



# Spatial Patterns and Determinants of Diabetes Mellitus in Indian Adult Population: a Secondary Data Analysis from Nationally Representative Surveys

Yuvaraj Krishnamoorthy · Sathish Rajaa · Madhur Verma ·  
Rakesh Kakkar · Sanjay Kalra

Received: September 9, 2022 / Accepted: October 18, 2022 / Published online: November 3, 2022  
© The Author(s) 2022

## ABSTRACT

**Introduction:** Diabetes mellitus (DM) is a major public health challenge around the world. It is crucial to understand the geographic distribution of the disease in order to pinpoint high-priority locations and focus intervention on the

---

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s13300-022-01329-6>.

---

Y. Krishnamoorthy · S. Rajaa  
Department of Community Medicine, ESIC Medical College and PGIMS, K.K. Nagar, Chennai 600078, India

Y. Krishnamoorthy  
e-mail: yuvi.1130@gmail.com

S. Rajaa  
e-mail: psrajaa2410@gmail.com

M. Verma · R. Kakkar  
Department of Community and Family Medicine, All India Institute of Medical Sciences-Bathinda, Bathinda, Punjab, India

M. Verma  
e-mail: drmadhurverma@gmail.com

R. Kakkar  
e-mail: drrakesh75@rediffmail.com

S. Kalra (✉)  
Department of Endocrinology, Bharti Hospital, Karnal, Haryana, India  
e-mail: brideknl@gmail.com

target populations. Hence, this study was carried out to determine the spatial pattern and determinants of type-2 DM in an Indian population using National Family Health Survey-4 (NFHS-4) and Longitudinal Aging Survey in India (LASI).

**Methods:** We have adopted an ecological approach, wherein geospatial analysis was performed using aggregated district-level data from NFHS-4 (613 districts) and LASI survey datasets (632 districts). Moran's *I* statistic was determined and Local Indicators of Spatial Association (LISA) maps were created to understand the spatial clustering pattern of DM. Spatial regression models were run to determine the spatial factors associated with DM.

**Results:** Prevalence of self-reported DM among males (15–50 years) and females (15–49 years) was 2.1% [95% confidence interval (CI) 2.0–2.3%] and 1.7% (95% CI 1.6–1.8%), respectively. Prevalence of self-reported DM among males and females aged 45 years and above was 12.5% (95% CI 11.5–13.5%) and 10.9% (95% CI 9.8–12%). Positive spatial autocorrelation with significant Moran's *I* was found for both males and females in both NFHS-4 and LASI data. High-prevalence clustering (hot-spots) was maximum among the districts belonging to southern states such as Kerala, Tamil Nadu, Karnataka, and Andhra Pradesh. Northern and central states like Madhya Pradesh, Chhattisgarh, and Haryana mostly had

clustering of cold spots (i.e., lower prevalence clustered in the neighboring regions).

**Conclusion:** DM burden in India is spatially clustered. Southern states had the highest level of spatial clustering. Targeted interventions with intersectoral coordination are necessary across the geographically clustered hotspots of DM.

**Keywords:** Diabetes mellitus; Geographic information systems; India; Spatial analysis

### Key Summary Points

Prevalence of type-2 diabetes mellitus was higher among males compared with females across most states and districts of the country.

Burden of type-2 diabetes mellitus in India is spatially clustered among both males and females of all age groups.

High-prevalence clustering (hotspots) was maximum among the districts belonging to southern states such as Kerala, Tamil Nadu, Karnataka, and Andhra Pradesh.

Northern and central states like Madhya Pradesh, Chhattisgarh, and Haryana mostly had clustering of cold spots (i.e., lower prevalence clustered in the neighboring regions).

Targeted interventions with intersectoral coordination are necessary across the geographically clustered hotspots of DM.

Asian Region alone contributes to nearly 10% of this burden, with India being one of the major contributors to the total prevalence [2]. Estimates in 2019 have reported that India alone accounts for nearly 77 million patients with DM, with the numbers expected to rise to 134 million by 2045 [3]. Given its burden and projections, DM is a significant public health issue and remains a major health indicator under the Sustainable Development Goals (SDGs).

Despite being one of the most common conditions affecting the adult population in the country, the local patterns and the levels of the condition are not well established and studied in India. Understanding the geographic patterns of the disease in the country is necessary to identify the areas of high priority and direct healthcare interventions to address this problem in target groups. Only one previous study has tried to document the geospatial pattern and burden of DM at the district level using nationally representative data among the adult population in India [4]. However, the study was focused on only the southern part of the country and utilized data from 2012 to 2013 to document the findings. Hence, we tried to evaluate the spatial pattern and determinants of self-reported DM in India using two major nationally representative surveys, i.e., “National Family Health Survey-4 (NFHS-4)” conducted in 2015–2016 covering the 15–49-year age group and “Longitudinal Aging Survey in India (LASI)” conducted in 2017–2018 covering the 45-year-and-above age group. This study has adopted a macrolevel approach to identify the districts with higher rates of type-2 DM prevalence to facilitate intervention at the district level rather than focusing on the individual level.

## INTRODUCTION

Diabetes mellitus (DM) is a significant public health issue hindering social and economic advancement around the world. This places a heavy burden on healthcare and social welfare systems worldwide [1]. The global burden has doubled since the year 1980, increasing from 4.7% to 8.5% of the adult population, especially among developing countries [2]. The Southeast

## METHODS

### Data Sources

The “Indian Institute of Population Sciences (IIPS),” Mumbai, conducted the NFHS-4 survey among a representative sample of Indian households. It is a large-scale survey, collecting data from all the states and union territories in

India [5]. District-level data (for around 640 districts) are also available for almost all the indicators in the NFHS survey. This information will help the government in monitoring and evaluating various health-related programs and public health policies in the country.

LASI was a biennial panel survey of a nationally representative sample of India's middle-aged and elderly population. It was created with the intention of giving a thorough evaluation of the significant health outcomes for older adults in India. Between April 2017 and December 2018, the first wave of the survey was conducted in 30 states and six union territories (UTs) [6].

### Statement of Ethical Compliance

Since the study utilized publicly available datasets, ethical approval was not required. Appropriate permissions were obtained before utilizing the data for the study.

### Study Design and Study Participants

We have adopted an ecological approach, wherein geospatial analysis was performed using aggregated district-level data from NFHS-4 and LASI survey datasets. Women aged 15–49 years and males aged 15–54 years who participated in the NFHS-4 survey, as well as males and females aged 45 years and above who participated in the LASI survey, made up the study population, with districts serving as the unit of analysis.

### Sampling Strategy

#### *NFHS-4*

This survey was conducted using a two-stage sampling technique for selecting the primary sampling unit (PSU) in rural and urban regions. Within each rural and urban stratum, villages and census enumeration blocks (CEBs) were chosen using “probability proportional to size” (PPS) sampling. Complete household listing and mapping operations were carried out prior to the survey in each of the selected urban and rural PSUs. Selected PSUs were divided into

groups of 100–150 households each, totaling about 300 households. Two of these segments were randomly chosen for the survey using systematic sampling with a probability proportionate to segment size. As a result, clusters in NFHS-4 can be either a PSU or a portion of a PSU.

#### *LASI*

The final sampling unit of observation (adults aged  $\geq 45$  years and their spouses) was chosen using a multistage stratified cluster sampling procedure. The survey utilized a three-stage sampling strategy in rural areas and a four-stage sampling approach in urban areas. A total of 44,462 age-eligible households (households containing a member who is at least 45 years old) were chosen, and 42,949 of these households had interviews conducted. From these households, 82,650 people who were age eligible were found, and 72,250 of them underwent individual interviews [6].

### Data Collection Process

#### *NFHS-4*

Between 20 January 2015, and 4 December 2016, 789 field teams collected data throughout a two-phase period. Three female interviewers, one male interviewer, two health investigators, a driver, and one field supervisor made up each team. The sample size determined the number of interviewing teams in each state. The chosen field agencies in each state picked interviewers based on their credentials, including their training, work history, and other relevant expertise. The survey coordinators from each field agency decided which PSUs would be assigned to each team and made other logistical arrangements. If no eligible informant was available for the household interview or if an eligible woman or male in the family was not home when the interviewer visited, they had to call back at least three times.

The overall direction of the field teams fell within the purview of the field supervisor. The field supervisor also performed spot checks to ensure the accuracy of data, particularly in regard to the eligibility of respondents. IIPS

additionally selected one or more project officers or senior project officers in each state for monitoring and supervision of data collection methods and data quality. The following personnel also performed supervisory field visits during data collection process: project directors, senior staffs in field agencies, and faculty coordinators from IIPS. Details on the data collection process and data validation have all been thoroughly explained and published in a separate report elsewhere [5].

### **LASI**

State-level field agencies were subcontracted to carry out LASI fieldwork at the state level. A total of 130 field teams made up of six people carried out the primary field survey over the course of three phases. A field supervisor, two female investigators, two male investigators, and a health investigator were the team members. The number of interviewing teams in each state varied depending on the sample size of the state, and the interviewers were chosen by the state field agencies on the basis of their educational background, prior experience conducting extensive surveys, and other pertinent credentials. Each supervisor was in charge of gathering sample household lists, maps, and logistics for travel and lodging for each location in which their team was operating.

Supervisors were also in charge of procuring all supplies and equipment required for their teams to carry out the assigned interviews, communicating any field issues to the senior coordinators at IIPS, and contacting local authorities to inform them about the survey and secure their support and cooperation. Teams interacted with local politicians and community leaders before beginning fieldwork in order to raise awareness and improve response rates. They also distributed printed informational brochures, including press releases in local newspapers. A three-tiered supervision and monitoring structure was created to reduce non-sampling error and guarantee data quality because of the survey's scale and complexity.

### **Data Variables**

Independent variables in the NFHS-4 spatial analysis included poverty (households in poorest/poorer wealth index quintile), proportion of men and women with no formal education, proportion of obese men and women (body mass index  $\geq 25$  kg/m<sup>2</sup>), proportion of tobacco users, and proportion of alcohol users. All these variables except obesity (data not available) were included as independent variables in the LASI spatial analysis. The dependent variable was the prevalence of self-reported type-2 DM.

### **Statistical Analysis**

STATA 14.2 was used for the analysis (Stata-Corp, College Station, TX, USA). To account for the varied probability of selection and participation, sampling weights were used in the study. After taking the stratification and clustering into consideration in the sample design, the NFHS dataset was declared to be a survey dataset using the "svyset" command. Prevalence estimates were presented with a 95% confidence interval (CI).

We used the cluster-level geographical information system (GIS) data of NFHS-4 received through the demographic health survey (DHS) program to conduct a geospatial analysis. GIS coordinates were available for 674 districts in India. However, the values for relevant indicators were available for 613 districts in the NFHS-4 survey and 632 districts in the LASI survey. The district-level prevalence of type-2 DM and proportions of independent variables such as obesity, tobacco use, alcohol use, poverty, and illiterate population (no formal education) were retrieved from the dataset. We merged these aggregate data from survey datasets into the geospatial dataset and performed the analysis in "GeoDa software version 1.14." Similar analysis and methodologies were used in previous studies [7, 8].

### **Global Spatial Autocorrelation**

Before starting the analysis, Queen's first-order contiguity matrix was used to produce spatial

weights. To determine if the spatial pattern of DM was clustered, spread, or distributed randomly, global autocorrelation was assessed using the global Moran's  $I$  statistic. This was further tested by checking the Moran's  $I$  value following randomization with nearly 999 permutations. If the pseudo  $p$  value generated by comparing the observed distribution with the reference distribution was less than 0.05, then spatial autocorrelation was confirmed. Direction of Moran's  $I$  value determines the direction of autocorrelation. A negative value indicates that data points were different from the neighboring clusters (labeled as spatial outliers), while positive values signify that data points were similar to their neighboring clusters (labeled as spatial clustering) [9].

### Local Spatial Autocorrelation

The location of clusters (i.e., hotspots/cold spots/spatial outliers) were identified by using "Local Indicators of Spatial Association (LISA)." Both univariate (correlation of DM with lag value of DM i.e., values at neighboring clusters) and bivariate local Moran's scatter plot (correlation of DM with independent variable at neighboring clusters) were generated. The extent and nature of the spatial clustering were also determined using LISA significance and cluster maps. Getis Ord and Local Geary statistics, as well as other spatial markers of local autocorrelation, were also produced [9].

### Spatial Regression

We executed the ordinary least square (OLS) regression, keeping DM as the dependent variable, and poverty, illiteracy, obesity (only for females in NFHS survey data), tobacco, and alcohol use as independent variables. OLS regression runs under the assumption that random errors are uncorrelated, distributed normally, and of constant variance. The following regression diagnostics were checked: multicollinearity condition numbers (check the correlation in the random errors with value  $> 10$ , indicating multicollinearity), Jarque–Bera test (to check for the normality of errors with

$p$  value less than 0.05, indicating that the errors are not distributed normally) and Breusch–Pagan test (to check whether the variance is constant with  $p$  value less than 0.05, indicating heteroskedasticity). If these basic assumptions for the OLS estimates are not satisfied for any model, spatial regression models are run. The two spatial regression models that were used were the spatial lag model (which states that the independent variables in a district and its neighboring district have an impact on the prevalence of DM in that district) and the spatial error model (which states that the error terms across various spatial clusters are correlated) [10, 11]. We summarized the results from each model's coefficients and evaluated each model's robustness using the  $R^2$  and the log-likelihood value [12].

## RESULTS

On the basis of the individual-level data of NFHS-4, the prevalence of self-reported type-2 DM among males and females was 2.1% (95% CI 2.0–2.3%) and 1.7% (95% CI 1.6–1.8%) respectively. Among the major states, Kerala had the highest prevalence of DM among both males and females followed by Tamil Nadu, Andhra Pradesh, and Karnataka (Table 1).

On the basis of individual-level data of LASI, prevalence of type-2 DM among males and females aged 45 years and above was 12.5% (95% CI 11.5–13.5%) and 10.9% (95% CI 9.8–12%). Here also, southern states such as Kerala, Andhra Pradesh, Tamil Nadu, and Karnataka had the highest prevalence of DM among both males and females (Table 1).

### Spatial Pattern and Determinants of DM among Males and Females in 15–50 Years Age Group

#### Global Spatial Autocorrelation

We found a significant positive spatial autocorrelation among males and females with Moran's  $I$  value of 0.284 and 0.554, respectively ( $p$  value  $< 0.001$ ) (Supplementary Material—Supplementary Figs. 1 and 2). We found that poverty, illiteracy, and tobacco use had a

**Table 1** State-wise prevalence of self-reported DM among males and females in NFHS and LASI survey

State/union territory	Weighted prevalence of self-reported DM (95%CI)			
	NFHS-4 (2015–2016)		LASI (2017–2018)	
	Males (15–50 years)	Females (15–49 years)	Males (≥ 45 years)	Females (≥ 45 years)
Andaman and Nicobar Islands	5.3 (3.0–9.0)	4.6 (3.4–5.9)	14.1 (10.8–18.2)	17.5 (13.6–22.4)
Andhra Pradesh	4.0 (3.1–5.3)	2.4 (2.1–2.8)	20.0 (17.6–22.7)	15.4 (13.6–17.5)
Arunachal Pradesh	1.1 (0.7–1.8)	1.2 (1.0–1.4)	8.7 (5.8–13.0)	4.3 (2.4–7.5)
Assam	1.4 (1.0–1.9)	1.0 (0.9–1.2)	8.7 (6.9–10.8)	5.9 (4.6–7.6)
Bihar	1.5 (1.2–1.9)	1.5 (1.3–1.9)	7.8 (6.5–9.5)	7.3 (5.7–9.4)
Chandigarh	1.5 (0.3–6.6)	2.3 (1.4–3.6)	20.7 (16.7–25.3)	20.3 (16.8–24.4)
Chhattisgarh	0.9 (0.6–1.4)	0.5 (0.4–0.7)	8.3 (6.5–10.4)	5.5 (4.3–7.2)
Dadra and Nagar Haveli	1.4 (0.3–6.2)	0.8 (0.3–2.1)	10.3 (7.6–13.9)	7.3 (5.2–10.2)
Daman and Diu	0.02 (0.002–0.2)	0.5 (0.2–1.1)	23.4 (18.8–28.7)	12.3 (9.5–15.9)
Goa	3.9 (2.7–5.7)	3.8 (3.0–4.7)	24.4 (20.5–28.8)	18.7 (15.9–21.9)
Gujarat	1.5 (1.2–2.0)	1.2 (1.0–1.4)	13.6 (11.1–16.6)	12.5 (10.4–14.9)
Haryana	0.6 (0.4–1.0)	1.1 (1.0–1.3)	8.8 (6.9–11.2)	6.9 (5.4–8.8)
Himachal Pradesh	1.4 (0.9–2.0)	1.6 (1.3–2.0)	11.4 (8.4–15.3)	12.2 (9.6–15.4)
Jammu and Kashmir	3.4 (2.5–4.4)	1.9 (1.7–2.2)	6.1 (4.1–9.0)	9.8 (7.8–12.4)
Jharkhand	2.3 (1.6–3.2)	0.6 (0.5–0.8)	11.3 (9.4–13.5)	6.7 (5.4–8.3)
Karnataka	2.7 (1.7–4.3)	2.6 (1.8–3.8)	18.5 (9.7–32.3)	21.9 (13.4–33.8)
Kerala	5.7 (4.5–7.2)	4.3 (3.9–4.8)	33.2 (28.9–37.7)	25.9 (22.5–29.6)
Lakshadweep	1.6 (0.5–5.2)	3.8 (2.7–5.2)	24.2 (18.9–30.3)	18.8 (15.3–22.7)
Madhya Pradesh	1.1 (0.9–1.4)	1.1 (1.0–1.3)	6.7 (4.8–9.1)	5.2 (3.9–6.9)
Maharashtra	1.6 (1.1–2.3)	1.4 (1.2–1.7)	14.4 (12.5–16.5)	10.8 (9.3–12.5)
Manipur	1.6 (1.1–2.3)	0.9 (0.8–1.1)	13.9 (10.4–18.2)	7.4 (5.4–10.1)
Meghalaya	5.8 (2.7–12.1)	1.1 (0.8–1.5)	2.3 (1.0–5.2)	4.0 (2.6–5.9)
Mizoram	1.0 (0.4–2.4)	1.2 (0.9–1.6)	6.8 (4.9–9.4)	6.5 (4.8–8.7)
Nagaland	1.3 (0.8–2.1)	0.8 (0.6–1.1)	5.3 (1.7–15.0)	7.3 (3.7–13.9)
New Delhi	1.7 (0.9–3.3)	2.2 (1.7–2.9)	20.5 (15.8–26.2)	15.0 (12.2–18.2)
Odisha	3.3 (2.6–4.0)	1.5 (1.3–1.7)	9.6 (8.0–11.5)	5.9 (4.8–7.2)
Puducherry	3.8 (1.9–7.4)	2.9 (1.9–4.5)	22.8 (18.8–27.3)	20.0 (17.0–23.4)
Punjab	1.6 (1.2–2.2)	1.7 (1.5–1.9)	13.1 (10.8–15.8)	16.2 (13.8–18.9)
Rajasthan	1.0 (0.8–1.4)	0.6 (0.5–0.8)	10.8 (8.6–13.4)	6.0 (4.7–7.7)

**Table 1** continued

State/union territory	Weighted prevalence of self-reported DM (95%CI)			
	NFHS-4 (2015–2016)		LASI (2017–2018)	
	Males (15–50 years)	Females (15–49 years)	Males (≥ 45 years)	Females (≥ 45 years)
Sikkim	1.5 (0.8–2.8)	1.6 (1.3–2.0)	NA	NA
Tamil Nadu	4.5 (3.4–6.1)	3.7 (3.3–4.1)	20.2 (17.6–23.1)	17.6 (15.6–19.8)
Telangana	2.9 (1.9–4.2)	2.2 (1.8–2.7)	14.4 (12.1–16.9)	10.3 (8.7–12.1)
Tripura	1.5 (0.8–2.8)	1.4 (1.0–1.9)	10.3 (7.6–13.9)	7.3 (5.5–9.6)
Uttar Pradesh	1.3 (1.1–1.6)	1.0 (0.9–1.1)	9.0 (7.6–10.5)	6.0 (4.9–7.2)
Uttarakhand	0.9 (0.5–1.5)	1.4 (1.1–1.6)	10.3 (7.7–13.6)	7.8 (6.1–10.0)
West Bengal	1.7 (1.1–2.4)	1.7 (1.5–2.0)	10.5 (8.7–12.5)	8.7 (7.4–10.2)

NA not available

significant and negative spatial autocorrelation with DM among both males and females, whereas alcohol use had significant positive spatial autocorrelation with DM among males and negative autocorrelation among females (Supplementary Table 1). These results suggest that there is significant local clustering in the distribution of prevalence of DM among both males and females in India.

#### Local Spatial Autocorrelation

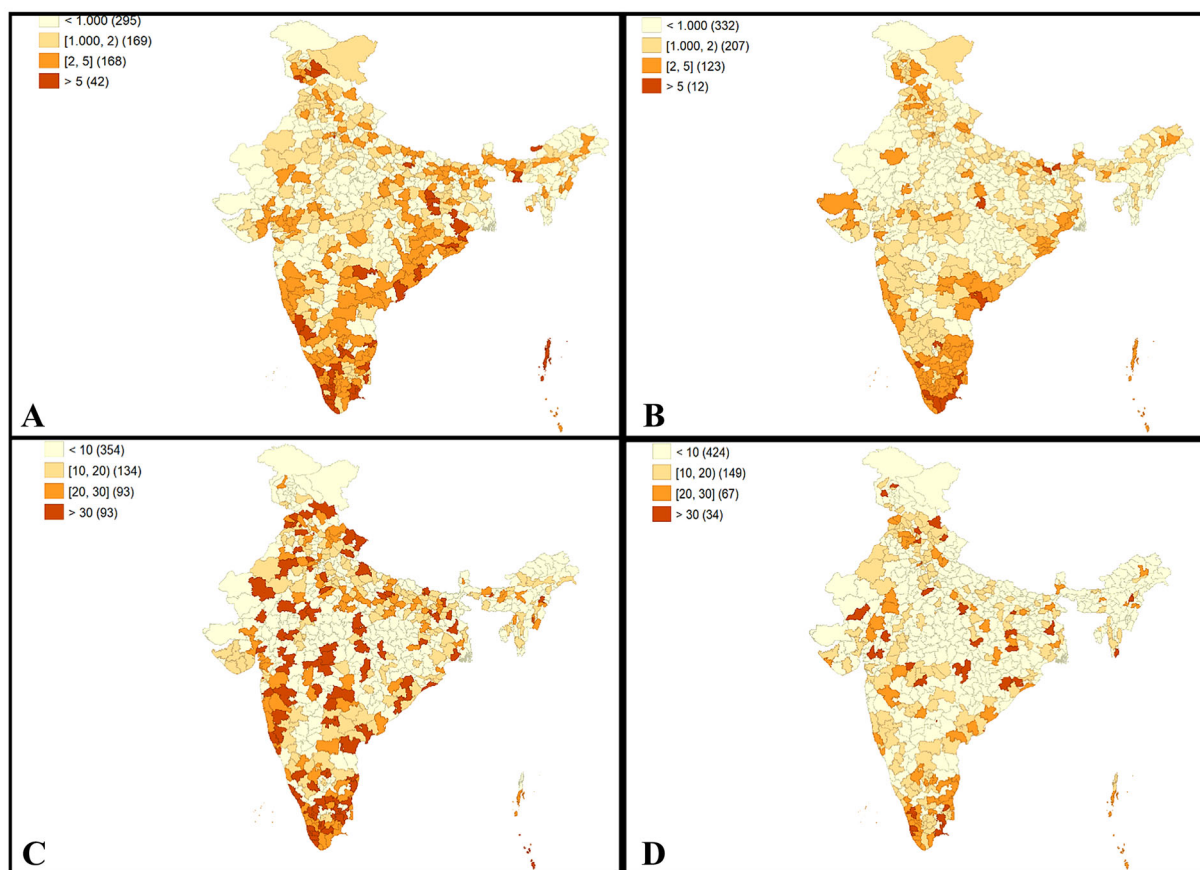
Figure 1A and B shows the district-wise prevalence of self-reported type-2 DM in India among males and females aged 15–50 years as per NFHS-4 survey. More than 6% of the districts had higher prevalence (> 5%) of DM among males, while only 2% of the districts had higher prevalence of DM among females. The majority of these high-prevalence districts were from the southern region of the country.

LISA significance map indicates that about 15% of the districts showed significant spatial clustering effect. LISA cluster map indicated that most of these significant hotspots (high-prevalence clustering) were present in southern states such as Kerala, Tamil Nadu, and Karnataka, while most of the significant cold spots (low-prevalence clustering) were present in

central states such as Madhya Pradesh (Figs. 2A and 3A).

LISA significance map showed that nearly 150 districts showed significant spatial clustering effect. Significant hotspots were more concentrated in southern states such as Kerala and Tamil Nadu, while the cold spots were concentrated in the central and eastern regions of the country (Figs. 2B and 3B). Local Geary and Local Getis-Ord statistic maps also showed similar number of hotspots and cold spots for the prevalence of DM among males and females (Supplementary Figs. 3–10).

Supplementary Figs. 11–28 show the bivariate LISA significance and cluster map for burden of DM. We found that nearly 10% of the districts were reported to have hotspots with higher burden of DM and higher burden of poverty, illiteracy, tobacco, and alcohol use among males aged 15–50 years. These districts were mostly from the eastern and central states such as Odisha, Jharkhand, Bihar, and Chhattisgarh. Nearly 20% of the districts were reported to have cold spots with all the covariates, and the majority were found in the northern states such as Rajasthan, Delhi, Haryana, etc. Among females, more than 20% of the districts were reported to have hotspots with high prevalence of DM and higher prevalence of obesity. Almost all these hotspots were found in



**Fig. 1** Map of India showing the district-wise prevalence of DM in India: **A** males aged 15–50 years; **B** females aged 15–49 years; **C** males aged  $\geq 45$  years; **D** females aged  $\geq 45$  years

southern states such as Tamil Nadu, Kerala, and Karnataka. The other covariates showed a high number of cold spots, with the majority found in the northern states.

### **Spatial Regression**

To start, an OLS regression model with spatial weights was used to determine whether DM and mesoscale correlates were related. However, after doing model diagnostics, it was found that both the male and female models' residuals in the OLS model displayed spatial dependency. To account for the autocorrelation of the residuals, we fitted these data to spatial autoregressive models such as spatial lag and spatial error model. With a rho coefficient of 0.44 (for males) and 0.47 (for females), the spatial lag model demonstrated the considerable influence of geographically lagged variables. Among males,

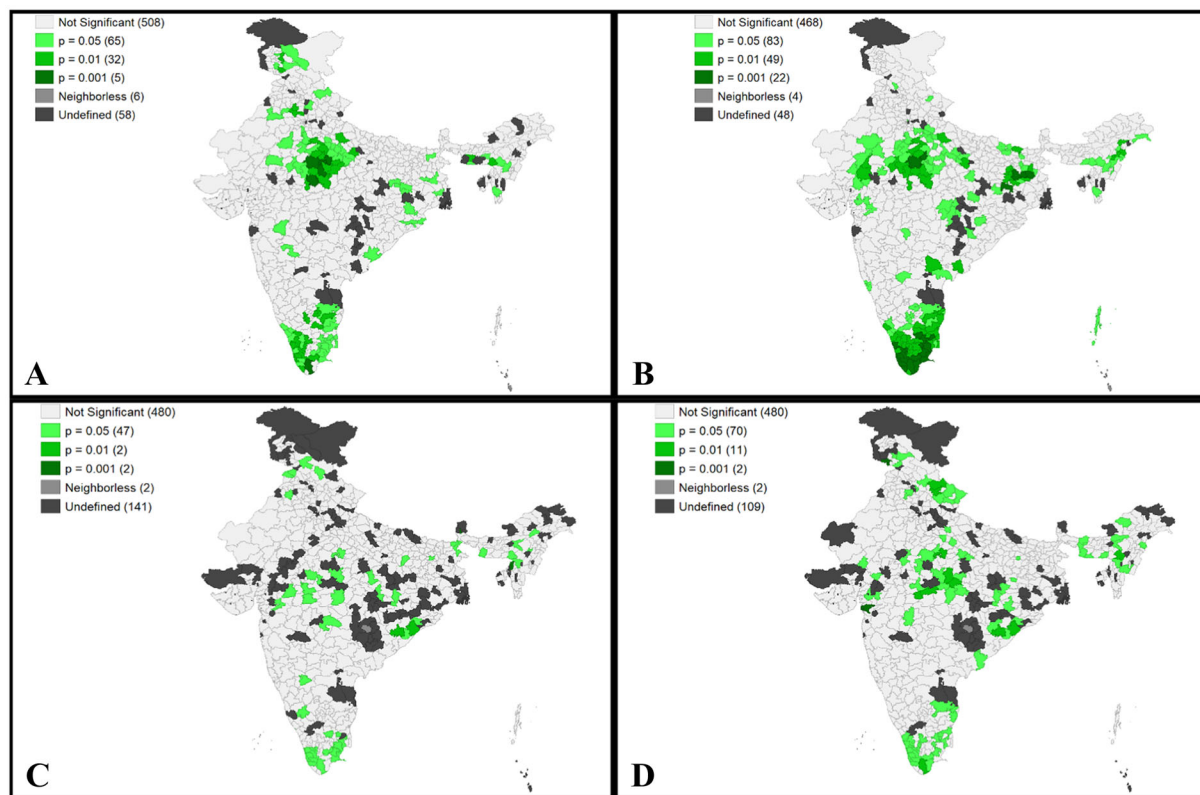
alcohol use was consistently shown as a significant spatial covariate for prevalence of DM across all three models, while obesity was shown as a significant spatial covariate among females in the OLS and spatial lag model (Supplementary Tables 2 and 3).

### **Spatial Pattern and Determinants of DM among Males and Females in $\geq 45$ Years Age Group**

#### **Global Spatial Autocorrelation**

We found a significant positive spatial autocorrelation among males and females with Moran's  $I$  value of 0.099 and 0.126, respectively ( $p$  value  $< 0.001$ ). Findings were similar to the NFHS global autocorrelation results as the covariates such as poverty, illiteracy, and tobacco use had a significant and negative





**Fig. 2** Univariate LISA significance map of India: **A** males aged 15–50 years; **B** females aged 15–49 years; **C** males aged  $\geq 45$  years; **D** females aged  $\geq 45$  years

spatial autocorrelation with DM among both males and females, whereas alcohol use had significant positive spatial autocorrelation with DM among males and negative autocorrelation among females (Supplementary Table 4). These results suggest that there is significant local clustering in the distribution of prevalence of DM among both males and females aged  $\geq 45$  years in India.

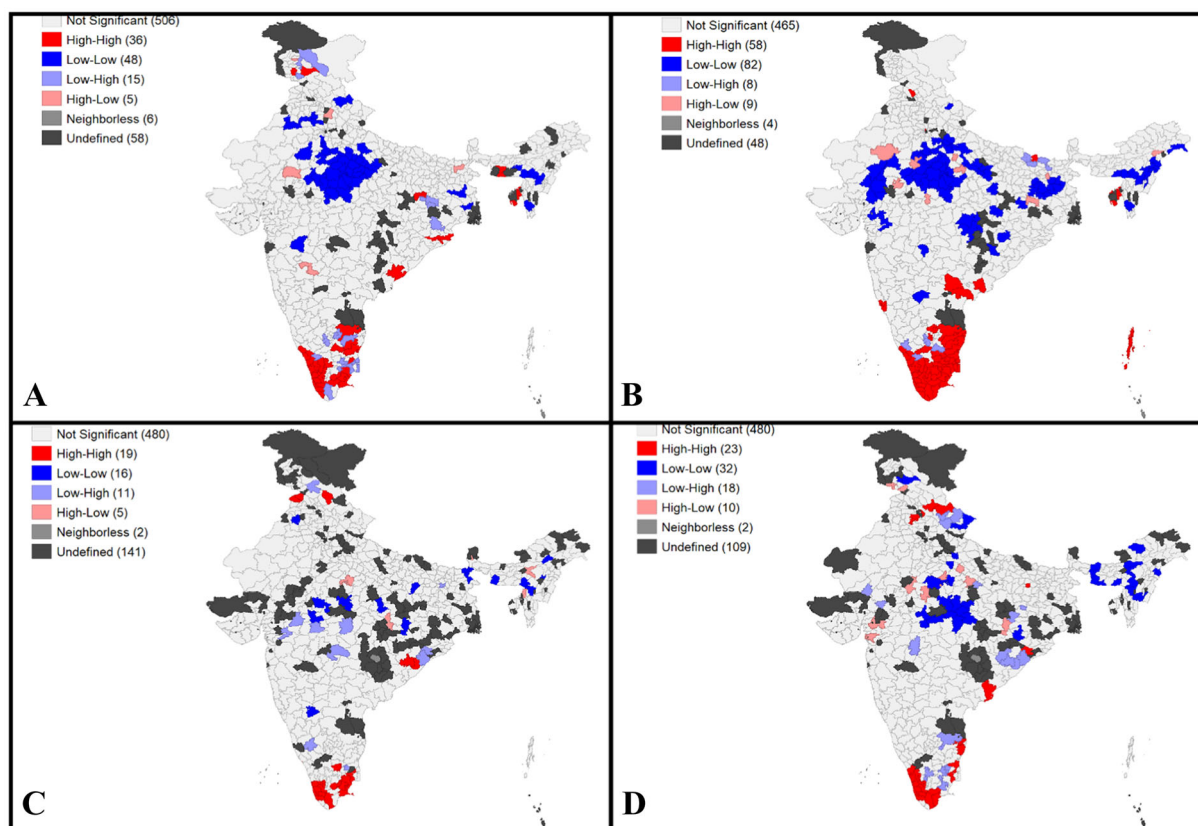
**Local Spatial Autocorrelation**

Figure 1C and D shows the district-wise prevalence of self-reported DM among males and females aged  $\geq 45$  years in India as per LASI survey. There was wide variation in the prevalence across the districts. About 93 districts had more than 30% prevalence of DM amongst males and these high burden districts were spread throughout the country. However, most districts were present in the southern and western regions of the country. Among females,

only 34 districts had very high prevalence of DM ( $> 30\%$ ), and these districts were spread across the country.

Figures 2C and 3C show the univariate LISA significance and cluster map for the burden of DM among males aged  $\geq 45$  years. LISA significance map indicates that only 51 districts showed significant spatial clustering effect. LISA cluster map indicated that most of these significant clusters were a mixture of hotspots and cold spots. Hotspots were present in southern states such as Kerala and Tamil Nadu, while most of the significant cold spots were present in central and eastern regions of the country.

Figures 2D and 3D show the univariate LISA significance and cluster map for the burden of DM among females. About 83 districts showed significant spatial clustering effect. High-prevalence clustering was found majorly in southern states, while low-prevalence clustering was found in central and eastern states. Local Geyr



**Fig. 3** Univariate LISA cluster map of India showing geographical clustering of hotspots and cold spots: **A** males aged 15–50 years; **B** females aged 15–49 years; **C** males aged  $\geq 45$  years; **D** females aged  $\geq 45$  years

and Local Getis-Ord statistic maps also showed similar number of hotspots and cold spots for the prevalence of DM among males and females aged  $\geq 45$  years (Supplementary Figs. 31–38). Supplementary Figs. 39–54 show the bivariate LISA significance and cluster map for prevalence of DM among males and females aged  $\geq 45$  years, with spatial lag of its covariates. Only 10–20 districts had high-prevalence clustering of DM and its covariates among males and females.

### **Spatial Regression**

Illiteracy, tobacco use, and alcohol use were found as significant spatial determinants of DM among males across all the models, while illiteracy and tobacco use were found as significant determinants across all three models among females (Supplementary Tables 5 and 6).

## **DISCUSSION**

DM is a significant public health problem in India with growing burden across all age groups. This shows that the national health programs and interventions are not effective enough to control the burden of DM. By targeting the districts with the highest burden and clustering of DM within India, it is helpful to position the DM prevention and control resources geographically and also address various causes of DM in the regions. Our study also addresses an important operational limitation for DM control on the Indian subcontinent by revealing the distribution of DM in both males and females across all age groups within the country.

Though the prevalence of DM was higher among males compared with females across most states and districts of the country, the

current study findings showed a significant spatial pattern among both males and females across all age groups. Moran's *I* statistic confirmed the presence of spatial dependence and geographical gradient of DM in India. Though high prevalence of DM was found in selected states and union territories, high-prevalence clustering (hotspots) was maximum among the districts belonging to southern states such as Kerala, Tamil Nadu, Karnataka, and Andhra Pradesh. Northern and central states such as Madhya Pradesh, Chhattisgarh, and Haryana mostly had clustering of cold spots (i.e., lower prevalence clustered in the neighboring regions). Similar results were found for males and females across all age groups. These findings were in line with previous studies showing significant spatial clustering and higher number of hotspots in the southern part of the country [13–16]. However, previous studies have focused only on particular regions or particular age groups (mostly reproductive age groups). Our study was able to confirm and reiterate the fact that the southern region remains the main target area for control of DM across all the age groups among both males and females.

Bivariate analysis using spatial lag covariates showed poverty, illiteracy, tobacco, alcohol use, and obesity (among females) showed high-prevalence clustering in eastern and central states such as Odisha, Jharkhand, Bihar, and Chhattisgarh among males and in southern states such as Kerala and Tamil Nadu among females. Apart from the clustering, spatial regression also found that illiteracy, alcohol use, and obesity were significantly associated with DM across all age groups.

First, among these factors, obesity had the highest association with DM prevalence among females and alcohol use had the highest association among males. Our findings are consistent with previous study findings, which establishes the impact of behavioral and anthropometric risk factors on DM prevalence [13, 15–17]. Prioritizing the regions with high-prevalence clustering of DM with alcohol use and obesity would be very helpful in reducing the overall burden of DM in India. We found an inverse association in terms of education as the districts having higher prevalence of illiteracy

are at significantly lower risk of having higher DM prevalence. Previous studies conducted in India and other countries [4, 16, 18] also reported that illiteracy was inversely associated with DM prevalence. Though we did not find significant association with poverty, the lack of pattern found in this study does not refute a relationship, but suggests that the socioeconomic patterning of DM (considered by many a “disease of affluence”) in developing countries is in transition as the urban environment of India.

Overall, we were able to explain 20–30% of the variance in DM burden among population aged 15–50 years in India, while only 10% of the variance in DM burden among population aged  $\geq 45$  years were explained by the covariates added in the model. The rest of the unexplained variance may be due to either the factors not accounted in the model or covariates not related to location. Dietary consumption pattern, sociocultural practices, and lifestyle patterns could account for the remaining unexplained variability. Hence, our study findings reinforce the need to have further large-scale geospatial surveys to find out how various factors affect DM prevalence at smaller spatial scales.

Our study has the following set of limitations. To identify the parameters linked to DM, we first employed a number of sociodemographic, behavioral, and anthropometric covariates as a proxy. This approach may give a rough estimate of exposure to the outcome, which may lead to regression dilution bias and underestimation of the observed results [19]. Second, despite the biological validity of the associations found in our investigation, it is impossible to estimate their size or scope since the use of aggregated data adds ecological fallacy into the analysis [20].

Despite these issues, our study findings have important implications. We identified the geographical variation in the DM burden on the basis of age group and gender. This knowledge will be helpful in allocating the resources for specific districts on the basis of the differential spatial findings based on age group and gender. We found how demographic and social factors such as poverty, literacy, and behavioral habits influence the burden of DM prevalence across

the region. The clustering of DM with spatial lag correlates at the district level was used to map and identify the hotspots and cold spots. This will assist in reviewing the justification for the intervention plans created at the national and regional levels for DM. The Government of India will be able to take an integrated approach with multisectoral coordinated efforts to lessen the total burden of DM throughout the nation by using these findings.

## CONCLUSION

DM burden in India is spatially clustered. Southern states had the highest level of spatial clustering. Targeted interventions with intersectoral coordination are necessary across the geographically clustered hotspots of DM.

## ACKNOWLEDGEMENTS

**Funding.** No funding was received for the publication of this study.

**Authorship.** All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

**Author Contributions.** Yuvaraj Krishnamoorthy—Conception of the work, data analysis and interpretation, drafting the article, critical revision of the article, final approval of the version to be submitted; Sathish Rajaa S—Data analysis and interpretation, drafting the article, critical revision of the article, final approval of the version to be submitted; Madhur Verma – Design of the work, Data analysis and interpretation, critical revision of the article, final approval of the version to be submitted; Rakesh Kakkar – Design of the work, data analysis and interpretation, critical revision of article, final approval of the version to be submitted; Sanjay Kalra—Design of the work, data analysis and interpretation, critical revision of

article, final approval of the version to be submitted;

**Disclosures.** Yuvaraj Krishnamoorthy, Sathish Rajaa, Madhur Verma, Rakesh Kakkar and Sanjay Kalra have nothing to disclose.

**Compliance with Ethics Guidelines.** Since the study utilized publicly available datasets, ethical approval was not required. Appropriate permissions were obtained before utilizing the data for the study.

**Data Availability.** The datasets generated during and/or analyzed during the current study are available in the Demographic Health Survey (DHS) repository and International Institute of Population Sciences repository, [WEB LINK TO DATASETS: <https://dhsprogram.com/data/available-datasets.cfm>; [https://iipsindia.ac.in/sites/default/files/LASI\\_DataRequestForm\\_0.pdf](https://iipsindia.ac.in/sites/default/files/LASI_DataRequestForm_0.pdf)].

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

## REFERENCES

1. Al-Lawati JA. Diabetes mellitus: a local and global public health emergency! *Oman Med J.* 2017;32(3):

- 177–9. <https://doi.org/10.5001/omj.2017.34.PMID:28584596;PMCID:PMC5447787>.
2. Cho NH, et al. IDF Diabetes Atlas: global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res Clin Pract.* 2018;138:271–81.
  3. Pradeepa R, Mohan V. Epidemiology of type 2 diabetes in India. *Indian J Ophthalmol.* 2021;69(11):2932.
  4. Barua S, Saikia N, Sk R. Spatial pattern and determinants of diagnosed diabetes in southern India: evidence from a 2012–13 population-based survey. *J Biosoc Sci.* 2021;53(4):623–38. <https://doi.org/10.1017/S0021932020000449>.
  5. International Institute for Population Sciences (IIPS). (2017) National Family Health Survey- 4 [Internet]. <http://rchiips.org/NFHS/nfhs4.shtml>. Accessed Feb 2020
  6. Arokiasamy P, Bloom D, Lee J, et al. Longitudinal Aging Study in India: Vision, Design, Implementation, and Preliminary Findings. In: National Research Council (US) Panel on Policy Research and Data Needs to Meet the Challenge of Aging in Asia; Smith JP, Majmundar M, editors. *Aging in Asia: Findings From New and Emerging Data Initiatives*. Washington (DC): National Academies Press (US); 2012. 3. <https://www.ncbi.nlm.nih.gov/books/NBK109220/>
  7. Krishnamoorthy Y, Ganesh K. Spatial pattern and determinants of tobacco use among females in India: evidence from a nationally representative survey. *Nicotine Tob Res.* 2020;22(12):2231–7. <https://doi.org/10.1093/ntr/ntaa137>.
  8. Krishnamoorthy Y, Majella MG, Rajaa S, Bharathi A, Saya GK. Spatial pattern and determinants of HIV infection among adults aged 15 to 54 years in India—evidence from National Family Health Survey-4 (2015–16). *Trop Med Int Health.* 2021;26(5):546–56. <https://doi.org/10.1111/tmi.13551>.
  9. Anselin L, Syabri I, Kho Y. GeoDa: an introduction to spatial data analysis. *Geogr Anal.* 2006;38:5–22.
  10. Pace RK, LeSage JP. Omitted variable biases of OLS and spatial lag models. In: *Progress in spatial analysis*. Berlin, Heidelberg: Springer; 2010. p. 17–28.
  11. Anselin L, Kelejian HH. Testing for spatial error autocorrelation in the presence of endogenous regressors. *Int Reg Sci Rev.* 1997;20:153–82.
  12. Anselin L. Spatial regression. In: *The SAGE handbook of spatial analysis*, vol. 1. 2009. p. 255–76.
  13. Biradar RA, Singh DP. Spatial clustering of diabetes among reproductive age women and its spatial determinants at the district level in southern India. *Clin Epidemiol Global Health.* 2020;8(3):791–6.
  14. Hernandez AM, Jia P, Kim HY, Cuadros DF. Geographic variation and associated covariates of diabetes prevalence in India. *JAMA Netw Open.* 2020;3(5): e203865. <https://doi.org/10.1001/jamanetworkopen.2020.3865>.
  15. Singh S, Puri P, Subramanian SV. Identifying spatial variation in the burden of diabetes among women across 640 districts in India: a cross-sectional study. *J Diabetes Metab Disord.* 2020;19(1):523–33.
  16. Ghosh K, Dhillon P, Agrawal G. Prevalence and detecting spatial clustering of diabetes at the district level in India. *J Public Health.* 2020;28(5):535–45.
  17. Wu J, Wang Y, Xiao X, Shang X, He M, Zhang L. Spatial analysis of incidence of diagnosed type 2 diabetes mellitus and its association with obesity and physical inactivity. *Front Endocrinol.* 2021. <https://doi.org/10.3389/fendo.2021.755575>.
  18. Faka A, Chalkias C, Montano D, Georgousopoulou EN, Tripitsidis A, Koloverou E, Tousoulis D, Pitsavos C, Panagiotakos DB. Association of socio-environmental determinants with diabetes prevalence in the Athens Metropolitan Area, Greece: a spatial analysis. *Rev Diabet Stud.* 2018;14(4):381–9. <https://doi.org/10.1900/RDS.2017.14.381>.
  19. Hutcheon JA, Chiolerio A, Hanley JA. Random measurement error and regression dilution bias. *BMJ.* 2001;340: c2289.
  20. Portnov BA, Dubnov J, Barchana M. On ecological fallacy, assessment errors stemming from misguided variable selection, and the effect of aggregation on the outcome of epidemiological study. *J Expo Sci Environ Epidemiol.* 2007;17:106–21.