methods Med Res. 2018;27(7):2015–23.
15. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Stat Med. 2007;26(11):2389–430.
16. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. Annu Rev Public Health. 1999;20:145–57.
17. Joint Committee for Guideline Revision. 2016 Chinese guidelines for the management of dyslipidemia in adults. J Geriatr Cardiol. 2018;15(1):1–29.
18. Sun M, Lyu Z, Wang C, Li Y, Zhao D, Ran X. 2024 Update of Chinese Guidelines for Diagnosis and Treatment of Hyperuricemia and Gout Part I: Recommendations for General Patients. Int J Rheum Dis. 2025;28(7):e70375.
19. Zhou BF, Cooperative Meta-Analysis Group of the Working Group on obesity in China. Predictive values of body mass index and waist circumference for risk factors of certain related diseases in Chinese adults. Biomed Environ Sci. 2002;15(1):83–96.
20. American Diabetes Association Professional Practice Committee. 2. Diagnosis and Classification of Diabetes: Standards of Care in Diabetes-2024. Diabetes Care. 2024;47(Suppl 1):S20–42.
21. Chinese Elderly Type 2 Diabetes Prevention and Treatment of Clinical Guidelines Writing Group; Geriatric Endocrinology and Metabolism Branch of Chinese Geriatric Society; Geriatric Endocrinology and Metabolism Branch of Chinese Geriatric Health Care Society; Geriatric Professional Committee of Beijing

Medical Award Foundation; National Clinical Medical Research Center for Geriatric Diseases (PLA General Hospital). Clinical guidelines for prevention and treatment of type 2 diabetes mellitus in the elderly in China (2022 edition). Zhonghua nei ke za zhi. 2022;61(1):12–50. Chinese.

22. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: a tutorial paper. Behav Res Methods. 2018;50(1):195–212.
23. Csardi G, Nepusz T, Traag V et al. igraph: Network Analysis and Visualization in R. R package version 2.1.1. 2024. https://CRAN.R-project.org/package=igraph
24. Field CA, Welsh AH. Bootstrapping clustered data. J R Stat Soc Ser B Stat Methodol. 2007;69(3):369–90.
25. Halekoh U, Højsgaard S, Yan J. The R package geepack for generalized estimating equations. J Stat Softw. 2006;15(2):1–11.
26. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986;73(1):13–22.
27. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol. 1995;57(1):289–300.
28. Gabadinho A, Ritschard G, Müller NS, et al. Analyzing and visualizing state sequences in R with TraMineR. J Stat Softw. 2011;40(4):1–37.
29. Proust-Lima C, Philipps V, Liquet B. Estimation of extended mixed models using latent classes and latent processes: the R package lcmm. J Stat Softw. 2017;78(2):1–56.
30. Steinley D. Properties of the Hubert-Arabie adjusted Rand index. Psychol Methods. 2004;9(3):386–96.
31. Jackson CH. Multi-State Models for Panel Data: The msm Package for R. J Stat Softw. 2011;38(8):1–29.
32. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. New York: Springer; 2000.
33. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. Ann Stat. 1982;10(4):1100–20.
34. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Struct Equ Model. 2007;14(4):535–69.
35. Linzer DA, Lewis JB. poLCA: an R package for polytomous variable latent class analysis. J Stat Softw. 2011;42(10):1–29.
36. King C, Lanaspa MA, Jensen T, Tolan DR, Sánchez-Lozada LG, Johnson RJ. Uric Acid as a Cause of the Metabolic Syndrome. Contrib Nephrol. 2018;192:88–102.
37. Sluijs I, Holmes MV, van der Schouw YT, Beulens JW, Asselbergs FW, Huerta JM. A Mendelian Randomization Study of Circulating Uric Acid and Type 2 Diabetes. Diabetes. 2015;64(8):3028–36.
38. Choi HK, Ford ES. Haemoglobin A1c, fasting glucose, serum C-peptide and insulin resistance in relation to serum uric acid levels—the Third National Health and Nutrition Examination Survey. Rheumatology. 2008;47(5):713–7.
39. Ginsberg HN, Packard CJ, Chapman MJ, Borén J, Aguilar-Salinas CA, Averna M, et al. Triglyceride–rich lipoproteins and their remnants: metabolic insights, role in atherosclerotic cardiovascular disease, and emerging therapeutic strategies—a consensus statement from the European Atherosclerosis Society. Eur Heart J. 2021;42(47):4791–806.

40. Shoelson SE, Lee J, Goldfine AB. Inflammation and insulin resistance. J Clin Invest. 2006;116(7):1793 – 801. doi: 10.1172/JCI29069. Erratum in: J Clin Invest. 2006;116(8):2308.
41. Ference BA, Kastelein JJP, Ray KK, Ginsberg HN, Chapman MJ, Packard CJ, et al. Association of triglyceride–lowering LPL variants and LDL–C–lowering LDLR variants with risk of coronary heart disease. JAMA. 2019;321(4):364–73.
42. Gong Q, Zhang P, Wang J, Ma J, An Y, Chen Y, et al. Morbidity and mortality after lifestyle intervention for people with impaired glucose tolerance: 30–year results of the Da Qing Diabetes Prevention Outcome Study. Lancet Diabetes Endocrinol. 2019;7(6):452–61.
43. Zhou X, Ji L, Ran X, Su B, Ji Q, Pan C, et al. Prevalence of dyslipidemia and its influence on achieving cardiometabolic goals in Chinese type 2 diabetes patients: an analysis of CCMR–3B study. PLoS ONE. 2016;11(1):e0144179.
44. Abdullah A, Peeters A, de Courten M, Stoelwinder J. The magnitude of association between overweight and obesity and the risk of diabetes: a meta–analysis of prospective cohort studies. Diabetes Res Clin Pract. 2010;89(3):309–19.
45. Diabetes Prevention Program Research Group. Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the Diabetes Prevention Program Outcomes Study. Lancet Diabetes Endocrinol. 2015;3(11):866–75.
46. Li G, Zhang P, Wang J, Gregg EW, Yang W, Gong Q, et al. The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. Lancet. 2008;371(9626):1783–9.
47. Whitty CJM, MacEwen C, Goddard A, Alderson D, Marshall M, Calderwood C, et al. Rising to the challenge of multimorbidity. BMJ. 2020;368:l6964.
48. Zhang D, Tang X, Shen P, Si Y, Liu X, Xu Z, et al. Multimorbidity of cardiometabolic diseases: prevalence and risk for mortality from one million Chinese adults in a longitudinal cohort study. BMJ Open. 2019;9(3):e024476.
49. Yan J, Wang M, Cao Y. Patterns of multimorbidity in association with falls among the middle-aged and older adults: results from the China Health and Retirement Longitudinal Study. BMC Public Health. 2022;22(1):1814.
50. Major TJ, Dalbeth N, Stahl EA, Merriman TR. An update on the genetics of hyperuricaemia and gout. Nat Rev Rheumatol. 2018;14(6):341–53.
51. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol. 2018;14(2):88–98.
52. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016;183(8):758–64.
53. Bringmann LF, Elmer T, Epskamp S, Krause RW, Schoch D, Wichers M, et al. What do centrality measures measure in psychological networks? J Abnorm Psychol. 2019;128(8):892–903.

Figures

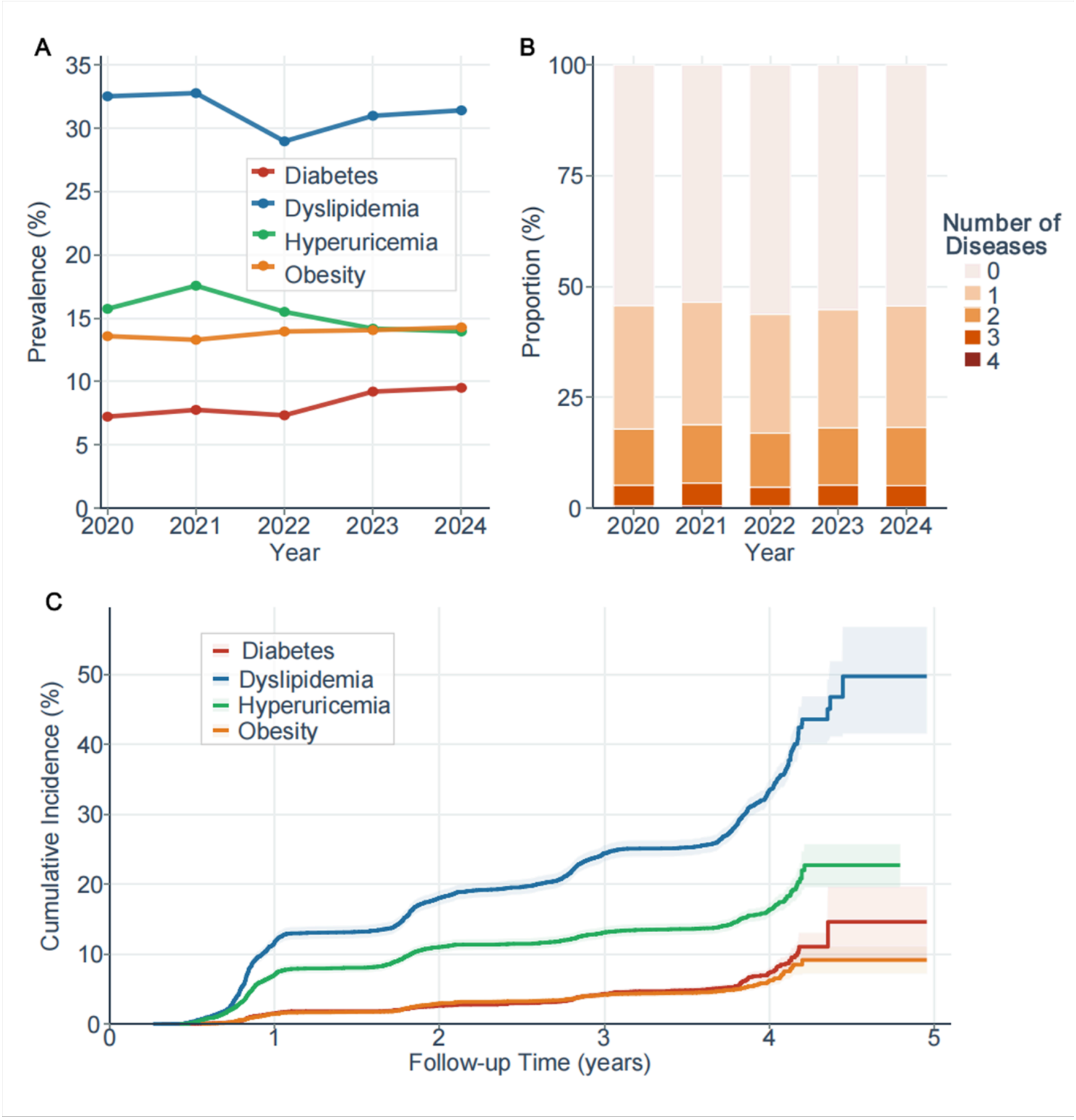


Figure 1

Prevalence trends and comorbidity burden of metabolic diseases, 2020–2024. (A) Annual prevalence of four metabolic diseases among 8,383 participants with 5 annual health examinations. (B) Stacked bar chart showing the distribution of metabolic comorbidity burden (number of co-occurring diseases: 0, 1, 2, 3, or 4) by calendar year. (C) Kaplan-Meier cumulative incidence curves for each metabolic disease among participants disease-free at baseline.

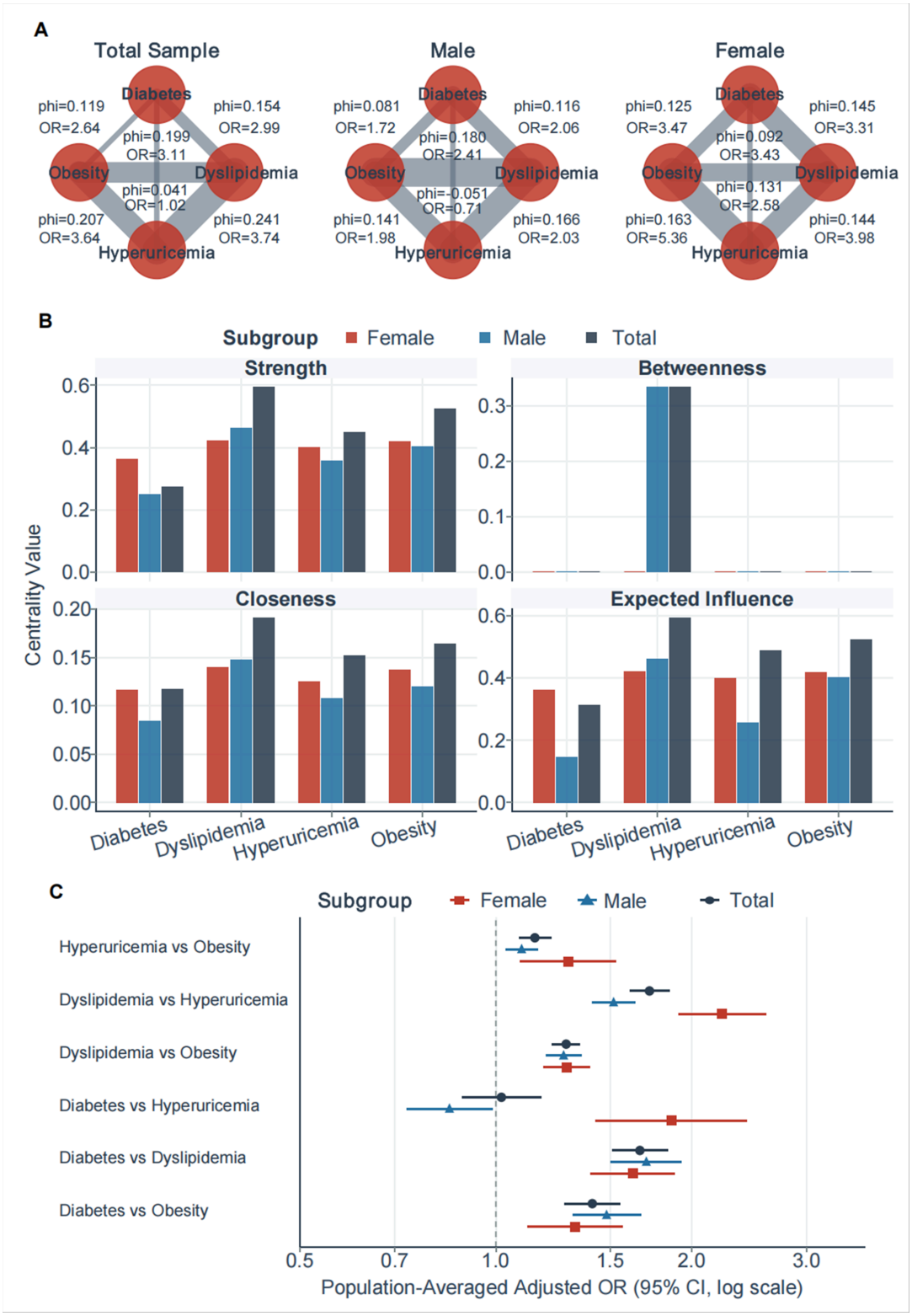


Figure 2

Comorbidity network analysis: pairwise associations, centrality metrics, and network structure. (A) Network diagrams for the total sample, males, and females. (B) Network centrality metrics (Strength, Betweenness, Closeness, and Expected Influence) by sex subgroup. (C) Forest plot of population-averaged adjusted ORs from GEE with exchangeable working correlation structure. Total sample models were adjusted for age and sex; sex-stratified models were adjusted for age only.

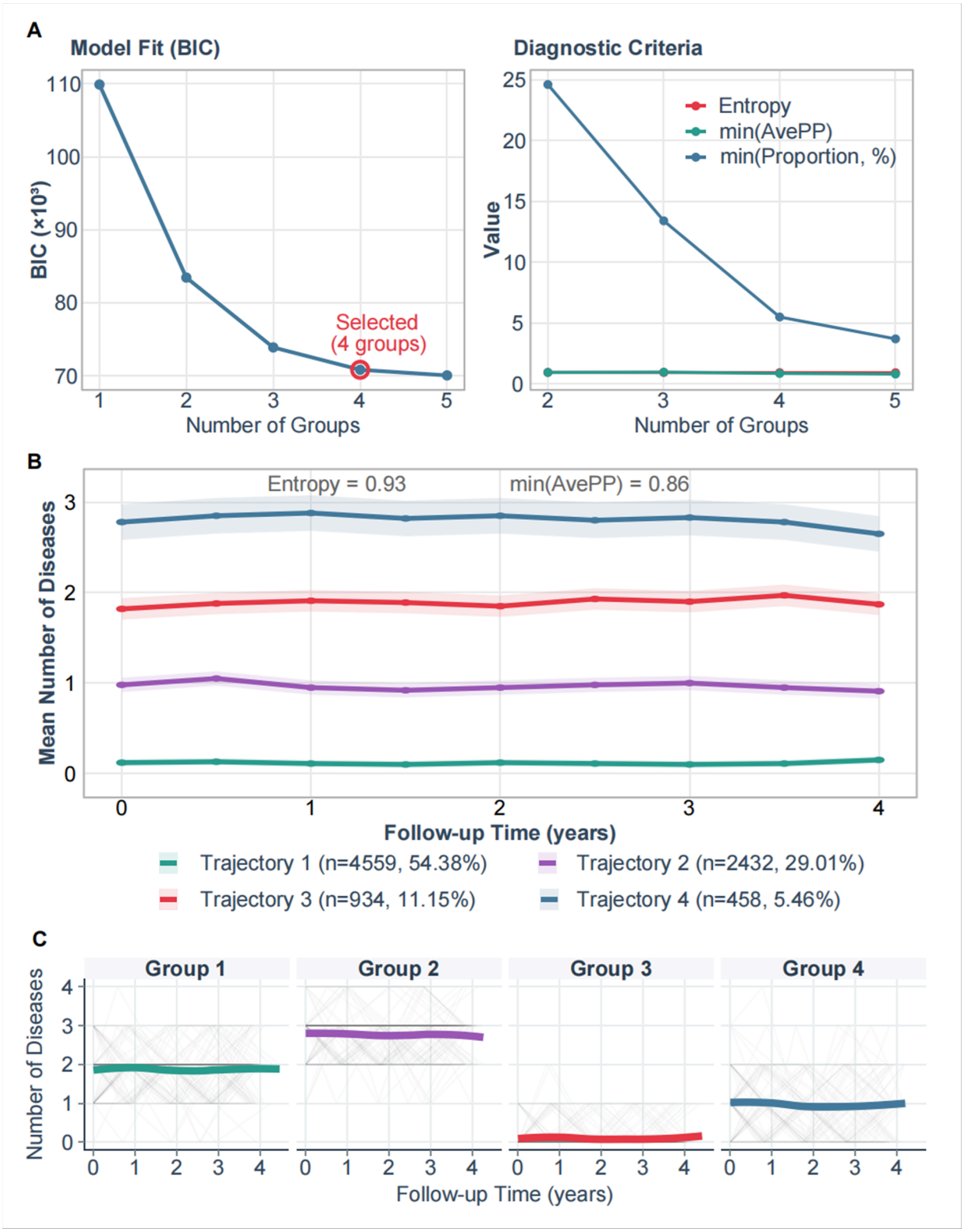


Figure 3

GBTM of metabolic comorbidity count. (A) BIC and diagnostic criteria across model specifications. (B) Mean trajectory of the number of metabolic diseases over 5 years of follow-up for each of the four identified groups. (C) Spaghetti plots showing individual-level trajectories (gray lines, random subsample of 200 per group) overlaid with group-level locally estimated scatterplot smoothing (LOESS) (colored lines) for each trajectory group.

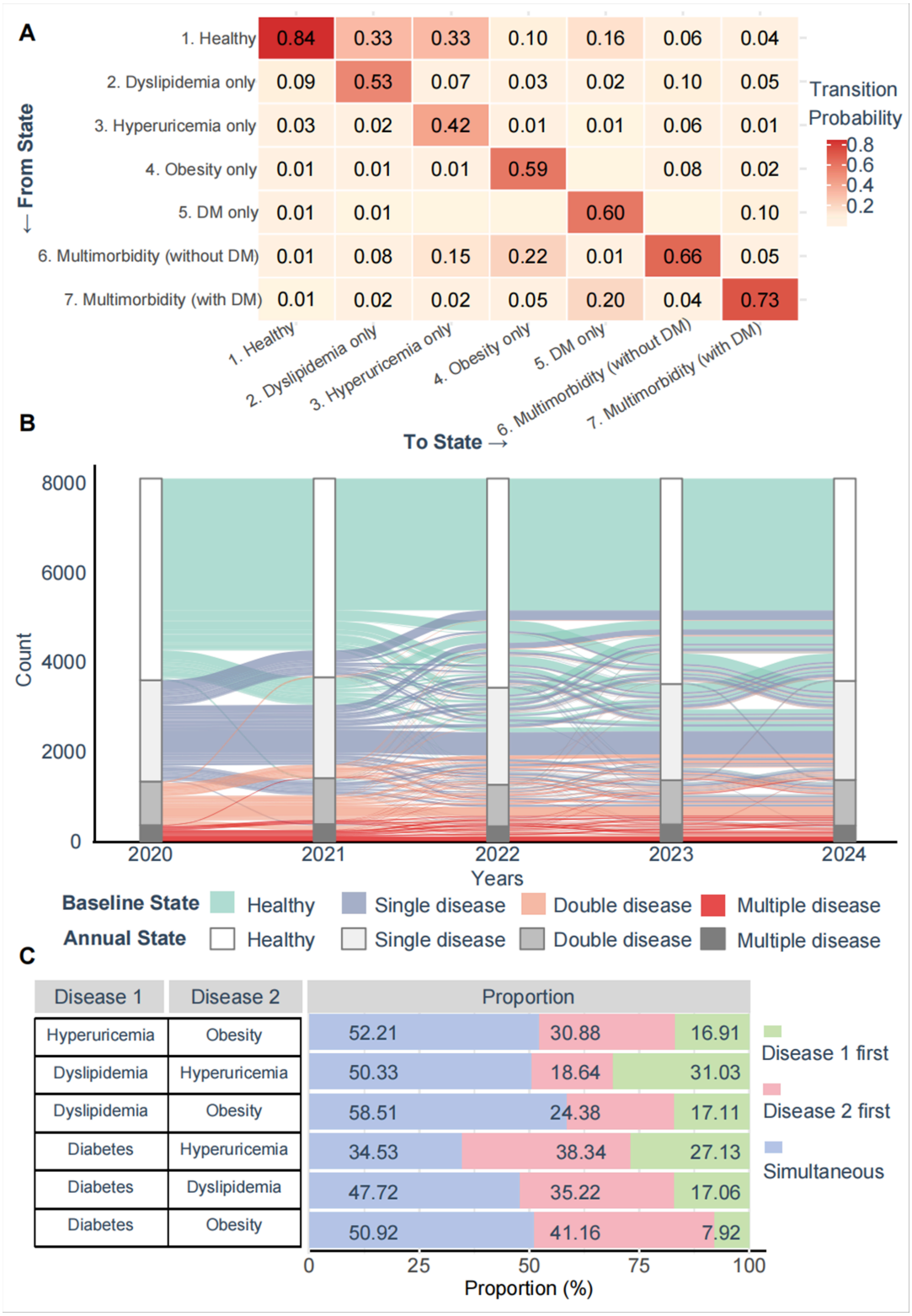


Figure 4

Multistate disease transitions, temporal progression patterns, and disease onset sequencing. (A) One-year transition probability matrix from the continuous-time Markov model with seven states. (B) Alluvial diagram illustrating individual-level transitions between comorbidity states (Healthy, Single disease, Double disease, Multiple disease) across 2020–2024. (C) Temporal ordering of disease pairs among individuals who developed both diseases during follow-up.

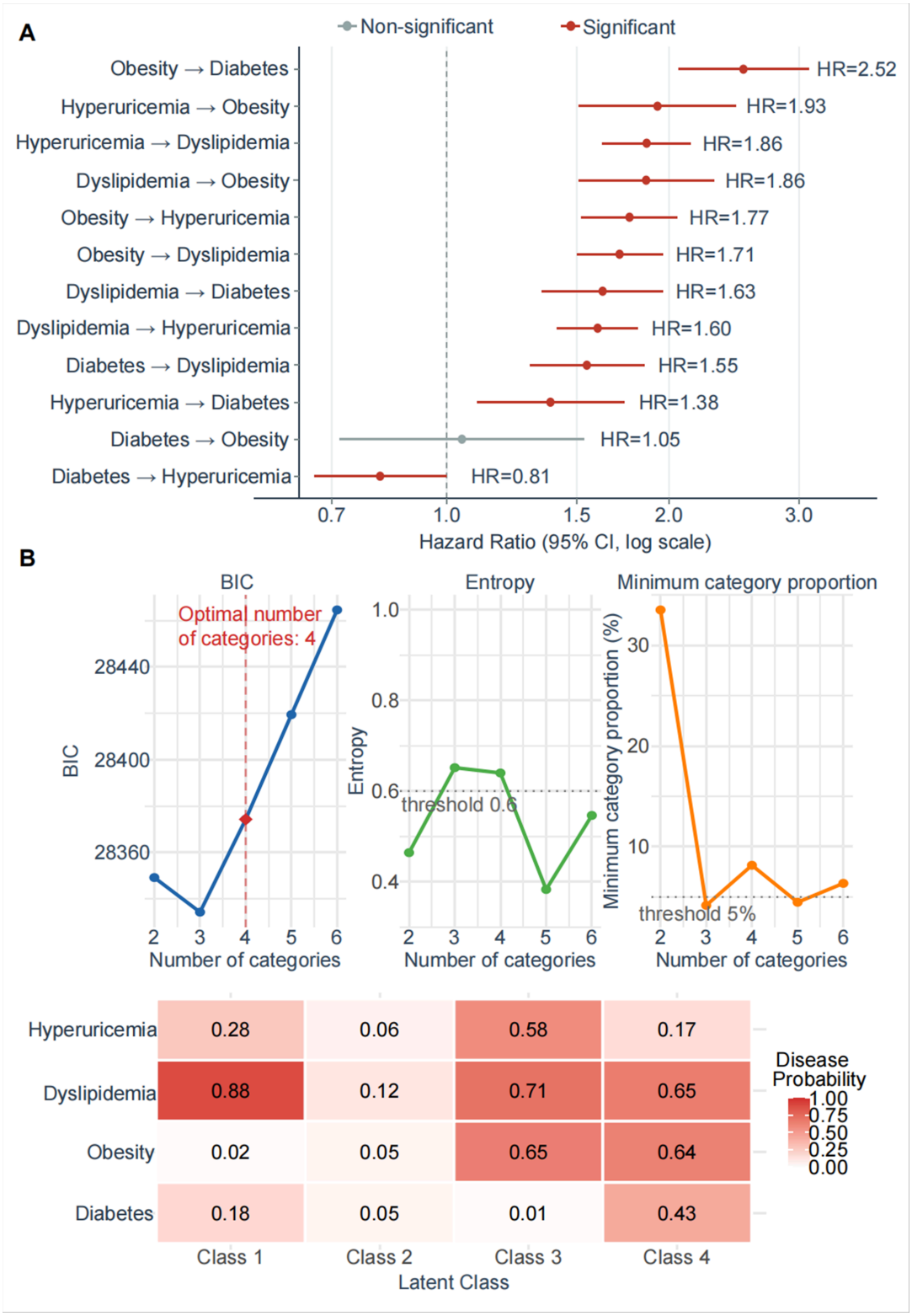


Figure 5

Longitudinal predictive effects between metabolic diseases, cumulative incidence, and LCA. (A) Forest plot of HRs from time-varying covariate Cox proportional hazards models. Each of the 12 directional pathways represents the predictive effect of a pre-existing metabolic disease (time-varying exposure) on the new onset of another disease, adjusted for age, sex, and all remaining diseases as time-varying covariates. Red points indicate statistical significance after Benjamini–Hochberg FDR correction (*p*-FDR < 0.05); gray points indicate

non-significance. (B) Diagnostic indices for optimal class enumeration in LCA. (C) LCA identifying four multimorbidity patterns. Heatmap shows conditional probabilities of each disease within each latent class.

Supplementary Files

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

- SupplementaryTableS17.xlsx