

Global burden structure of chronic kidney disease due to type 2 diabetes: a SHAP-based spatiotemporal model-attribution analysis across 203 countries and territories

Zeyu Guo

Chongqing Medical University

Qais Ahmad Naseer

Chongqing Medical University

Zhaoyu Zhai

Chongqing Medical University

Yichen Guo

Chongqing Medical University

Fengyuan Yang

Chongqing Medical University

Ying Ye

Second Affiliated Hospital of Chongqing Medical University

Jianbo Pan

pan.jianbo@cqmu.edu.cn

Chongqing Medical University

Research Article

Keywords: Chronic kidney disease (CKD), Type 2 diabetes mellitus (T2DM), Global Health, Comorbidity, Machine learning, Shapley Additive Explanations (SHAP)

Posted Date: July 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10032043/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

Chronic kidney disease due to type 2 diabetes (CKD-T2DM) is an increasing contributor to the global kidney disease burden, yet its country-level burden structure remains incompletely characterised. Using Global Burden of Disease 2021 estimates, we analysed adults aged 45 years or older in 203 countries and territories from 1990 to 2021. T2DM-related and non-diabetic CKD-related indicators were incorporated into an XGBoost model with CKD-T2DM disability-adjusted life-years as the outcome, and SHAP was used to quantify model-attributed feature contributions. We developed the relative driver structure ratio (RDSR) to compare these contributions across countries and over time. Global RDSR increased from 0.21 (95% credible interval 0.14–0.27) in 1990 to 0.43 (0.37–0.49) in 2021, with higher values at higher Socio-demographic Index. Pacific Island countries and territories showed the highest CKD-T2DM burden and elevated body-mass index exposure. These findings describe model-attributed burden patterns rather than causal effects and may inform integrated diabetes-kidney surveillance.

Introduction

Chronic kidney disease (CKD) is a major and growing public-health challenge, affecting more than 10% of the global population, causing more than 1 million deaths annually, and contributing substantially to disability-adjusted life-year (DALY) loss[1]. Recent Global Burden of Disease (GBD) 2021 analyses indicate that diabetes and kidney diseases are among the non-communicable disease categories with increasing age-standardised DALY rates worldwide, underscoring their persistent relevance for health systems[2]. Type 2 diabetes mellitus (T2DM) is a major metabolic contributor to CKD burden, accounting for an estimated 30–50% of global CKD burden[3–5]. At the same time, population ageing and the obesity epidemic have increased the importance of non-diabetic kidney injury, including obesity-related nephropathy, among populations with diabetes[6]. These overlapping metabolic and kidney-disease processes contribute to haemodynamic overload, systemic inflammation and glomerulosclerosis, accelerating CKD progression and increasing the complexity of prevention and management[7–9]. CKD due to T2DM (CKD-T2DM) therefore represents a clinically important and health-system-relevant CKD subtype with substantial implications for long-term morbidity, mortality and resource use[10, 11].

Previous studies have provided important estimates of the overall burden of CKD-T2DM[12]. However, less is known about how the country-level burden structure of CKD-T2DM relates to T2DM-related and non-diabetic CKD-related indicators over time. This distinction is important because countries with similar overall CKD-T2DM burden may differ in the relative balance of disability, premature mortality and underlying metabolic or kidney-disease profiles. Global variation in Socio-demographic Index, metabolic risk exposure and health-system capacity may further shape how CKD-T2DM burden is distributed across regions. A clearer description of these country-level burden structures could support more targeted surveillance and help identify settings where integrated diabetes-kidney strategies may be most relevant.

Interpretable machine-learning methods provide one approach for summarising complex country-level relationships among related burden indicators. In this setting, SHAP-based model attribution can describe how selected T2DM-related and CKD-nDM-related indicators contribute to modelled CKD-T2DM DALY patterns. To summarise the relative balance of these model-attributed contributions, we used the relative driver structure ratio (RDSR), a SHAP-derived index comparing T2DM-related and CKD-nDM-related feature contributions across countries and over time. Because both predictors and outcomes are derived from GBD estimates, these analyses should be interpreted as descriptive and model-attributed rather than causal. Nevertheless, when framed appropriately, this approach can help characterise spatial and temporal heterogeneity in CKD-T2DM burden and provide a reproducible framework for comparing countries and regions.

In this study, we used GBD 2021 data to characterise the burden structure of CKD-T2DM among adults aged 45 years or older across 203 countries and territories from 1990 to 2021. We assessed the relative model-attributed contributions of T2DM-related and CKD-nDM-related indicators using XGBoost and SHAP, applied RDSR to summarise their balance, and examined spatial clustering and temporal trajectories through 2050. By describing these country-level burden patterns, this study aims to support more precise CKD-T2DM surveillance and inform integrated public-health planning for diabetes and kidney disease.

Methods

Overview

We conducted a country-level spatiotemporal analysis using Global Burden of Disease (GBD) 2021 estimates to characterise the burden structure of chronic kidney disease due to type 2 diabetes (CKD-T2DM) among adults aged 45 years or older in 203 countries and territories from 1990 to 2021. The primary outcome was age-standardised disability-adjusted life-years (DALYs) for CKD-T2DM. We compiled country-level burden indicators for type 2 diabetes mellitus (T2DM) and non-diabetic chronic kidney disease (CKD-nDM), including incidence, prevalence, mortality, years of life lost (YLLs) and years lived with disability (YLDs). These indicators were used in a workflow that combined feature screening, Extreme Gradient Boosting (XGBoost) modelling and Shapley additive explanations (SHAP) to summarise model-attributed feature contributions. Complementary regression analyses were used to examine statistical interaction patterns between T2DM-related and CKD-nDM-related indicators. We then assessed the spatial distribution and temporal evolution of CKD-T2DM burden structures across countries and territories, including projected trajectories to 2050. Because the predictors and outcome were derived from GBD estimates, all model outputs were interpreted as descriptive, model-attributed patterns rather than causal effects.

DALYs for CKD-T2DM and burden indicators for T2DM and CKD-nDM were obtained from the GBD 2021 Results Tool. CKD-nDM was defined as CKD due to glomerulonephritis, hypertension and other or unspecified causes, aggregated according to the GBD classification[13]. We extracted estimates for adults aged 45 years or older across 203 countries and territories between 1990 and 2021, and restandardised all indicators to the GBD global standard population

aged 45–95 years[14]. To support interpretation of country-level patterns, we also extracted 2021 exposure levels for the ten leading CKD-T2DM risk factors and the Socio-demographic Index (SDI) for each country or territory.

Spatial boundaries were obtained from the ArcGIS World Vector Map Set and linked to country-level and territory-level burden indicators. Maps were generated for 203 analytical units. Tokelau was omitted from mapping because of spatial incompatibility, consistent with previous work[15]. All data used in this study are publicly available through the GBD 2021 Results Tool.

Feature screening

We extracted five burden indicator classes—incidence, prevalence, mortality, years of life lost (YLLs) and years lived with disability (YLDs)—for T2DM and CKD-nDM with age-standardised CKD-T2DM DALYs defined as the modelling outcome. To harmonise comparisons across disease categories and burden measures, source-specific T2DM and CKD-nDM indicators were standardised and evaluated within their corresponding candidate burden domains.

Feature screening was conducted in a stepwise framework before machine-learning analysis. Spearman correlation analysis was first used to assess monotonic associations between candidate burden domains and CKD-T2DM DALYs. Least absolute shrinkage and selection operator (LASSO) regression was then applied to reduce non-informative variables, and variance inflation factor (VIF) assessment was used to identify and remove collinear variables. The remaining candidate domains were further evaluated using negative binomial regression. Indicator domains retained after this screening process were carried forward for subsequent modelling; where applicable, retained domains were decomposed into their source-specific T2DM-related and CKD-nDM-related indicators for XGBoost-SHAP analyses and spatial assessment. These procedures were used to select informative model inputs and were not interpreted as evidence of causal effects

Predictive model performance and feature attribution

To characterise country-level CKD-T2DM burden patterns, retained burden domains were decomposed into source-specific T2DM-related and CKD-nDM-related features and incorporated into an Extreme Gradient Boosting (XGBoost) model with age-standardised CKD-T2DM DALYs as the outcome. The XGBoost model was used to capture non-linear relationships among country-level burden indicators and to generate model-based predictions of CKD-T2DM DALYs. Model performance was evaluated using year-wise five-fold cross-validation, with the coefficient of determination (R^2), mean absolute error (MAE) and root mean squared error (RMSE) used to summarise predictive accuracy across countries and territories.

To assess temporal robustness, the model trained on 2021 data was applied without retraining to each earlier calendar year from 2010 to 2020. Predictive performance was evaluated across countries and territories using R^2 , MAE and RMSE. Hyperparameters were tuned using RandomizedSearchCV with cross-validation on the training data; the tuning space and final 2021 model settings are reported in Supplementary Table 1. Prespecified model questions and success criteria for within-year accuracy, temporal robustness, feature-attribution stability and interpretability are summarised in Supplementary Box 1.

Model-attributed feature contributions were quantified using Shapley additive explanations (SHAP). SHAP values were used to describe the direction and magnitude of each retained T2DM-related and CKD-nDM-related feature's contribution to the fitted XGBoost model. To summarise the relative contribution of T2DM-related versus CKD-nDM-related features, SHAP values were further evaluated using complementary regression analysis and LMG variance decomposition. Temporal stability of feature attribution was assessed by comparing SHAP- and LMG-based feature rankings from 2017 to 2020 with those from 2021, and by evaluating concordance in the direction of SHAP values across country-year observations. All SHAP-based outputs were interpreted as descriptive model-attributed patterns rather than causal effects.

Hierarchical clustering and interaction regression

To identify countries and territories with similar CKD-T2DM burden-structure profiles, we performed hierarchical clustering using SHAP-derived model-attribution patterns from the retained T2DM-related and CKD-nDM-related features. Clustering was conducted at the country and territory level to group analytical units with similar relative contributions of T2DM-related and CKD-nDM-related indicators to the fitted XGBoost model. The resulting SHAP-derived clusters were used to describe heterogeneity in model-attributed CKD-T2DM burden structures across countries and territories.

Within each SHAP-derived cluster, complementary regression models were fitted to examine statistical interaction patterns between T2DM-related and CKD-nDM-related model-attributed features. Interaction terms were used to assess whether the association between one feature group and modelled CKD-T2DM DALYs differed according to the level of the other feature group within the same cluster. Statistical support for interaction terms was evaluated using p values, with $p < 0.05$ considered evidence of a statistically supported interaction pattern. These analyses were interpreted as within-cluster statistical patterns of model attribution rather than causal or biological interactions. The clustering and interaction-regression analyses were used to support interpretation of country-level heterogeneity in CKD-T2DM burden structure and to provide context for subsequent spatial and temporal analyses of the relative driver structure ratio (RDSR).

Calculation of the relative driver structure ratio (RDSR)

To summarise the relative balance of T2DM-related and CKD-nDM-related model-attributed contributions to CKD-T2DM burden, we defined the relative driver structure ratio (RDSR). RDSR was designed as a SHAP-derived comparative index for cross-country and temporal analyses, rather than as a causal measure of disease attribution. For each country or territory k , we first summed the absolute TreeSHAP contributions of predictors linked to T2DM and CKD-nDM as:

$$\phi_{k,j} = \sum_{j \in A} \phi_{k,j} \text{ and } \phi_{k,j} = \sum_{j \in B} \phi_{k,j}$$

where $\phi_{k,j}$ denotes the TreeSHAP value of predictor j for country or territory k , and A and B denote the sets of T2DM-related and CKD-nDM-related predictors, respectively. RDSR for country or territory k was then defined as:

$$R_{k,D} = \frac{\phi_{k,D} - \phi_{k,C}}{\phi_{k,D} + \phi_{k,C}}$$

..

RDSR values range from -1 to $+1$. Values greater than 0 indicate a larger relative contribution from T2DM-related features within the fitted model, whereas values less than 0 indicate a larger relative contribution from CKD-nDM-related features. Values closer to 0 indicate a more balanced model-attribution structure. Supplementary Fig. 1 provides a schematic illustration of the RDSR calculation.

To evaluate the stability of RDSR under scaling assumptions, sensitivity analyses were conducted by linearly rescaling predictor sets and CKD-T2DM DALYs. RDSR was unchanged when only predictor sets were rescaled and varied by approximately 20% or less when DALYs were rescaled, while the relative ordering of clusters remained stable (Supplementary Text 2; Supplementary Table 2). These results support the use of RDSR as a descriptive index for comparing model-attributed burden structures across countries and years.

Spatial analysis of cluster-specific burden patterns

We mapped countries and territories according to SHAP-derived CKD-T2DM burden-structure clusters and summarised cluster composition across the five 2021 Socio-demographic Index (SDI) categories. The association between SDI and RDSR was assessed using Pearson correlation. To compare cluster-level exposure profiles, we calculated mean summary exposure values (SEVs) for ten major CKD-T2DM risk factors in 2021, normalised each cluster-level estimate to the corresponding global average, and visualised relative exposure patterns using a heatmap. These analyses were used to describe how model-attributed CKD-T2DM burden structures varied across socioeconomic contexts and risk-factor exposure profiles.

Temporal trend modelling and long-term projections

Year-specific XGBoost models were trained for each calendar year from 1990 to 2021, and SHAP values were used to compute annual country-level and territory-level RDSR values. We then fitted Bayesian hierarchical time-series models using integrated nested Laplace approximation (INLA) to estimate historical RDSR trajectories from 1990 to 2021 and project trajectories to 2050. Calendar year was modelled as the temporal component, with country and territory included as a random intercept to account for repeated observations within each analytical unit. Uncertainty in the estimated and projected trajectories was summarised using posterior credible intervals. Model fit was assessed using standardised residuals, Kolmogorov-Smirnov tests, residual histograms and Q-Q plots.

AI Use Statement

AI-assisted language editing was used to improve clarity, structure and scientific presentation during manuscript revision. No AI tools were used to generate data, conduct statistical analyses, produce figures or derive results. All scientific content, analyses, interpretations and final wording were reviewed and approved by the authors.

Results

Feature screening and model performance

We analysed age-standardised CKD-T2DM DALY rates and related T2DM and CKD-nDM burden indicators among adults aged 45 years or older across 203 countries and territories. Among the five candidate burden indicator domains assessed; incidence, prevalence, mortality, YLLs and YLDs, the incidence, YLDs and YLLs met the screening criteria and were retained for subsequent modelling with CKD-T2DM DALYs as the outcome. In negative binomial regression, YLDs (RR 1.25, 95% CI 1.07–1.47) and YLLs (RR 1.73, 95% CI 1.54–1.95) were statistically supported predictors, whereas incidence was not (RR 1.04, 95% CI 0.90–1.20; Fig. 1A). These findings indicate that cross-country variation in CKD-T2DM DALYs was more strongly captured by disability and premature-mortality burden domains than by incidence.

The 2021 XGBoost model, using T2DM-related and CKD-nDM-related YLL and YLD features, showed good five-fold cross-validation performance, with an R^2 of 0.754, MAE of 211.2 and RMSE of 313.4 per 100,000 population. Predicted values closely tracked observed CKD-T2DM DALY rates without clear systematic bias (Fig. 1B). Across 1990–2021, year-wise cross-validation performance remained stable, with annual R^2 values ranging from 0.73 to 0.80 (Supplementary Fig. 2). Temporal validation using the 2021-trained model to predict earlier calendar years from 2010 to 2020 yielded a median annual R^2 of 0.86 (range 0.85–0.93), with MAE and RMSE 21.0% and 12.5% lower than the corresponding within-year cross-validation estimates. Detailed year-wise cross-validation and temporal-validation metrics are provided in Supplementary Table 3 and visualised in Supplementary Fig. 3a,b.

SHAP-based feature attribution

In 2021, SHAP analysis showed clear differences in the direction and magnitude of the four retained feature groups, T2DM-related YLLs, T2DM-related YLDs, CKD-nDM-related YLLs and CKD-nDM-related YLDs in the fitted CKD-T2DM DALY model (Fig. 1C). Feature-attribution rankings were stable across recent years, with Spearman correlations of 0.80-1.00 when 2017–2020 rankings were compared with 2021. SHAP sign concordance was at least 84% in each year, with median yearly concordance ranging from 87% to 97%. T2DM-related features showed larger and more consistently positive model-attributed contributions, whereas CKD-nDM-related features were smaller and more heterogeneous. LMG variance decomposition supported this pattern, assigning 65.3% of explained variance to T2DM-related features and 34.7% to CKD-nDM-related features. Detailed SHAP sign-concordance and feature-ranking stability results are provided in Supplementary Table 4 and visualised in Supplementary Fig. 3c,d.

SHAP-derived burden-structure clusters

Using SHAP-derived profiles from T2DM-related and CKD-nDM-related YLL and YLD features, we grouped the 203 countries and territories into four burden-structure clusters (Fig. 2). Age-standardised CKD-T2DM DALY rates increased stepwise across clusters: Cluster 2 (n = 63), 162.71 per 100,000 population (IQR 112.29-236.54); Cluster 1 (n = 71), 547.35 (405.17-685.28); Cluster 4 (n = 60), 1063.88 (825.35-1278.63); and Cluster 3 (n = 9), 2687.13 (2545.59-2944.44). Cluster 2 showed low or negative SHAP contributions for both T2DM-related and CKD-nDM-related components. Clusters 1 and 4 had broadly similar component burdens but differed in the direction of SHAP-attributed contributions. Cluster 3 showed the highest burden profile, with pronounced T2DM-related model-attributed contributions.

Within-cluster regression analyses showed burden-structure-specific statistical interaction patterns (Table 1). No statistically supported T2DM-related × CKD-nDM-related interaction terms were observed in the low-burden Cluster 2 (all $p > 0.05$). Cluster 1 showed two negative statistical interaction patterns, involving T2DM_YLLs × CKD-nDM_YLLs and T2DM_YLDs × CKD-nDM_YLLs. In the higher-burden Clusters 3 and 4, two positive statistical interaction patterns were observed, involving T2DM_YLDs × CKD-nDM_YLLs and T2DM_YLDs × CKD-nDM_YLDs. These interaction results describe within-cluster statistical patterns of model attribution rather than causal or biological interactions. Full cluster-specific interaction-regression outputs are provided in Supplementary Table 5.

Table 1
Cluster-specific statistical interaction patterns between T2DM-related and CKD-nDM-related SHAP features

Cluster	SHAP_T2DM	SHAP_CKD-nDM	Std.Error	Statistic	P.Value	Significance
Cluster 1	YLLs	YLLs	0.0012	-3.5913	0.0006	***
		YLDs	0.0018	-1.7632	0.0824	
	YLDs	YLLs	0.0016	-3.4506	0.0010	***
		YLDs	0.0022	-0.6955	0.4891	
Cluster 2	YLLs	YLLs	0.0041	-0.8658	0.3901	
		YLDs	0.0015	0.1054	0.9164	
	YLDs	YLLs	0.0095	0.5081	0.6133	
		YLDs	0.005	-0.3178	0.7518	
Cluster 3 and 4	YLLs	YLLs	0.0019	1.5449	0.1272	
		YLDs	0.0007	-0.7145	0.4775	
	YLDs	YLLs	0.0019	3.4649	0.0009	***
		YLDs	0.0005	4.0702	0.0001	***

Interaction terms were evaluated using cluster-specific regression models with age-standardised CKD-T2DM DALYs as the outcome and SHAP values of T2DM-related and CKD-nDM-related YLL and YLD features as predictors. The table reports the standard error, test statistic, p value and significance level for each interaction term. Statistical significance is indicated as $p < 0.05$, $p < 0.01$ and $p < 0.001$. These interaction results describe statistical patterns within the fitted model and should not be interpreted as causal or biological interactions. Abbreviations: CKD-T2DM, chronic kidney disease due to type 2 diabetes mellitus; CKD-nDM, non-diabetic chronic kidney disease; DALYs, disability-adjusted life-years; SHAP, Shapley additive explanations; T2DM, type 2 diabetes mellitus; YLDs, years lived with disability; YLLs, years of life lost.

Spatial distribution, SDI and risk-factor profiles

Spatial mapping showed distinct geographic distributions of the SHAP-derived clusters (Fig. 3; SDC Table 1). Cluster 2 was concentrated in Western and Northern Europe and parts of East Asia, including countries with low CKD-T2DM DALY rates such as Iceland (74.72 per 100,000) and Japan (263.59 per 100,000). Clusters 1 and 4 overlapped geographically across Latin America, South and Southeast Asia, and sub-Saharan Africa, but were stratified by burden; for example, Tanzania was classified in Cluster 1 with 363.11 DALYs per 100,000, whereas Uganda was classified in Cluster 4 with 813.98 DALYs per 100,000. Cluster 3 consisted mainly of Pacific Island countries and territories and had the highest burden levels, including American Samoa (3653.14 per 100,000) and the Federated States of Micronesia (2779.64 per 100,000).

Cluster composition differed across 2021 Socio-demographic Index (SDI) categories (Fig. 4A). Low-to-moderate SDI countries predominated in Cluster 1 (73.24%) and Cluster 4 (78.33%), whereas Cluster 2 was mainly composed of high or above-moderate SDI countries (79.37%). Cluster 3 was largely

composed of moderate-SDI settings and did not show a clear spatial pattern. SDI was positively correlated with the relative SHAP ratio comparing T2DM-related and CKD-nDM-related contributions ($r = 0.32$, 95% CI 0.20–0.44; $p = 2.38 \times 10^{-6}$; Supplementary Fig. 4). The ratio was concentrated around -0.2 to $+0.5$ at low-to-middle SDI levels, but increased at high-middle SDI and above, where most values were between $+0.5$ and $+0.8$.

Risk-factor exposure profiles also differed by cluster (Fig. 4B). Clusters 1 and 4 showed similar relative exposure patterns, including kidney dysfunction at 1.06 and 1.04 times the global mean, respectively, and high fasting plasma glucose at 1.16 and 1.10 times the global mean. Cluster 3 showed markedly elevated high body-mass index and high fasting plasma glucose exposures, at 1.96 and 2.25 times the global mean, respectively. Cluster 2 had higher exposure to processed meat and red meat intake, at 1.80 and 1.31 times the global mean, respectively, but lower high fasting plasma glucose exposure at 0.87 times the global mean. Cluster-level risk-factor exposure summaries are provided in Supplementary Table 7.

RDSR trajectories and projections

Residual diagnostics for the Bayesian INLA trajectory models are shown in Supplementary Fig. 5. Standardised residuals for the global model and all four cluster-specific models were approximately symmetric, with broadly bell-shaped histograms. Q-Q plots showed no major deviations from the 45° reference line, although slight tail deviations were observed. Kolmogorov-Smirnov D statistics ranged from 0.315 to 0.337, remaining below the prespecified diagnostic threshold of 0.35, although all tests were statistically significant ($p < 0.001$). Supplementary Table 6 provides 2021 country-level DALY, SHAP and RDSR summaries, whereas Fig. 5 presents Bayesian INLA-fitted RDSR trajectories and projections; therefore, these outputs were interpreted as complementary descriptive and trajectory-based summaries rather than interchangeable estimates.

Projected RDSR trajectories showed a sustained global increase in the relative model-attributed contribution of T2DM-related features (Fig. 5). Global RDSR increased from 0.21 (95% credible interval 0.14–0.27) in 1990 to 0.43 (0.37–0.49) in 2021 and was projected to reach 0.64 (0.58–0.71) by 2050. Cluster 2 remained the highest trajectory, increasing from 0.37 in 1990 to 0.62 in 2021 and 0.85 by 2050. Cluster 4 increased from 0.15 to 0.41 and 0.64 over the same period. Cluster 1 rose more gradually, from 0.10 in 1990 to 0.31 in 2021 and 0.50 by 2050. Cluster 3 showed the smallest change, increasing from 0.27 in 1990 to 0.30 in 2021 and 0.33 by 2050. These trajectories describe model-attributed RDSR patterns and should be interpreted as projected trends rather than causal forecasts. Year-specific global and cluster-level RDSR trajectory estimates are provided in Supplementary Table 8. Sensitivity projections based on $\pm 10\%$ random perturbation of SHAP values showed qualitatively similar global and cluster-specific RDSR trajectories, supporting the robustness of the projected burden-structure patterns (Supplementary Fig. 7). Country-level RDSR changes from 1990 to 2021 showed substantial within-cluster heterogeneity (Supplementary Fig. 6). Many countries and territories shifted toward higher T2DM-related model-attributed contribution, but the magnitude and direction of change varied across clusters. Several countries showed stable or decreasing RDSR patterns, indicating that cluster-average trajectories masked important country-level variation in CKD-T2DM burden structure.

Discussion

Using GBD 2021 data from 203 countries and territories, we systematically characterised the spatiotemporal burden structure of CKD-T2DM. A data-driven analysis identified YLLs and YLDs as the most informative burden domains, and we built predictive and clustering models using T2DM-related and CKD-nDM-related burden indicators as core inputs. Globally, the relative driver structure ratio (RDSR) increased from 0.21 to 0.43 between 1990 and 2021 and was projected to increase further by 2050. We further identified four national burden-structure patterns and quantified cluster-specific statistical interaction patterns between T2DM-related and CKD-nDM-related features. Overall, the rising RDSR and spatial heterogeneity suggest an increasing T2DM-related model-attributed contribution, together with regional differences in metabolic risk exposure and chronic-disease management, shaping the current distribution and projected trajectory of CKD-T2DM burden.

In the feature-screening analysis, YLLs and YLDs were the most informative burden domains for CKD-T2DM DALYs, whereas incidence was not statistically supported and prevalence and mortality were not retained for downstream modelling. This pattern may reflect disease natural history and metric definitions: CKD complications in T2DM often emerge years after diagnosis, with approximately 40% of patients developing clinically detectable kidney-function decline only after 5–10 years[16]. Such delayed progression is better captured by cumulative health-loss measures that reflect the gradual accumulation of complications over time, whereas annual incidence and point prevalence may be less sensitive to evolving burden. Cause-specific mortality may also incompletely capture CKD-T2DM burden because death estimates rely on underlying-cause-of-death coding[17], and T2DM more often contributes to long-term disability and multimorbidity than to direct death attribution[18]. Compared with crude mortality, YLLs capture premature mortality by weighting deaths at younger ages more heavily, while YLDs capture accumulated non-fatal disability regardless of cause-of-death attribution[17].

T2DM-related features showed a larger and increasingly prominent model-attributed contribution to CKD-T2DM burden, consistent with GBD evidence that diabetes-related DALYs have increased substantially since 1990 and are projected to continue rising by 2050[19]. We observed a steady increase in RDSR toward T2DM-related feature contribution; in high-SDI settings, higher RDSR may reflect persistent metabolic exposures, including obesity and lifestyle or dietary transitions, alongside longer survival and stronger detection within more developed health systems, rather than uniformly higher absolute CKD-T2DM burden[20]. Overall, this pattern supports earlier prevention, metabolic risk reduction and more systematic diabetes-kidney surveillance.

Cluster 3, composed mainly of Pacific Island countries and territories, showed a distinctive high-burden profile: T2DM-related DALYs and BMI exposure were high, whereas total CKD-nDM DALYs were below the global median. CKD-nDM-related features nevertheless contributed substantially to the SHAP-based attribution pattern. Decomposition of CKD-nDM DALYs suggested that this pattern may be partly explained by relatively lower YLLs than the global median (959.11 vs. 1764.92 per 100,000) but higher YLDs (321.18 vs. 170.88 per 100,000), with SHAP indicating a greater contribution from YLDs than YLLs. This pattern is consistent with a disability-weighted CKD-nDM burden structure in settings with high metabolic exposure and potential constraints in long-term chronic-disease management. Prior evidence is consistent with this interpretation: T2DM burden has increased across Pacific Island nations from 1990 to

2019, with Fiji among the highest-burden settings worldwide[21, 22], and ESKD incidence among Native Hawaiians and Pacific Islanders has been reported at approximately 921 per million, with earlier dialysis initiation but relatively better survival[23].

Clusters 1 and 4 both mainly comprised low-to-moderate SDI countries and territories and had broadly similar individual burdens of T2DM and CKD-nDM, yet CKD-T2DM burden in Cluster 4 was nearly twice that in Cluster 1. This contrast may reflect differences in health-system capacity, screening coverage, diagnostic timing and integrated chronic-disease management. Cluster 4 showed statistically supported positive T2DM-related $\times$ CKD-nDM-related interaction terms, consistent with an associative pattern in which the combined burden structure was higher within the fitted model. This pattern may align with settings where CKD screening coverage is limited, diagnoses are delayed, resources are fragmented and integrated management is lacking. In low- and middle-income countries, fewer than 30% of patients with T2DM receive CKD screening, with many cases detected at mid-to-advanced stages[24]. By contrast, Cluster 1 showed statistically supported negative interaction terms, suggesting a different model-attributed burden structure rather than biological “buffering.” Integrated diabetes-kidney care can improve coordination, and primary-care-based joint service frameworks have been proposed for resource-limited settings[25, 26]. These observations suggest that differences in chronic-disease service organisation may partly contribute to the contrast between Clusters 1 and 4, although this could not be directly tested using GBD data.

Cluster 2 comprised mainly higher-SDI countries and territories, with age-standardised DALY rates for both T2DM and CKD-nDM below the global average, despite dietary patterns in high-income settings such as higher red and processed meat consumption[27]. Interaction analysis showed no statistically supported positive interaction pattern, which may be consistent with more established detection and management pathways. CKD screening in higher-resource settings is often embedded in routine check-ups and chronic-disease workflows, rather than relying solely on T2DM or hypertension pathways[28]. Although T2DM-related burden was comparatively lower in this cluster, it still contributed to CKD-T2DM burden; therefore, health systems should continue strengthening targeted early detection, metabolic risk reduction and standardised T2DM management to limit future increases in CKD-T2DM burden.

Despite these findings, several limitations merit note. First, although SHAP helped identify influential model features and statistical interaction patterns, GBD estimates are observational and country-level; therefore, the relationships described in our analyses should be interpreted as associative rather than causal. Second, although RDSR provides a useful summary of the relative balance between T2DM-related and CKD-nDM-related model-attributed contributions, its magnitude depends on the XGBoost-TreeSHAP specification, feature selection and included predictors. In addition, summed absolute SHAP values capture the relative strength of model-attributed contributions but not biological directionality. Finally, systematic error, under-ascertainment or sparse primary data in GBD estimates and covariates may propagate into RDSR and influence cross-country comparisons, particularly in settings with limited health-data infrastructure.

From a public-health perspective, Pacific Island countries and territories may face a sustained burden of long-term disability and dialysis care; therefore, prevention strategies should address functional decline and long-term management in addition to mortality reduction. In these settings, a dual-focus strategy may be appropriate: first, strengthening early diabetes screening, glycaemic control and access to community-based education and care; and second, addressing CKD-nDM-related disability through earlier detection, nutritional and blood-pressure management, rehabilitation and primary-care capacity to reduce disability and improve quality of life. Elevated BMI also represents an important upstream intervention point. Prior studies show that lowering BMI can attenuate T2DM onset and improve kidney-function parameters[29, 30]. A comprehensive approach that combines metabolic-risk reduction with kidney-disability prevention may help mitigate the growing CKD-T2DM burden in these high-risk settings.

Based on these findings, countries and territories with Cluster 4-like profiles may benefit from three integrated actions: (1) establishing diabetes-CKD care pathways in primary care, supported by shared records and coordinated referral and follow-up; (2) embedding CKD screening into routine diabetes care while jointly managing hypertension, dyslipidaemia and obesity; and (3) improving affordability and access to renoprotective and metabolically beneficial therapies through mechanisms such as public procurement and generics, which may improve cost-effectiveness in low- and middle-income countries[31]. Strengthening integrated care to expand screening, coordination and treatment access may help reduce fragmentation and slow the accumulation of CKD-T2DM-related DALYs in high-burden settings.

Nonetheless, by combining GBD data with interpretable machine-learning models, we provide a model-attributed characterisation of evolving CKD-T2DM burden structure. Future work should incorporate patient-level clinical data, health-system indicators and implementation outcomes to validate these ecological patterns and identify actionable strategies for region-specific integrated care. Overall, locally tailored frameworks supported by service coordination and medication accessibility may help mitigate the global rise in CKD-T2DM burden.

Conclusion

In conclusion, CKD-T2DM burden has increased globally since 1990, accompanied by a rising relative model-attributed contribution from T2DM-related features. We observed marked spatial heterogeneity, with higher-SDI settings showing lower absolute burden but higher relative T2DM-related attribution, and several lower-SDI settings showing higher CKD-T2DM burden with positive statistical interaction patterns between T2DM-related and CKD-nDM-related features. These findings describe country-level burden structures rather than causal mechanisms, but they support the need for locally tailored diabetes-kidney surveillance, earlier CKD screening among people with T2DM, integrated chronic-disease management, improved access to renoprotective therapies and upstream metabolic-risk reduction to help limit future CKD-T2DM health loss.

Declarations

Ethics approval and consent to participate

This study used publicly available, aggregated and de-identified secondary data from the Global Burden of Disease study. No individual-level identifiable data were accessed. Therefore, additional ethical approval and informed consent were not required for this analysis.

Consent for publication

Not applicable.

Competing interests

The authors declare no financial or non-financial competing interests.

AI use statement

AI-assisted language editing was used to improve clarity, structure and scientific presentation during manuscript revision. No AI tools were used to generate data, conduct statistical analyses, produce figures or derive results. All scientific content, analyses, interpretations and final wording were reviewed and approved by the authors.

Funding

This work was supported by the National Natural Science Foundation of China (32470699) and the Research Startup Funds of Chongqing Medical University.

Author Contribution

J.P. conceived and planned the study and supervised the analyses. Z.G. conducted the analyses and wrote the first draft. Q.A.N. reviewed and edited the manuscript. Y.G., Z.Z. and F.Y. critically reviewed the integrity and plausibility of the data analysis. J.P. revised the manuscript and was responsible for the integrity of data acquisition and statistical analyses. Y.Y. provided administrative, technical and material support, supervision and mentorship. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Data Availability

The data analysed in this study were obtained from the Global Burden of Disease 2021 Results Tool available through the Global Health Data Exchange. These data are publicly available and can be accessed without restriction. The processed supplementary data supporting the findings of this study are provided in the supplementary files.

References

1. Kovesdy CP. Epidemiology of chronic kidney disease: an update 2022. *Kidney Int Suppl* (2011). 2022;12(1):7–11.
2. Li J, Pandian V, Davidson PM, Song Y, Chen N, Fong DYT. Burden and attributable risk factors of non-communicable diseases and subtypes in 204 countries and territories, 1990–2021: a systematic analysis for the global burden of disease study 2021. *Int J Surg*. 2025;111(3):2385–97.
3. Li H, Lu W, Wang A, Jiang H, Lyu J. Changing epidemiology of chronic kidney disease as a result of type 2 diabetes mellitus from 1990 to 2017: Estimates from Global Burden of Disease 2017. *J Diabetes Investig*. 2021;12(3):346–56.
4. Collaboration NCDRF. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet*. 2016;387(10027):1513–30.
5. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2014;384(9945):766–81.
6. Arabi T, Shafqat A, Sabbah BN, Ashraf N, Shah H, Abdulkader H, et al. Obesity-related kidney disease: Beyond hypertension and insulin-resistance. *Front Endocrinol (Lausanne)*. 2022;13:1095211.
7. Wang M, Li J, Li Y, Yao S, Zhao M, Wang C, et al. The effects of hypertension and diabetes on new-onset chronic kidney disease: A prospective cohort study. *J Clin Hypertens (Greenwich)*. 2020;22(1):39–46.
8. Zhuo L, Zou G, Li W, Lu J, Ren W. Prevalence of diabetic nephropathy complicating non-diabetic renal disease among Chinese patients with type 2 diabetes mellitus. *Eur J Med Res*. 2013;18(1):4.
9. Ameer OZ. Hypertension in chronic kidney disease: What lies behind the scene. *Front Pharmacol*. 2022;13:949260.
10. Kim K, Crook J, Lu CC, Nyman H, Sarker J, Nelson R, et al. Healthcare Costs Across Diabetic Kidney Disease Stages: A Veterans Affairs Study. *Kidney Med*. 2024;6(9):100873.

11. Xie D, Ma T, Cui H, Li J, Zhang A, Sheng Z, et al. Global burden and influencing factors of chronic kidney disease due to type 2 diabetes in adults aged 20–59 years, 1990–2019. *Sci Rep.* 2023;13(1):20234.
12. Deng Y, Li N, Wu Y, Wang M, Yang S, Zheng Y, et al. Global, Regional, and National Burden of Diabetes-Related Chronic Kidney Disease From 1990 to 2019. *Front Endocrinol (Lausanne).* 2021;12:672350.
13. Diseases GBD, Injuries C. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet.* 2024;403(10440):2133–61.
14. Collaborators GBDD. Global age-sex-specific mortality, life expectancy, and population estimates in 204 countries and territories and 811 subnational locations, 1950–2021, and the impact of the COVID-19 pandemic: a comprehensive demographic analysis for the Global Burden of Disease Study 2021. *Lancet.* 2024;403(10440):1989–2056.
15. An X, Liu Z, Zhang L, Zhao J, Gu Q, Han W, et al. Co-occurrence patterns and related risk factors of ischaemic heart disease and ischaemic stroke across 203 countries and territories: a spatial correspondence and systematic analysis. *Lancet Glob Health.* 2025;13(5):e808–19.
16. Alicic RZ, Rooney MT, Tuttle KR. Diabetic Kidney Disease: Challenges, Progress, and Possibilities. *Clin J Am Soc Nephrol.* 2017;12(12):2032–45.
17. Mortality GBD, Causes of Death C. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet.* 2016;388(10053):1459–544.
18. Magliano DJ, Boyko EJ. *IDF Diabetes Atlas*. 10th ed. Brussels: International Diabetes Federation; 2021.
19. Collaborators GBDD. 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.* 2023;402(10397):203–34.
20. Chan JCN, Lim LL, Wareham NJ, Shaw JE, Orchard TJ, Zhang P, et al. The Lancet Commission on diabetes: using data to transform diabetes care and patient lives. *Lancet.* 2021;396(10267):2019–82.
21. Li Y, Zhang H, Jiang Y. Research on the changes and predictions of the burden of type 2 diabetes mellitus in Pacific Island countries from 1990 to 2019. *PLoS ONE.* 2023;18(12):e0293681.
22. Safiri S, Karamzad N, Kaufman JS, Bell AW, Nejadghaderi SA, Sullman MJM, et al. Prevalence, Deaths and Disability-Adjusted-Life-Years (DALYs) Due to Type 2 Diabetes and Its Attributable Risk Factors in 204 Countries and Territories, 1990–2019: Results From the Global Burden of Disease Study 2019. *Front Endocrinol (Lausanne).* 2022;13:838027.
23. Na'ai D, Raphael KL. CKD in Native Hawaiians and Pacific Islanders: Trouble in Paradise. *Clin J Am Soc Nephrol.* 2019;14(11):1661–3.
24. George C, Echouffo-Tcheugui JB, Jaar BG, Okpechi IG, Kengne AP. The need for screening, early diagnosis, and prediction of chronic kidney disease in people with diabetes in low- and middle-income countries-a review of the current literature. *BMC Med.* 2022;20(1):247.
25. Zimbudzi E, Lo C, Robinson T, Ranasinha S, Teede HJ, Usherwood T, et al. The impact of an integrated diabetes and kidney service on patients, primary and specialist health professionals in Australia: A qualitative study. *PLoS ONE.* 2019;14(7):e0219685.
26. Tonelli M, Nkunu V, Varghese C, Abu-Alfa AK, Alrukhaimi MN, Fox L et al. Framework for establishing integrated kidney care programs in low- and middle-income countries. *Kidney Int Suppl (2011).* 2020;10(1):e19–e23.
27. Mirzababaei A, Abaj F, Roumi Z, Khosroshahi RA, Aali Y, Clark CCT, et al. Consumption of red, white, and processed meat and odds of developing kidney damage and diabetic nephropathy (DN) in women: a case control study. *Sci Rep.* 2024;14(1):10344.
28. Farrell DR, Vassalotti JA. Screening, identifying, and treating chronic kidney disease: why, who, when, how, and what? *BMC Nephrol.* 2024;25(1):34.
29. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med.* 2002;346(6):393–403.
30. Look ARG. Effect of a long-term behavioural weight loss intervention on nephropathy in overweight or obese adults with type 2 diabetes: a secondary analysis of the Look AHEAD randomised clinical trial. *Lancet Diabetes Endocrinol.* 2014;2(10):801–9.
31. Global H. Population Project on Access to Care for Cardiometabolic D. Reduced insulin use and diabetes complications upon introduction of SGLT-2 inhibitors and GLP1-receptor agonists in low- and middle-income countries: A microsimulation. *PLoS Med.* 2025;22(4):e1004559.

Figures

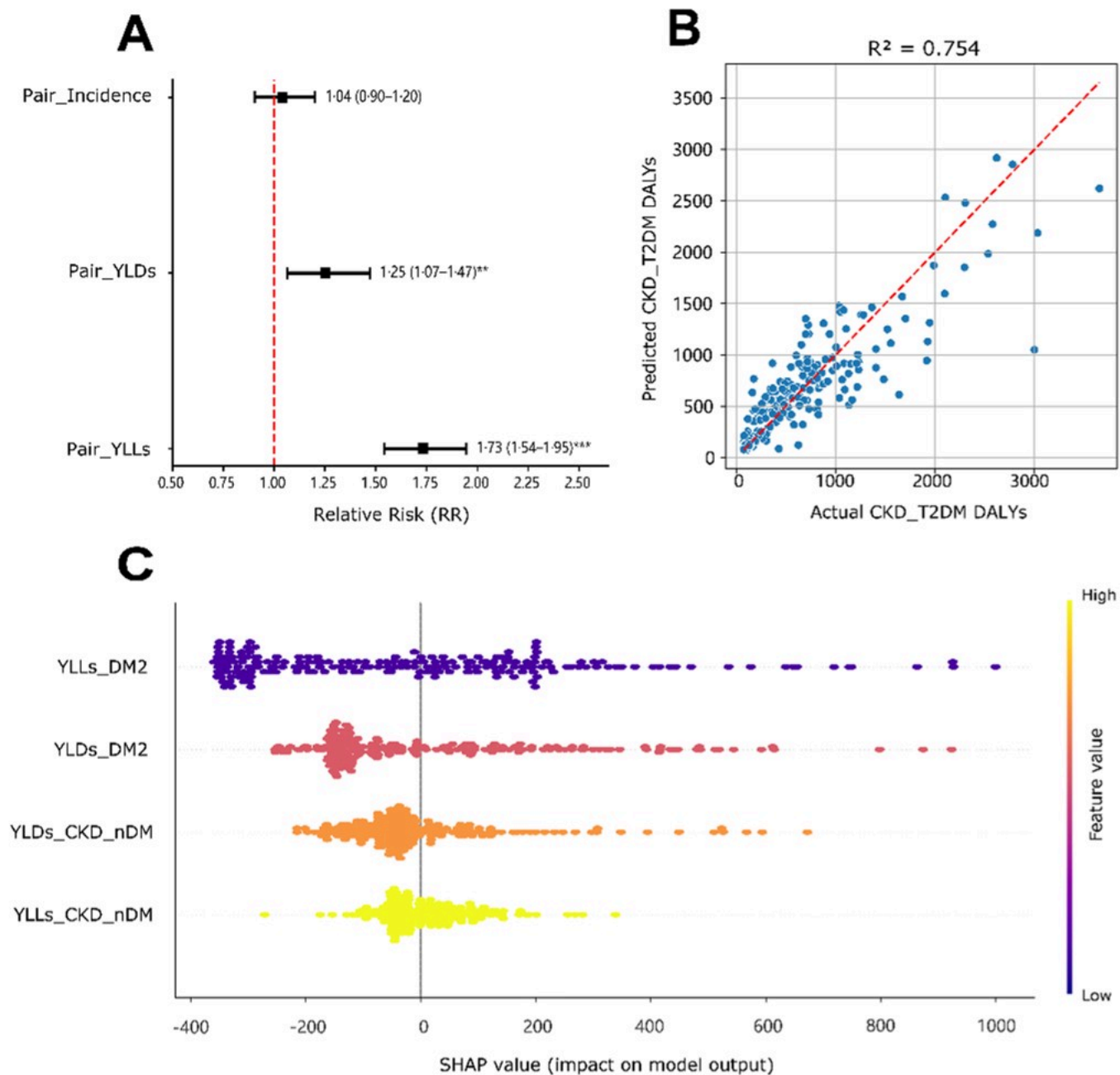


Figure 1

Feature screening, XGBoost model performance and SHAP-based attribution for CKD-T2DM DALYs. **A**, Relative risks (RRs) and 95% confidence intervals (CIs) from negative binomial regression evaluating retained burden domains in relation to age-standardised CKD-T2DM DALYs. Statistical significance is indicated as $p < 0.05$, $p < 0.01$ and $p < 0.001$. **B**, Observed versus predicted age-standardised CKD-T2DM DALY rates from the 2021 XGBoost model. **C**, SHAP summary plot showing the direction and magnitude of model-attributed contributions for the retained T2DM-related and CKD-nDM-related YLL and YLD features. Each point represents a country or territory, and colours indicate feature values from low to high. Abbreviations: CKD-T2DM, chronic kidney disease due to type 2 diabetes mellitus; CKD-nDM, non-diabetic chronic kidney disease; CI, confidence interval; DALYs, disability-adjusted life-years; RR, relative risk; SHAP, Shapley additive explanations; T2DM, type 2 diabetes mellitus; XGBoost, Extreme Gradient Boosting; YLDs, years lived with disability; YLLs, years of life lost.

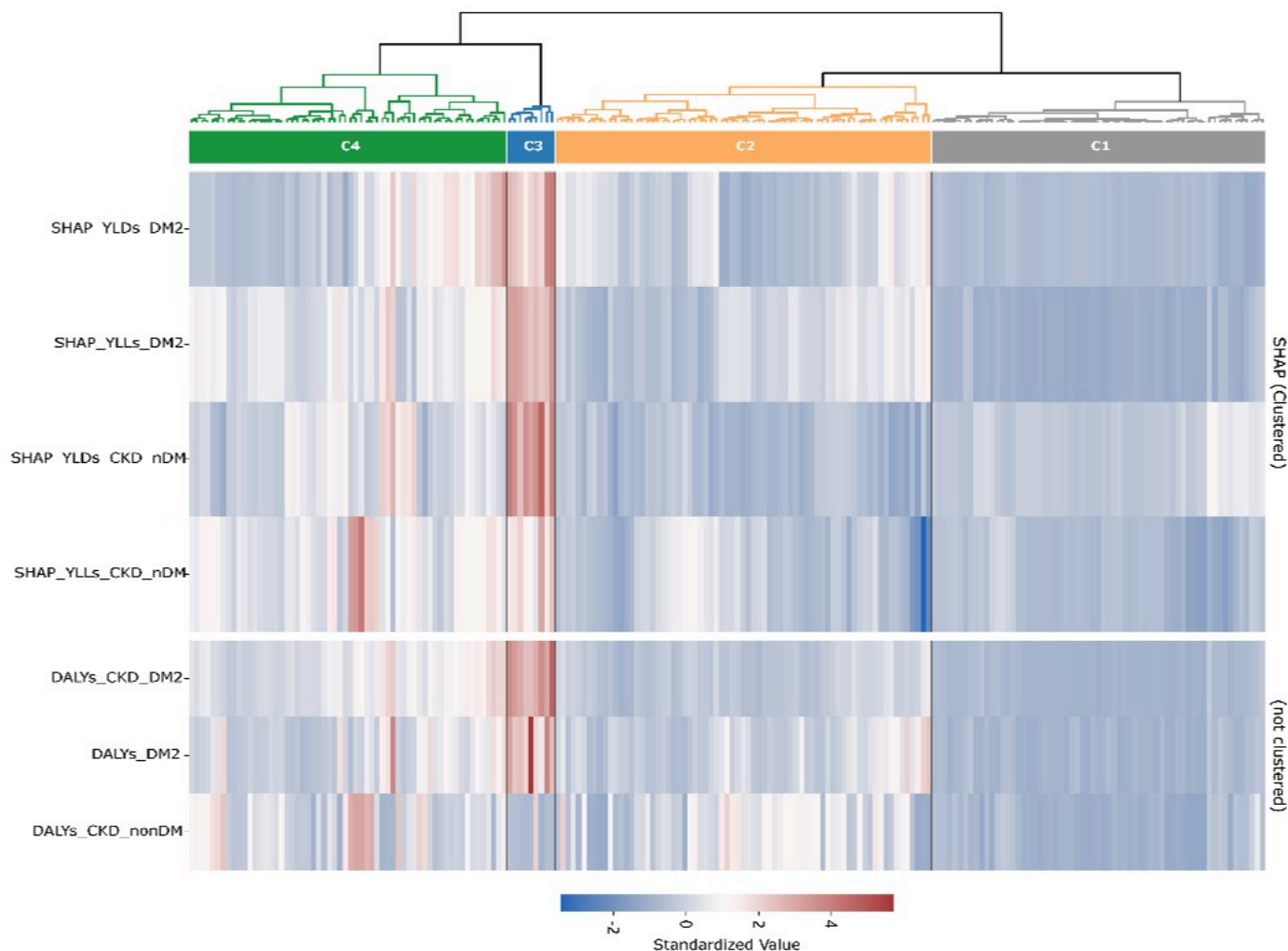


Figure 2

SHAP-derived hierarchical clustering of CKD-T2DM burden-structure profiles across countries and territories. Hierarchical clustering was performed using Ward's method based on SHAP values from the four retained model-attributed features: T2DM-related YLLs, T2DM-related YLDs, CKD-nDM-related YLLs and CKD-nDM-related YLDs. Age-standardised CKD-T2DM DALY rates, shown on the right, were not included in the clustering process and are displayed only for interpretation. C1-C4 denote Clusters 1-4. Abbreviations: CKD-T2DM, chronic kidney disease due to type 2 diabetes mellitus; CKD-nDM, non-diabetic chronic kidney disease; DALYs, disability-adjusted life-years; SHAP, Shapley additive explanations; T2DM, type 2 diabetes mellitus; YLDs, years lived with disability; YLLs, years of life lost.

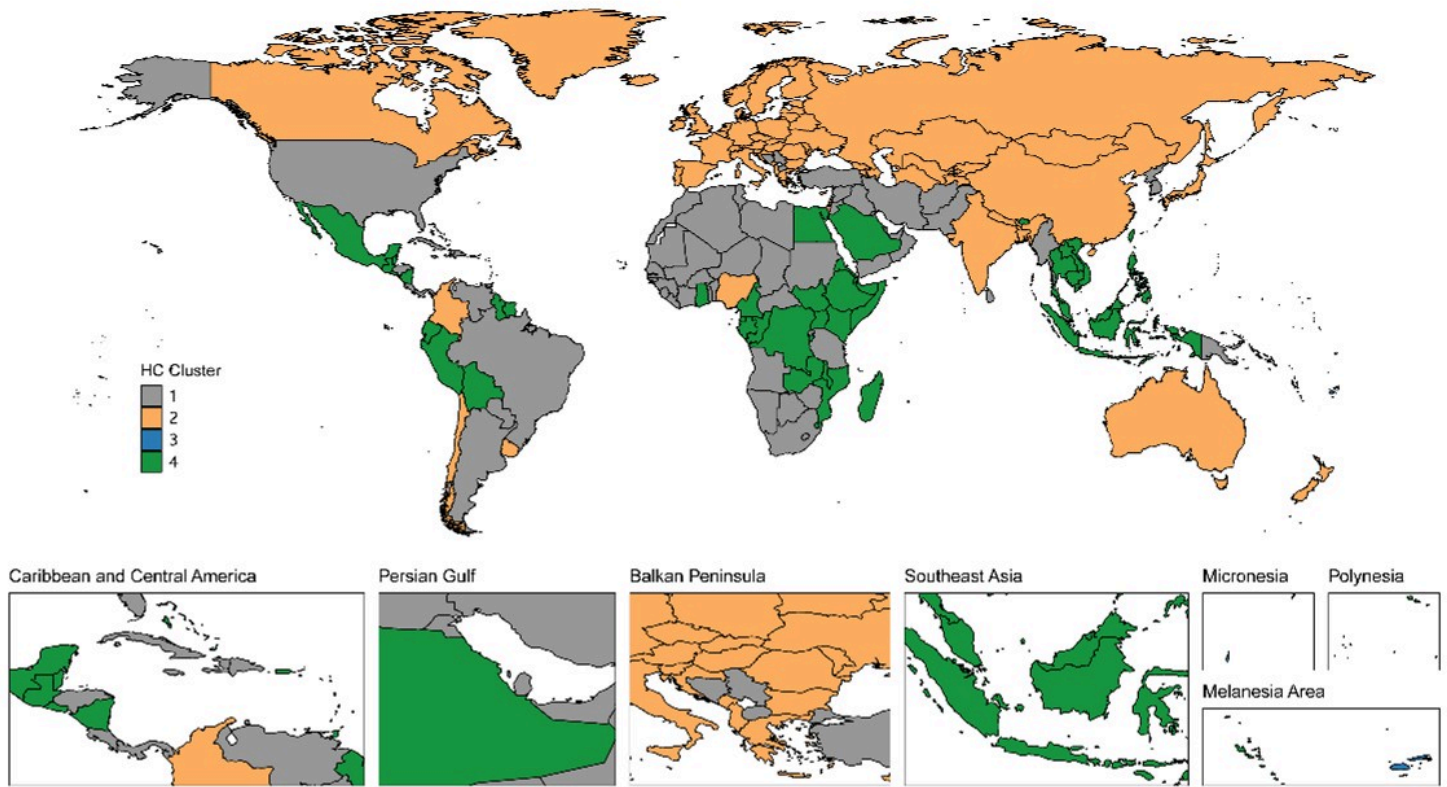


Figure 3

Global distribution of SHAP-derived CKD-T2DM burden-structure clusters. World map showing the geographic distribution of the four SHAP-derived burden-structure clusters across 203 countries and territories. Clusters were defined using SHAP values from the retained T2DM-related and CKD-nDM-related YLL and YLD features. Country and territory cluster assignments are provided in Supplementary Table 6. Abbreviations: CKD-T2DM, chronic kidney disease due to type 2 diabetes mellitus; CKD-nDM, non-diabetic chronic kidney disease; SHAP, Shapley additive explanations; T2DM, type 2 diabetes mellitus; YLDs, years lived with disability; YLLs, years of life lost.

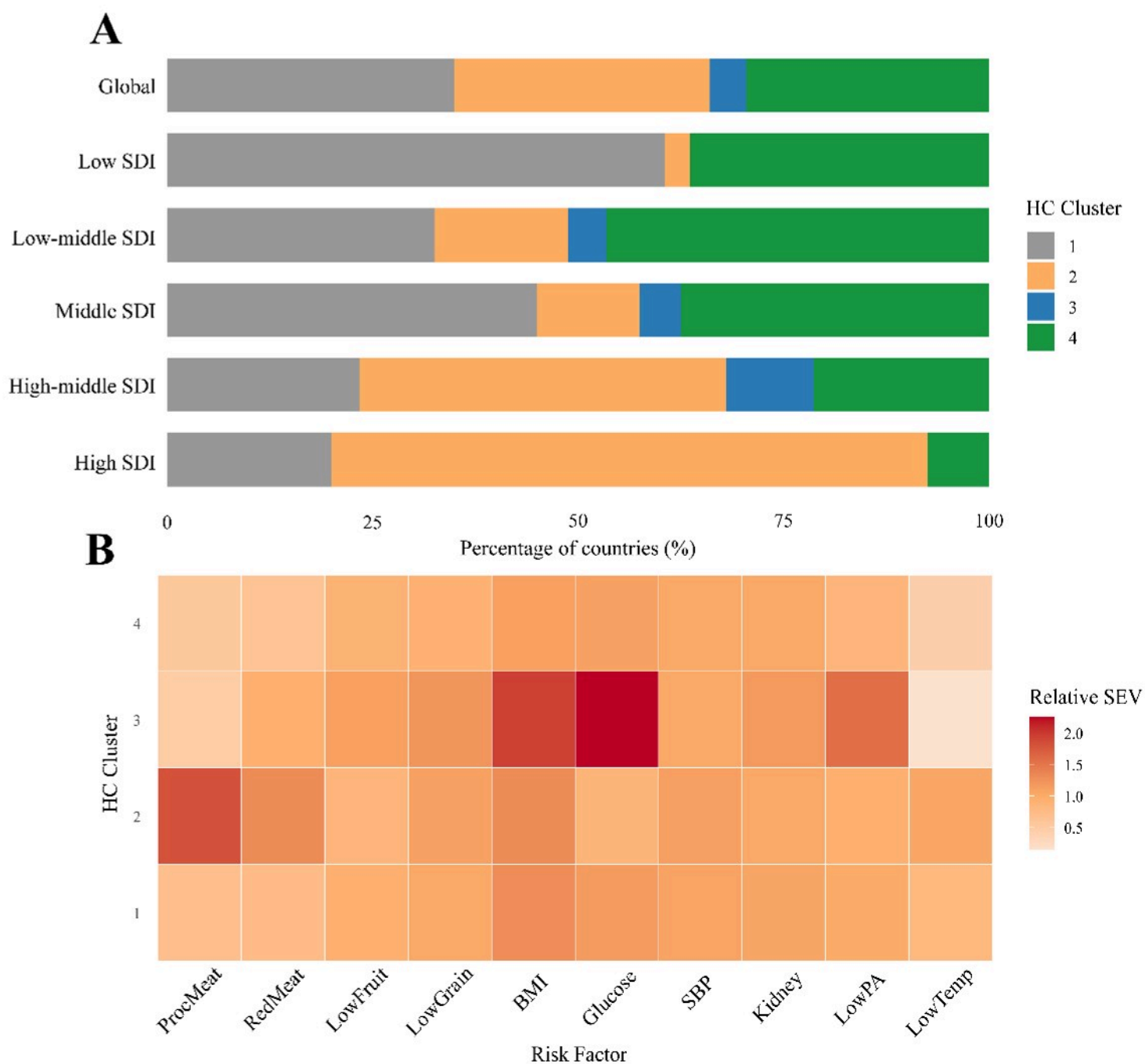


Figure 4

SDI composition and risk-factor exposure profiles across SHAP-derived CKD-T2DM burden-structure clusters. **A**, Stacked bars showing the distribution of the 203 countries and territories assigned to each SHAP-derived cluster across Socio-demographic Index (SDI) categories. The Global bar represents the overall distribution across all countries and territories, combining all SDI categories. **B**, Heatmap showing cluster-level relative summary exposure values (SEVs) for ten CKD-T2DM risk factors, normalised to the corresponding global average. Risk factors are labelled as follows: ProcMeat, diet high in processed meat; RedMeat, diet high in red meat; LowFruit, diet low in fruits; LowGrain, diet low in whole grains; BMI, high body-mass index; Glucose, high fasting plasma glucose; SBP, high systolic blood pressure; Kidney, kidney dysfunction; LowPA, low physical activity; LowTemp, low temperature. Abbreviations: CKD-T2DM, chronic kidney disease due to type 2 diabetes mellitus; SDI, Socio-demographic Index; SEV, summary exposure value; SHAP, Shapley additive explanations.

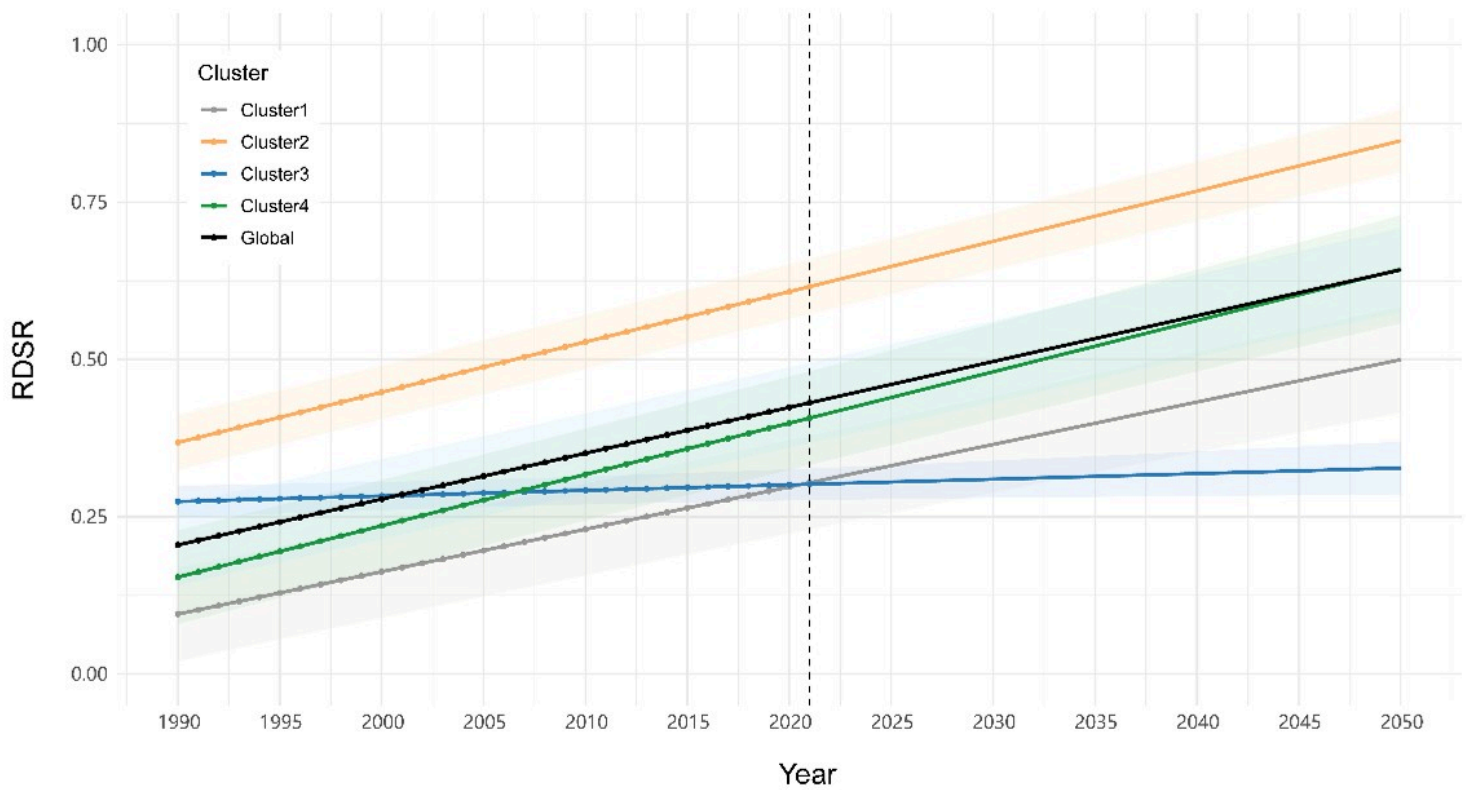


Figure 5

Bayesian INLA trajectories and projections of RDSR by SHAP-derived cluster and globally. Bayesian hierarchical models were fitted using integrated nested Laplace approximation (INLA), with calendar year modelled as the temporal component and country or territory included as a random intercept. The dashed vertical line at 2021 indicates the start of the projection period. The y-axis shows the relative driver structure ratio (RDSR), where higher values indicate a greater relative model-attributed contribution from T2DM-related features compared with CKD-nDM-related features. Shaded areas represent 95% posterior credible intervals. Abbreviations: CKD-T2DM, chronic kidney disease due to type 2 diabetes mellitus; CKD-nDM, non-diabetic chronic kidney disease; INLA, integrated nested Laplace approximation; RDSR, relative driver structure ratio; SHAP, Shapley additive explanations; T2DM, type 2 diabetes mellitus.

Supplementary Files

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

- [Supplementaryappendix.docx](#)