



OPEN A retrospective cluster analysis of regional disparities and healthcare factors influencing causes of death certification and mortality statistics in India

Khushwant Singh¹, Sanghamitra Pati² & Ashoo Grover¹✉

Reliable cause-specific mortality statistics are crucial for defining health priorities, public health programs, allocating resources, designing and implementing policies to improve healthcare quality and accessibility. India accounts for almost 18 percent of the world's population. The 2020 report from the Office of the Registrar General of India indicates that the Medical Certification of Cause of Death (MCCD) rate is only 22.5%, with a minimal improvement of just 2.5% over the past decade. This study is the first to provide a comprehensive evaluation of MCCD-patterns across India over the past 15 years, addressing a critical-gap in the literature by identifying regional patterns, disparities, and healthcare variables that have previously been underexplored. Based on MCCD-trends over this period, the states and Union-Territories of India can be categorized into three clusters. Cluster1 includes 23 states with the lowest-average MCCD-rate of 18%, attributed to a low 0.14 doctors per 1000 people, with only 27.4% of hospitals actively reporting-MCCD. In contrast, Clusters2 and 3 have higher-average MCCD-rates of 63% and 60%, respectively, supported by higher 0.27 and 0.33 doctors per 1000 people, with over 80% of hospitals actively reporting-MCCD. Although, the findings indicate that active MCCD-reporting is a major factor associated with MCCD rates, other factors including healthcare infrastructure, state-specific healthcare policies, socioeconomic factors, and administrative management also influence MCCD-rates.

Keywords Life expectancy, Chronic disease prevalence, Death registration system, Healthcare infrastructure, Healthcare accessibility, Elbow method in model selection, Health resource allocation, Cause of Death, India health statistics, Healthcare policy and planning

Abbreviations

CI	Confidence interval
COD	Coefficient of determination
CRS	Civil registration system
ICMR	Indian Council of Medical Research
LCL	Lower confidence limit
MCCD	Medically Certified Causes of Death
MoHFW	Ministry of Health and Family Welfare
NCT	National Capital Territory
ORGI	Office of the Registrar General & Census Commissioner, India
RMSE	Root mean squared error
SD	Standard deviations
SE	Standard error
SRS	Sample Registration System
UCL	Upper confidence limit

¹Division of Delivery Research, Indian Council of Medical Research (ICMR) Headquarters, Ministry of Health & Family Welfare (MoHFW), Govt. of India, Ansari Nagar, New Delhi 110029, India. ²Indian Council of Medical Research (ICMR), Headquarters, New Delhi 110029, India. ✉email: grovera.hq@icmr.gov.in

UTs	Union Territories
WCSS	Within-cluster sum of squares
WHO	World Health Organization

Healthcare advancements have significantly increased global life expectancy, with India witnessing a notable rise of 9%, from 62.1 years in 2000 to 67.7 years in 2024¹. As life expectancy grows, the effective allocation of healthcare resources becomes increasingly critical^{2,3}. Accurate data on the leading causes of death at both national and state levels is crucial for guiding public health policies and resource allocation⁴. In this context, a robust registry of deaths and Medically Certified Causes of Death (MCCD) is vital. India faces unique challenges in death registration and MCCD-reporting, exacerbated due to a fragmented healthcare system resulting in disparities in healthcare access, the number of doctors, hospitals, and MCCD-reporting across states and union territories (UTs). According to the 2022 Sample Registration System (SRS) report, 8,759,522 deaths were reported in 2020, with 80% having registered death certificates^{5–7}. However, the Civil Registration System (CRS) report revealed that only 22.5% of these deaths were medically certified⁸. This substantial gap between registered deaths and MCCD underscores a critical challenge in tracking health outcomes and impeding the development of effective public health interventions. To improve health outcomes, India must better identify the leading causes of death and address challenges in death registration and certification across regions. This study aims to: (1) analyze regional disparities in MCCD-patterns across India, (2) classify states based on MCCD-reporting rates, and (3) identify healthcare factors contributing to low MCCD-rates in underperforming states.

Results

Regional disparities and variations in medically certified deaths in India

As evident from supplementary Table 1, significant regional variations exist in MCCD-rates in different regions of India. The overall national-average MCCD-percentage has steadily increased from 2006–2010 (15.33%) to 2016–2020 (20.91%), indicating progress in death registration and certification practices across India (Fig. 1A). In UTs, Lakshadweep consistently exhibited the highest MCCD-percentage (94–95%), demonstrating near-complete certification, followed by Puducherry and Chandigarh, with almost 70% MCCD as shown in supplementary Table 1. Additionally, Delhi, the capital of India, demonstrated an overall improvement in MCCD from 57.46% to 59.71% over the past years. However, the North India region reported the lowest-average MCCD (13%) among all regions during 2015–2020. The overall trend change presents a mixed picture, with Punjab (from 10.55 to 16.54%) and Himachal Pradesh reporting consistent increases, while Haryana and Uttar Pradesh lag behind with incomplete data reporting between 2006 and 2015. In contrast, the South India region leads the country in overall improvement in MCCD reporting. Tamil Nadu shows a significant rise from 28.34 to 43.46%, making it one of the highest MCCD states in the southern region. Furthermore, Karnataka and Andhra Pradesh follow closely with consistently high levels of certification. Interestingly, Kerala, despite improvements, continues to have comparatively lower MCCD-rates compared to other southern states.

Similarly, the East India region exhibits varying trends, with Chhattisgarh demonstrating remarkable improvement, jumping from 7.35 to 19.99% in the past 15 years, followed by West Bengal, with a notable increase from 4.27 to 12.29% (Fig. 1B). However, Jharkhand remains one of the lowest-certified states in the region, with only 4.99% of deaths medically certified in the 2016–2020 period. On the other hand, with the average MCCD increasing from 33.95 to 35.75%, West India reveals significant disparities and trends. With consistently high MCCD-rates maintaining an exceptional 100% certification, Goa reflects near-perfect death certification practices, while Rajasthan and Madhya Pradesh reported relatively lower MCCD-percentages. Additionally, among all regions, the Northeast region presents a diverse scenario with an average MCCD of 37.61% in the past years (Fig. 1B), with Manipur standing out with an impressive 68.88% MCCD-rate, followed by Arunachal Pradesh and Meghalaya. Conversely, Nagaland and Sikkim demonstrate significantly lower MCCD-percentages.

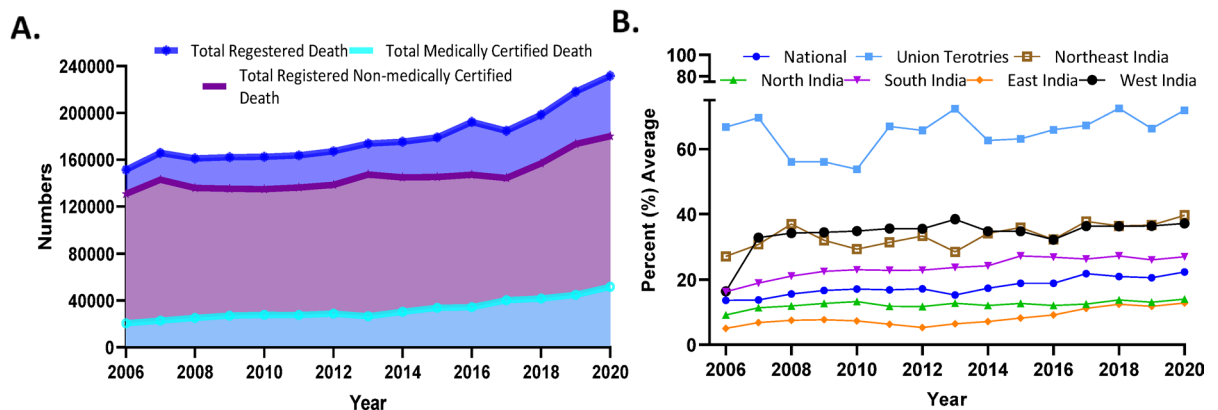


Fig. 1. (A) Trends in the total registered deaths, medically certified deaths, and registered non-medically certified deaths from 2006 to 2020, demonstrating the gap in medical certification of the cause of death and total registered deaths across India, (B) Trends in the national average MCCD percentage from 15% in 2006 to 22.5% in 2020, showing steady improvement and highlighting disparities across Union Territories and regions.

State-level disparities and clustering of states based on MCCD patterns

Overall, regional and UTs comprehensive analysis indicates that while certain regions, such as the South and West, have shown improvement in MCCD, other regions, especially parts of North and East India, lag behind in the improvement of MCCD. Although the data also highlights that within each region, certain states have comparably higher MCCD than others. Further, to group states with similar MCCD patterns, we performed a series of K-Means clustering analyses (2006–2020 time period) with different numbers of clusters (k = 2 to 5), and the total Within-Cluster Sum of Squares (WCSS), which reflects the compactness of the clusters was recorded for each. The recorded WCSS values were 170,255.98 for k=2; 2,401.62 for k=3; 81,305.94 for k=4; and 66,442.75 for k=5 (supplementary Table 2). To select optimum number of clusters, we applied the Elbow Method, and examined the reduction in WCSS as clusters (k) increases. As evident from supplementary Table 2, the WCSS sharply decreased from k=2 to k=3 (~47,854 units) and from k=3 to k=4 (~41,096 units), indicating a major improvement in clustering compactness. However, the decrease further diminished from k=4 to k=5 (~14,863 units), which is a considerably smaller reduction compared to the previous two. While k=4 and k=5 do exhibit lower WCSS values, the drop in WCSS from k=2 to k=3 suggests, it to be critical point where the majority of the variance in the data captured—an “elbow” in the WCSS plot. Therefore, k=3 was selected as the optimal number of clusters, effectively balancing data representation with model simplicity.

As shown in Fig. 2A, in the selected three clusters (k), the Cluster 1 comprises 23 states, making it the largest cluster. The observed Within-Cluster Sum of Squares (WCSS) of 41,965 reflects the overall variation within this cluster, consistent with its larger size. Despite this variability, the observed average distance of 36 units from the centroid suggests that the states within this cluster share common characteristics, forming a compact group. Meanwhile, the maximum distance of 82 units from the centroid indicates a few states exhibiting some degree

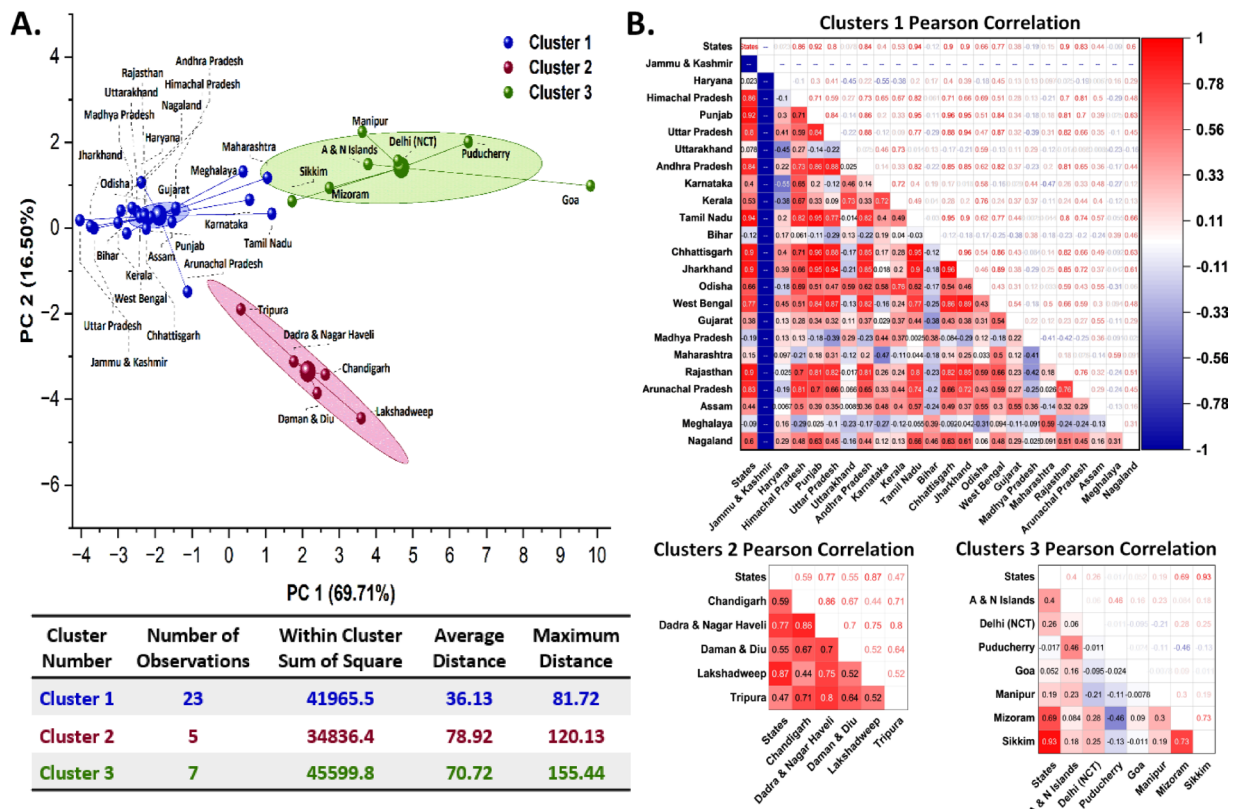


Fig. 2. (A) K-means Clustering of Indian States Based on MCCD Characteristics. Cluster 1 includes 23 states, characterized by a Within-Cluster Sum of Squares (WCSS) of 41,965, an average distance of 36 units from the centroid, and a maximum distance of 82 units, indicating moderate variation and compactness among states. Cluster 2 consists of 5 states, with a WCSS of 34,836 and an average distance of 79 units from the centroid, signifying significant divergence in reporting patterns and presence of two outlier states (Goa and Sikkim). Cluster 3 comprises 7 states and exhibiting a WCSS of 45,599, an average distance of 70 units, and a maximum distance of 155 units, reflecting substantial internal variability and the presence of outlier states (Tripura and Lakshadweep). (B) Pearson Correlation Analysis of MCCD Percentages among states within the identified clusters. Strong positive correlations (0.95 between Punjab and Tamil Nadu) indicate similar MCCD-reporting trends among certain states in Cluster 1. In contrast, negative correlations (−0.55 between Haryana and Andhra Pradesh) reveal divergent reporting patterns. In Cluster 2, states like Chandigarh and Dadra and Nagar Haveli show strong correlations (0.86), while Cluster 3 displays a mix of strong positive correlations (0.93 between A & N Islands and Sikkim) and negative correlations (−0.46 between Puducherry and Mizoram).

of deviation within Cluster 1. Cluster 2, on the other hand, comprises only five states. The WCSS of 34,836 and average distance of 79 units from the centroid, shows that at least one state diverges significantly from the others, potentially due to unique MCCD characteristics. In addition, Cluster3, placed between Cluster1 and 2, demonstrates a moderate size with seven states. With a WCSS of 45,599 indicating significant internal variability among the states, along with the average distance of 70 units from the centroid reflecting that the states are not tightly clustered, the maximum distance of 155 units from the centroid highlights the presence of extreme outlier states.

Divergent MCCD-reporting trends among clusters of Indian states

The observed high WCSS accompanied by higher maximum distance from the centroid indicate that certain states in Cluster2 and 3 have characteristics that diverge sharply from the others, leading to high variability within the group. As shown in Fig. 2A, there are two states in each cluster (Goa and Sikkim in Cluster2, and Tripura and Lakshadweep in Cluster3) diverge significantly from the other states. These observations prompted us to perform a correlation analysis within each cluster to determine the extent to which the states within each cluster correlated with one another in terms of MCCD-percentage.

The Pearson-correlation analysis of clusters revealed varied relationships between states in terms of MCCD-percentage (Fig. 2B). In Cluster1, states such as Punjab and Tamil Nadu (0.95), Jharkhand and Chhattisgarh (0.96), and Uttar Pradesh and Jharkhand (0.94) show strong positive correlations, indicating similar trends in MCCD-reporting across these regions. Conversely, negative correlations between Haryana and Andhra Pradesh (-0.55) and Haryana and Karnataka (-0.55) highlight divergent MCCD-reporting patterns, suggesting contrasting healthcare systems or data practices. Jammu & Kashmir and Uttarakhand exhibit weaker or insignificant correlations with others, showing minimal relationships with other states. However, in Cluster 2, Chandigarh, Dadra and Nagar Haveli, and Lakshadweep exhibit strong correlations with each other (0.86) and with Tripura (0.71 and 0.80, respectively), suggesting consistent MCCD-reporting patterns. Daman and Diu, shows moderate correlations with Dadra and Nagar Haveli (0.70) and Tripura (0.64), while weaker correlations with Chandigarh and Tripura (0.47) indicate slightly differing trends (Fig. 2B). Further, a mix of strong positive and weak or negative correlations can be seen in the Cluster 3 states (Fig. 2B). The strongest positive correlation between A & N Islands and Sikkim (0.93), followed by Mizoram and Sikkim (0.73), reflects similar MCCD-reporting trends in these states. Conversely, Puducherry displays negative correlations with Mizoram (-0.46), Sikkim (-0.13), Delhi (NCT) (-0.011), and Manipur (-0.11), indicating divergent reporting patterns.

Healthcare variables shaping the MCCD trends among clusters of Indian states

The varying Pearson correlation patterns observed across states—both within and between clusters—reflect differences in reporting practices. Multiple factors hinder MCCD rates in India, which can be broadly categorized into four key areas: (1) Systemic—lack of trained staff, infrastructure gaps, administrative load, and poor data quality, etc.; (2) Socio-demographic—geographic, cultural, religious beliefs, and economic barriers, etc.; (3) Healthcare system—doctors per 1000 people, hospital availability, and non-institutional deaths, etc.; (4) Policy and governance—Registration of Births and Deaths Act, 1969 and MCCD law enforcement, infrastructure reforms, national and state health policies, etc. Among all factors, the healthcare variables are pivotal, as they directly determine MCCD rates and simultaneously mediate the effects of systemic, socio-demographic, and policy-governance influences. Further, we selected four key healthcare variables—the number of doctors per 1000 population, the number of hospitals registered in the MCCD system, the number of hospitals actively reporting MCCD, and the number of registered hospitals not reporting MCCD per 10,000 population—to evaluate their impact on MCCD. The multiple linear regression was used to evaluate effect of healthcare variables across Indian using state-level data from 2008 to 2020 (excluding Jammu & Kashmir, Uttar Pradesh, Jharkhand, and Madhya Pradesh due to the unviability of data). As states exhibit significant variation in both MCCD rates (Fig. 3A showing average of past 13-years) and the proportion of hospitals reporting, we tested two regression models: one using the raw MCCD rate and a second using a weighted MCCD rate adjusted for reporting characteristics.

The multiple linear regression analysis for raw MCCD rate (percent) demonstrated a statistically significant intercept of 18.02 (SE = 2.08, $p < 0.001$) in Model 1, indicating the baseline MCCD rate when all independent variables are zero (Table 1). Among the healthcare variables, the number of doctors per 1000 population showed a strong positive association with MCCD reporting ($\beta = 45.27$, SE = 6.06, $p < 0.001$). Interestingly, the number of hospitals per 10,000 population exhibited a statistically significant negative association with the MCCD rate ($\beta = -70.77$, SE = 22.01, $p = 0.001$). This counterintuitive finding likely reflects that, on average, only about 40% of hospitals currently report MCCD data, indicating that hospital density alone does not ensure improved reporting unless those institutions are actively engaged in certification practices. In contrast, the number of hospitals actively reporting MCCD was found to be a strong positive predictor of the MCCD rate ($\beta = 397.54$, SE = 28.55, $p < 0.001$), highlighting the critical role of institutional participation in ensuring accurate cause-of-death documentation. Conversely, the number of hospitals not reporting MCCD per 1000 population demonstrated a negative association ($\beta = -53.09$, SE = 32.41), although this effect did not reach statistical significance ($p = 0.10$), suggesting a possible adverse influence of non-participating facilities on overall MCCD performance. Model 1 explained approximately 45% of the variance in MCCD rates ($R^2 = 0.45$; adjusted $R^2 = 0.45$), with statistically significant regression equation ($F(4,398) = 82.91$, $p < 0.0001$) (Table 1). Supporting the assumptions of linearity and homoscedasticity, residual analysis revealed that the residuals vs. predicted values plot displayed a random scatter around zero (Fig. 3B). Similarly, the histogram of residuals demonstrated an approximately symmetric, bell-shaped distribution centered near zero, with most residuals falling within ± 2 standard deviations ($\sim 95\%$), and with maximum and minimum residual values ranging from -72.6 to $+62.3$, suggesting minimal skewness (estimated skewness ~ 0.1 to 0.2) and kurtosis (~ 2.8 to 3.2), thereby confirming the appropriateness of model assumptions (Fig. 3D).

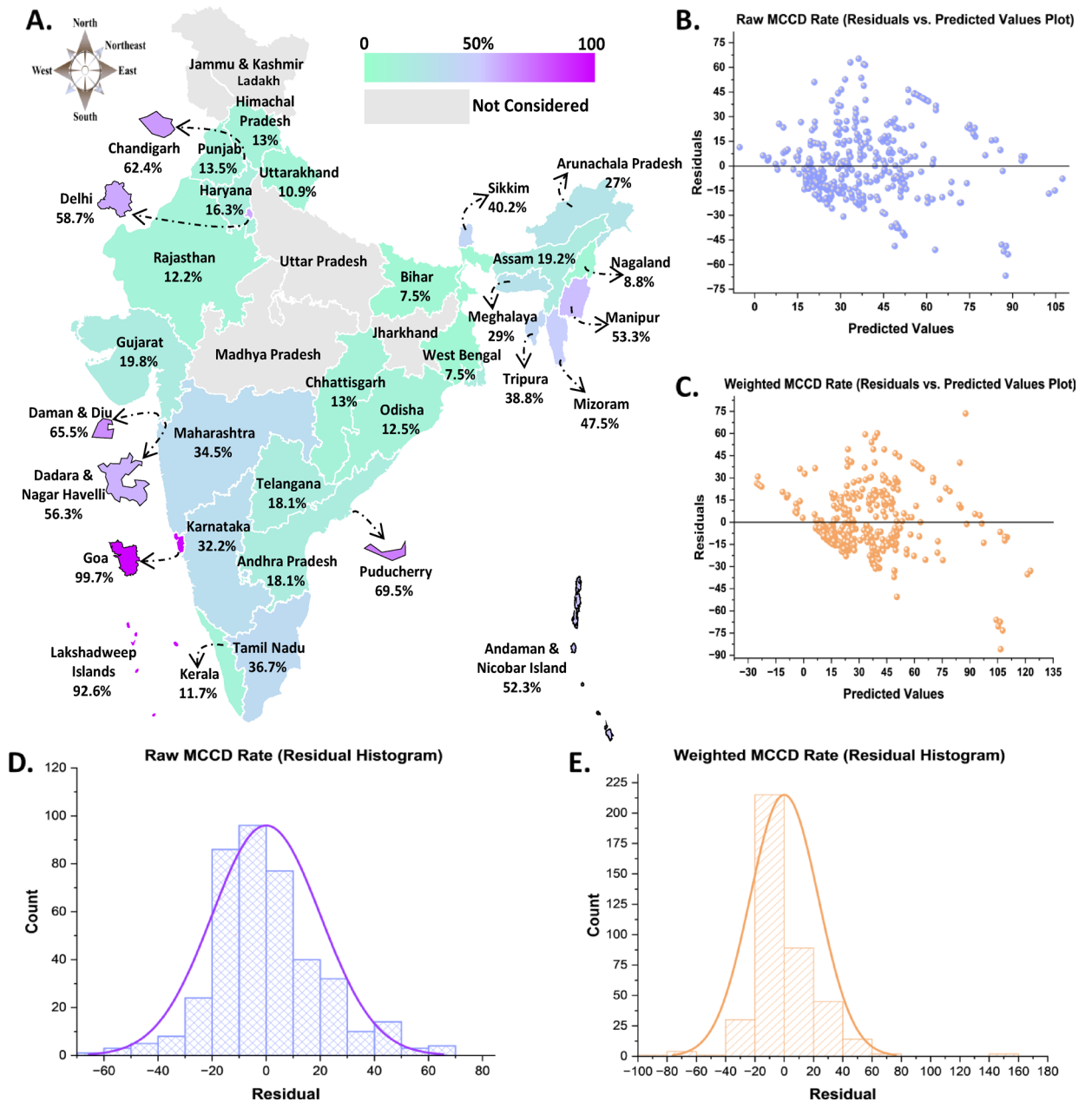


Fig. 3. (A) Map of India showing the average MCCD rates over the past 13 years (2008–2020) across different states, (B, C) Residuals vs. predicted plot for Model 1 (raw MCCD rate) and Model 2 (weighted MCCD rate), showing random scatter and indicating homoscedasticity, (D, E) Histograms of residuals for Model 1 and Model 2, displaying roughly normal, symmetric distributions centered near zero. Most residuals fall within ± 2 SD, with ranges from -85 to $+70$ (Model 1) and -72.6 to $+62.3$ (Model 2).

While Model 1 provided valuable insights into factors associated with the raw MCCD rate, it did not account for substantial variation in reporting completeness across Indian states. To address this limitation and reduce potential measurement bias, we calculated a Weighted MCCD Rate using the equation as:

$$\text{WeightedMCCD Rate(State)} = \left[\text{MCCD rate (\%)} \times \left(\frac{\text{Number of Hospitals Reporting MCCD (State)}}{\text{Number of Hospitals Registered to Report MCCD (State)}} \right) \right]$$

By using a weighted outcome variable that incorporates the proportion of hospitals actually reporting MCCD data, we aimed to reflect the true performance of each state in MCCD reporting by avoiding the assumption that non-reporting hospitals perform similarly to those that report MCCD.

As evident from Table 1, the weighted MCCD rate regression analyses (Model 2) provided more refined insights into the determinants of MCCD reporting compared to Model 1, the intercept value of 15.27 (SE = 2.41, $p < 0.001$) indicated the baseline MCCD rate when all independent variables are zero. Among the healthcare

Independent variables	Statistical summary	Raw MCCD rate (Model 1)	Weighted MCCD rate (Model 2)
Model fit Statistics	Reduced Chi-Sqr	410.11	551.61
	R Value	0.67	0.71
	R-Square (COD)	0.45	0.51
	Adj. R-Square	0.45	0.51
No variable	Intercept Value	18.02	15.27
	SE	2.08	2.41
	95% CI (LCL / UCL)	14 / 22.19	10.53 / 20.02
	<i>p</i> -value	<0.0001	<0.0001
Doctor per, 1000 People	Regression coefficients (β)	45.27	35.61
	SE	6.06	7.03
	95% CI (LCL / UCL)	33.33 / 57.20	21.77 / 49.45
	<i>p</i> -value	<0.0001	<0.0001
Hospitals per 10,000 People	Regression coefficients (β)	-70.77	-116.07
	SE	22.01	25.53
	95% CI (LCL / UCL)	-114.05 / -27.84	-166.27 / -65.87
	<i>p</i> -value	0.001	<0.0001
Hospitals per 10,000 People (Reporting MCCD)	Regression coefficients (β)	397.54	548.45
	Standard Error	28.55	33.11
	95% CI (LCL / UCL)	341.41 / 453.67	483.35 / 613.55
	<i>p</i> -value	<0.0001	<0.0001
Hospitals per 10,000 People (Not Reporting MCCD)	Regression coefficients (β)	-53.09	-183.13
	SE	32.41	37.59
	95% CI (LCL / UCL)	-116.82 / 10.64	-257.05 / -109.22
	<i>p</i> -value	0.10	<0,0001

Table 1. Multiple linear regression analysis and statistical Summary of Independent variables and dependent variable MCCD for Raw MCCD Rate and weighted MCCD rate. R value, Correlation coefficient; Adjusted R-Square, Adj. R-Square; COD, Coefficient of Determination; SE, Standard Error; LCL, Lower Confidence Limit; UCL, Upper Confidence Limit.

variables, the number of doctors per 1000 population again showed a strong positive association with the MCCD rate ($\beta = 35.61$, $SE = 7.03$, $p < 0.001$), suggesting that states with higher physician density tend to report more complete death certification. Reinforcing the earlier finding that hospital density alone does not improve MCCD unless accompanied by active reporting, the number of hospitals per 10,000 population exhibited a negative association ($\beta = -116.07$, $SE = 25.53$, $p < 0.001$). Conversely, the number of hospitals reporting MCCD emerged as the strongest positive predictor ($\beta = 548.45$, $SE = 33.11$, $p < 0.001$). In contrast, the number of hospitals not reporting MCCD showed a statistically significant negative association ($\beta = -183.13$, $SE = 37.59$, $p < 0.001$), implying that these institutions dilute overall MCCD performance and further highlighting the importance of institutional participation in improving death certification practices. Compared with Model 1, the Model 2 explained 51% of the variance in the weighted MCCD rate ($R^2 = 0.51$; adjusted $R^2 = 0.51$), with statistically significant model ($F(4,398) = 103.6$, $p < 0.0001$) (Table 1). Residual diagnostics further supported the model's assumptions, with the residuals vs. predicted values plot showing a random scatter around zero, confirming linearity and homoscedasticity (Fig. 3C). Similarly, the histogram of standardized residuals displayed a normal, bell-shaped curve centered around zero, with approximately 95% of residuals falling within ± 2 standard deviations (Fig. 3E). The estimated range of residuals was approximately -85 to $+70$, with visual inspection indicating mild skewness (~ 0.1 to 0.2) and near-normal kurtosis (~ 2.8 to 3.2), confirming that the assumptions of linear regression were adequately met.

Factors beyond healthcare variables shaping MCCD-rates in India

Given its superior explanatory power (adjusted $R^2 = 0.5335$ vs. 0.4490), stronger statistical significance of key predictors, and incorporation of reporting completeness through weighting, Model 2 was selected for all subsequent multiple linear regression analyses to better understand the factors among health of healthcare variables driving MCCD within three predefined clusters. As shown in Table 2 (and Fig. 4A, B, C, D, E, F), across all three clusters, multiple linear regression models significantly predicted weighted MCCD rates (all $p < 0.0001$) and satisfied key regression assumptions, however, explanatory power (R^2) and residual variability differed markedly. In Cluster 1, the model explained 34.5% of MCCD variance ($R^2 = 0.3448$; adj. $R^2 = 0.334$; $F(4,242) = 31.83$; $RMSE = 8.87$), with standardized residuals ranging from -25 to $+28$ (Fig. 4D). In contrast, Cluster2, accounted for 40% of variance ($R^2 = 0.4$; adj. $R^2 = 0.4$; $F(4,60) = 10$) and exhibited higher uncertainty with $RMSE = 33.3$; residuals: -70 to $+140$ (Fig. 4E). However, among the three clusters, Cluster 3, demonstrated the strongest fit, explaining 57.5% of variance ($R^2 = 0.5754$; adj. $R^2 = 0.556$; $F(4,86) = 29.14$) with $RMSE = 16.16$ (residuals ≈ -45 to $+40$; Fig. 4F). Additionally, all clusters, residuals vs. predicted plots (Fig. 4A, B, C) showed

Independent variable	Statistical summary	Cluster 1	Cluster 2	Cluster 3
	Average MCCD Rate	18%	63%	60%
Model fit Statistics	Reduced Chi-Sqr	78.7	1112.7	261.07
	R Value	0.58	0.65	0.76
	RMSE	8.87	33.3	16.16
	R-Square (COD)	0.34	0.4	0.58
	Adj. R-Square	0.33	0.4	0.56
No variable	Intercept value	9.63	54.5	29.26
	SE	1.04	8.4	8.23
	95% CI (LCL/ UCL)	7.58 / 11.68	37.6/ 71.2	12.89 / 45.62
	<i>p</i> -value	<0.001	<0.001	<0.001
Doctor per 1000 People	Average Number of doctors per 1000 People	0.144	0.266	0.332
	Regression coefficients (β)	- 3.46	9.6	5.51
	SE	3.83	22.7	15.3
	95% CI (LCL/ UCL)	- 11.01 / 4.09	- 36 / 55	- 24.90 / 35.92
	<i>p</i> -value	0.325	0.67	0.719
Hospitals per 10,000 people	Average Number of Hospitals per 10,000 People	0.713	0.646	0.646
	Regression coefficients (β)	1.45	- 31.9	14.66
	SE	1.05	14	18.19
	95% CI (LCL/ UCL)	- 0.62 / 3.54	- 59.9 / -3.8	- 21.50 / 50.82
	<i>p</i> -value	0.16	0.37	0.422
Hospitals Registered in MCCD (Per 10,000 People)		0.415	0.577	0.572
Hospitals per 10,000 people (registered and reporting MCCD)	Average Number of Hospitals Registered and Reporting MCCD per 10,000 People	0.217	0.56	0.528
	Regression coefficients (β)	15.91	61.3	37.3
	SE	1.96	10.9	14.67
	95% CI (LCL/ UCL)	12.03 / 19.78	39.4 / 83.2	8.12 / 66.47
	<i>p</i> -value	<0.001	<0.001	0.0128
Hospitals per 10,000 People (registered but not reporting MCCD)	Average Number of Hospitals Registered but Non-reporting MCCD per 10,000 People	0.199	0.0173	0.0435
	Regression coefficients (β)	- 8.10	- 47.6	- 135.09
	SE	1.46	53.2	30.33
	95% CI (LCL/ UCL)	- 10.98 / - 5.22	- 154 / 58.8	- 195.38 / - 74.80
	<i>p</i> -value	<0.001	0.37	<0.001

Table 2. Multiple linear regression analysis and statistical Summary of Independent variables and dependent variable MCCD for different clusters 1, 2, and 3. R value, Correlation coefficient; Adjusted R-Square, Adj. R-Square; RMSE, root mean squared error; COD, coefficient of determination; SE, standard error, LCL, lower confidence limit; UCL, upper confidence limit.

no pattern, as all residuals scattered randomly around zero across the spectrum of predicted values without any discernible pattern or funnel shape, supporting the assumptions of linearity and homoscedasticity (Histogram in Fig. 4D, E, F). Moreover, the distribution of standardized residuals was approximately bell-shaped with $\sim 95\%$ within ± 2 SD and cantered near zero among all clusters, confirming normality. These results suggest that healthcare and reporting variables accurately explain MCCD reporting completeness: Cluster 3 > Cluster 2 > Cluster 1.

The intercept values for Cluster 1, 2, and 3 revealed significant baseline differences in MCCD-rates across these groups, when all the independent healthcare variables were considered zero (Table 2). Cluster 1, with an average MCCD-rate of 18%, exhibited the lowest intercept (β) value of 9.63 (SE = 1.04, $p < 0.001$) among all clusters (Fig. 4G). This signifies that the expected MCCD-rate remains low even in the absence of healthcare variables. Conversely, Cluster 2, characterized by the highest average MCCD-rate of 63%, exhibited a substantially higher intercept of 54.5 (SE = 8.4, $p < 0.001$; Fig. 4H). Similarly, Cluster 3, characterized by average MCCD-rate of 60%, exhibited an intercept coefficient of 29.2 (SE = 8.23, $p < 0.001$; Fig. 4I), positioning it between Cluster 1 and Cluster 2 in terms of baseline MCCD-rates.

The critical role of active MCCD-reporting by hospitals in elevating MCCD-rates

For improving overall health in a population, a higher doctors per 1000 people or, more doctors per capita—is generally associated with better health outcomes. As evident from Table 2, significant differences exist in the doctors per 1000 people among the three clusters. With the lowest average of 0.14 doctors per 1000 people, the states in Cluster 1 have the fewest doctors. Although the *p* value for all three clusters for doctor density was not statistically significant, the observed coefficient (β) of - 3.46 (SE = 3.83, $p = 0.32$) for Cluster 1 indicates

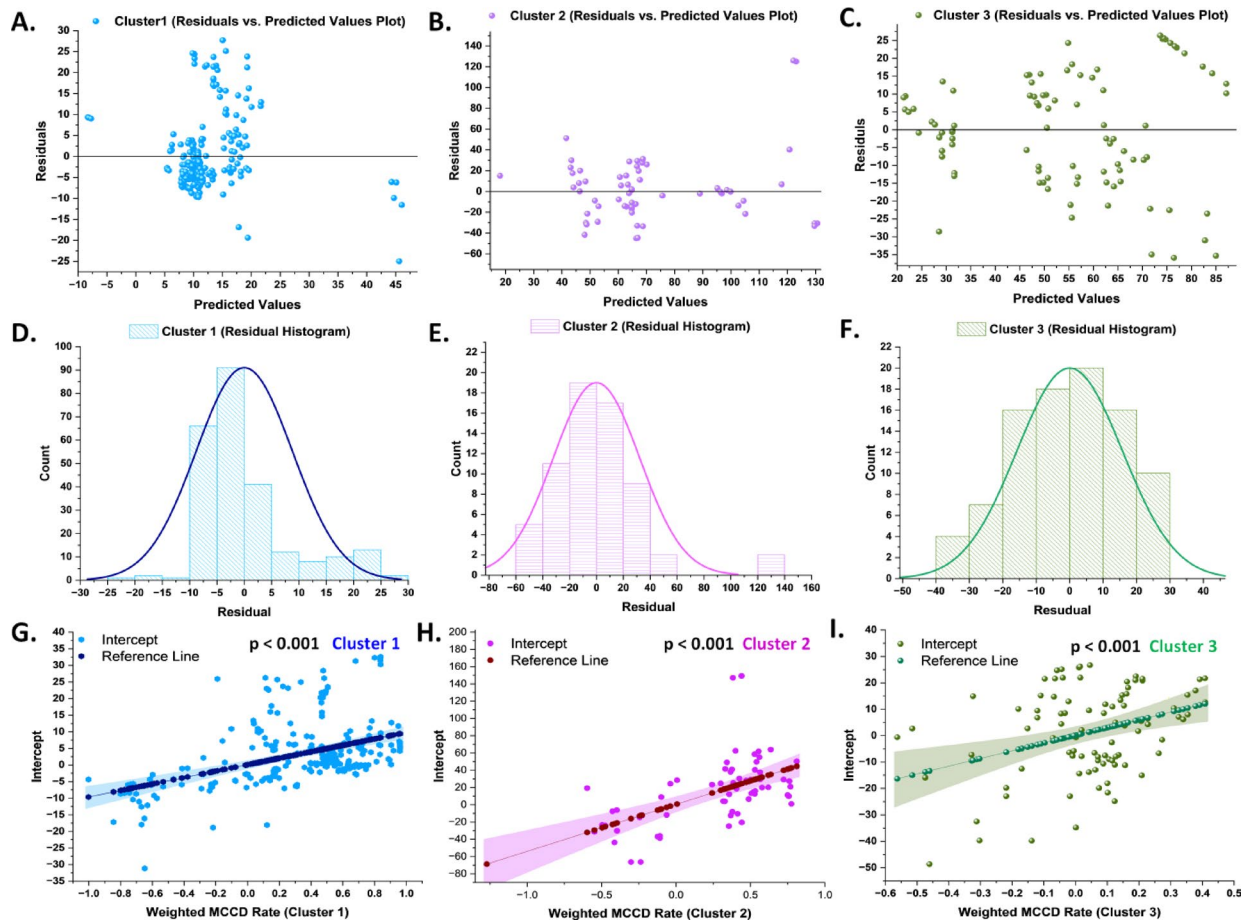


Fig. 4. Performance and residual analysis of three cluster-specific multiple linear regression models predicting weighted MCCD rates, (A–C) Residuals vs. predicted plots for clusters 1, 2, and 3, showing random scatter around zero, supporting linearity and homoscedasticity, (D–F) Histograms of standardized residuals for each cluster, displaying roughly normal, bell-shaped distributions with $\sim 95\%$ within ± 2 SD. Residual ranges: -25 to $+28$ (Cluster 1), -70 to $+140$ (Cluster 2), and -45 to $+40$ (Cluster 3), (G–I) Multiple linear regression analysis scatter plots illustrating the baseline MCCD-rates across clusters when healthcare variables are zero. Cluster 1, with 18% MCCD, has the lowest intercept (9.63, SE = 1.04, $p < 0.001$); Cluster 2, with 63% MCCD, has a higher intercept (54.5, SE = 8.4, $p < 0.001$), and Cluster 3, with 60%, has an intercept of 29.2 (SE = 8.23, $p < 0.001$), placing it between clusters 1 and 2.

that a lower number of doctors may negatively impact MCCD rates (Fig. 5A). In contrast, Cluster 2 (~ 0.27 doctors per 1000 people) and Cluster 3 (~ 0.33 doctors per 1000 people) showed positive coefficients of 9.6 (SE = 22.7, $p = 0.67$) and 5.51 (SE = 15.30, $p = 0.72$), respectively, suggesting that more doctors may positively influence MCCD rates (Fig. 5B, C). However, the high p -values indicate that variation in doctor density within each cluster alone does not significantly drive certification completeness. On the other hand, a higher hospital-to-population ratio indicates better healthcare infrastructure and is a key indicator of healthcare accessibility in a community. Although, among all the clusters (Table 2), Cluster 1 exhibited the highest average of 0.713 hospitals per 10,000 people (0.64 hospitals per 10,000 people in clusters 2 and 3), the non-significant p -values and the coefficients of 1.45 (SE = 1.01), -31.9 (SE = 14), and 14.66 (SE = 18.2) for clusters 1–3 indicate that sheer differences in the number of hospitals alone may not translate into better death-certification rates (Fig. 5D, E, F). These results prompted us to hypothesize that the variations in the MCCD rate among clusters are due to differences in the number of hospitals actively reporting MCCD, rather than the total number of hospitals in each cluster. Surprisingly, a significant disparity was observed among clusters for the number of hospitals present per 10,000 people and the number of hospitals registered for reporting the MCCD per 10,000 people (Table 2). With an average of 0.415 hospitals ($\sim 58.5\%$ of the total hospitals) registered for reporting MCCD out of the overall average of 0.713 hospitals, Cluster 1 demonstrates the lowest number of hospitals registered for reporting MCCD. In contrast, clusters 2 and 3 showed significantly higher registration rates, with 0.577 out of 0.646 hospitals ($\sim 89.3\%$) and 0.578 out of 0.646 hospitals ($\sim 88.5\%$), respectively.

Further, the analysis of the number of hospitals actively reporting MCCD per 10,000 people revealed that, on average, Cluster 1 has only 0.22 hospitals per 10,000 people reporting MCCD actively ($\sim 52\%$ of registered hospitals) (Table 1). In contrast, clusters 2 and 3 have higher averages, with 0.56 and 0.52 hospitals per 10,000 people actively reporting MCCD ($\sim 97\%$ and $\sim 92\%$ of registered hospitals), respectively. Further, the small

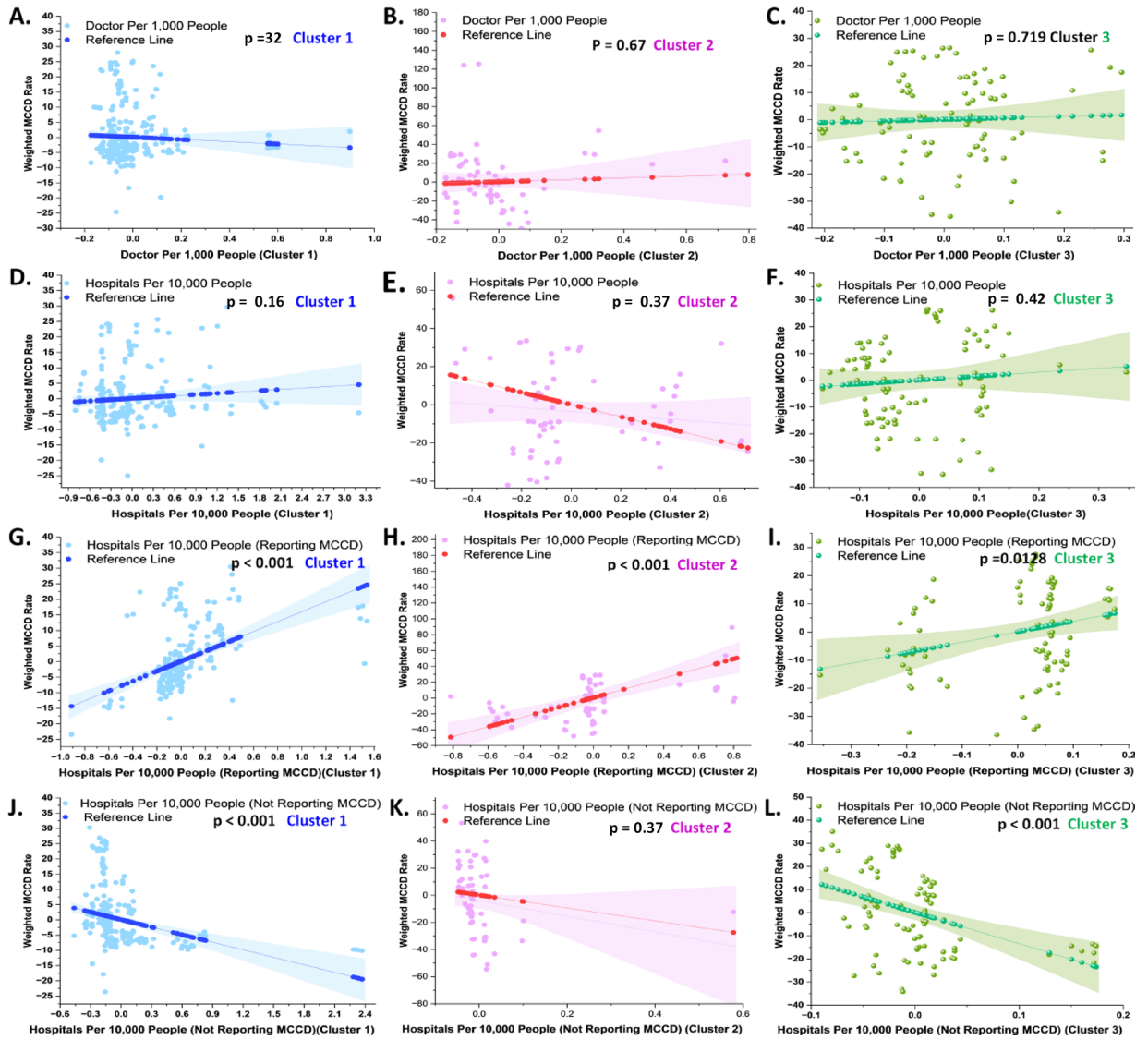


Fig. 5. Scatter plots illustrating multiple linear regression analysis of various independent variables on Medical Certification of Cause of Death (MCCD) rates, with data stratified into three clusters. (A–C) Display the relationship between the number of doctors per 1000 population and MCCD rates. Cluster 1 demonstrate a negative coefficient of -3.46 ($p=0.32$), suggesting a non-significant negative impact of fewer doctors on MCCD rates; Cluster 2 exhibits a positive coefficient of 9.6 ($p=0.67$), and Cluster 3 shows a positive coefficient of 5.51 ($p=0.72$), both indicating nonsignificant positive associations, (D–F) Shows the effect of population-to-hospital ratio, with Cluster 1 presenting a positive coefficient of 1.45 ($p=0.16$), Cluster 2 a negative coefficient of -31.9 ($p=0.37$), and Cluster 3 a positive coefficient of 14.66 ($p=0.42$); all suggest no significant impact, (G–I) Shows the effect of a number of hospitals actively reporting MCCD; Cluster 1 demonstrating a significant positive coefficient of 15.91 ($p<0.0001$), Cluster 2 a higher positive coefficient of 61.3 ($p<0.001$), and Cluster 3 a coefficient of 37.3 ($p=0.012$), indicating that increased reporting is associated with higher MCCD rates, (J–L) Analyze hospitals registered for MCCD but not actively reporting; Cluster 1 showing a significant negative coefficient of -8.10 ($p<0.001$), Cluster 2 a non-significant negative coefficient of -47.6 ($p=0.37$), and Cluster 3 a significant negative coefficient of -135.09 ($p=0.001$), demonstrating that hospitals not reporting MCCD data negatively influence MCCD rates across all clusters.

coefficient of 15.91 ($SE=1.96$; Fig. 5G) and a highly significant p -value ($p<0.0001$) for hospitals reporting MCCD in Cluster 1, in comparison to the coefficients of 61.3 ($SE=10.9$, $p<0.001$; Fig. 5H) and 37.30 ($SE=14.67$, $p=0.012$; Fig. 5I) for Clusters 2 and 3, respectively, underscores that institutional participation is the primary driver of higher MCCD rates. To further confirm this, we evaluated how hospitals registered for MCCD but not reporting MCCD affected the overall MCCD rate (Table 1 and Fig. 5), K, L). Surprisingly, Cluster 1 demonstrated the highest number of non-reporting hospitals per 10,000 people (~ 0.19 or 48% of hospitals), with a further negative coefficient of -8.10 ($SE=1.46$, $p<0.001$; Fig. 5J) indicating that non-reporting hospitals have a highly negative impact on MCCD reporting. Further, the importance of institutional participation in MCCD reporting

is mirrored by the negative coefficient of -47.6 ($SE = 53.2$, $p = 0.37$; Fig. 5K) and -135.09 ($SE = 30.33$, $p < 0.001$; Fig. 5L) for Clusters 2 and 3, which have only 0.017 (~3% of registered hospitals) and 0.043 (~8% of registered hospitals) non-reporting hospitals per 10,000 people, indicating that even small lapses in reporting substantially reduce certification completeness.

Discussion

For a country that accounts for 17.8% of the world's population, the existing MCCD-rate of 22.5%, with only a 2.5% improvement over the past decade, is strikingly concerning. By analyzing the trends in MCCD-reporting over the past 15 years across different states of India, this study reveals significant regional and state-level disparities in MCCD-reporting, highlighting both progress and persistent challenges in improving MCCD practices within the healthcare system.

The findings highlight stark differences in MCCD-rates across regions. Southern states, like Tamil Nadu, Karnataka, and Andhra Pradesh, demonstrate a consistent upward trend in MCCD-reporting, reflecting well-developed healthcare infrastructure and administrative efficiency. In contrast, North India's low average MCCD-rate of 13% during 2015–2020 illustrates persistent gaps in healthcare access and reporting practices. Similarly, while states like Chhattisgarh and West Bengal in East India have shown improvements, others like Jharkhand continue to underperform. Moreover, the contrasting performance of UTs like Chandigarh, Lakshadweep (94–95%) and Delhi (57–59%) further emphasizes the varying levels of administrative efficiency and healthcare system engagement in different regions. At the state level, these variations indicate that MCCD-rates are not only the regional issue but a state-specific and national problem, suggesting the presence of potential administrative barriers, reporting inefficiencies, and gaps in population awareness that necessitate localized strategies for improvement.

The K-means clustering analysis for data corresponding to 2006–2020 provides a nuanced understanding of these disparities by grouping states with shared MCCD characteristics in three clusters. Cluster1, the largest (23 states), shows high variability, with states such as Punjab and Tamil Nadu exhibiting strong positive correlations in MCCD trends, while outliers like Jammu & Kashmir and Uttarakhand display weaker correlations. Cluster2 (5 UTs) and Cluster3 (5 states and 2 UTs), though smaller, include notable outliers like Goa and Sikkim, which differ significantly from other states in their clusters. These findings highlight the need to understand both shared and unique factors affecting MCCD-reporting across regions. The underlying cause for the differential MCCD-reporting patterns among states may be attributed to India's decentralized healthcare system, which divides healthcare services between the central and state governments^{11,12}. While the central government oversees UTs and national programs, state governments manage healthcare in their regions, resulting in varied budget allocations based on state priorities, economic conditions, and requirements¹³. This emphasizes the need for state-specific interventions and cross-state learning to enhance the consistency and reliability of MCCD-reporting across India.

Although differences exist in healthcare system management between states, the healthcare variables that strengthen MCCD-reporting remain consistent across all states. The Government of India has made efforts to strengthen the health system over the past few years. The analysis of healthcare-variables reveals important insights into systemic gaps and opportunities for improvement in the healthcare system. The population-to-doctor ratio, a key indicator of healthcare adequacy, reflects access to medical care. As of December 2023, the doctor-patient ratio stands at 1:834, calculated using approximately 1.4 million registered allopathic doctors and 565,000 AYUSH doctors, exceeding the WHO standard of 1:1000^{14,15}. However, AYUSH doctors and all allopathic hospitals are currently not included in the MCCD registry program. Additionally, the number of available allopathic doctors per 1000 population varies significantly across different Indian states¹⁰. States in Cluster1, with fewer doctors (0.14 per 1000 people), demonstrate lower MCCD-rates, while regions with a higher doctor density (0.35–0.50 doctors per 1,000) show moderate improvements in MCCD-rates. However, the limited correlation highlights that merely increasing doctor numbers without addressing systemic challenges may not yield substantial improvements in reporting practices.

Furthermore, the density of hospitals within a population serves as an essential measure of access to hospital care in any region, and hospital density analysis per 10,000 people has emerged as a crucial determinant in MCCD-reporting¹⁶. While the healthcare dynamics remain consistent across all clusters, Cluster2 and 3 demonstrate better MCCD-rates, largely due to a significant portion of their populations residing in urban areas, which facilitates easier access to healthcare and certification processes¹⁷. Conversely, Cluster1 states, with lower MCCD-rates, struggle with underreporting, particularly due to a higher population in rural regions where access to certified medical facilities is limited, such as in Uttar Pradesh and Bihar¹⁶.

Improving MCCD-rates has far-reaching public health implications and accurate death certification data is essential for understanding disease burdens, guiding resource allocation, and shaping health policy. This underscores the importance of increasing hospital density and accessibility in rural areas, which are vital for improving MCCD-rates in states with lower reporting. Additionally, it is crucial to verify that such increases in hospital density and accessibility translates to active MCCD-reporting, as active hospital reporting has emerged as the greatest influencing healthcare variable affecting MCCD-rates among all clusters.

The regression analysis reveals that low-reporting states within Cluster 1 have only around 15% of hospitals actively reporting MCCD out of total hospitals, leading to an average MCCD rate of just 18%. This illustrates significant gaps in compliance and oversight. In comparison, clusters 2 and 3 have a much higher active reporting rate, with approximately 92–97 of hospitals actively reporting MCCD, resulting in significantly higher rates of 63% and 60%, respectively. The combination of a low number of registered hospitals for MCCD reporting and insufficient active participation among those registered in Cluster 1 is a major factor contributing to the lower MCCD rates observed in these states. This highlights the essential importance of active MCCD reporting by hospitals in improving overall MCCD rates. These findings collectively suggest that addressing regional and

state-level disparities in MCCD reporting through active participation by non-reporting hospitals is crucial for ensuring equitable healthcare outcomes and effectively responding to emerging health challenges. Efforts should focus on increasing the number of hospitals reporting MCCD, coupled with enhanced active MCCD reporting, particularly in the underperforming states of clusters 1 and 3, especially in regions such as North and East India. This can be achieved through state-level audits, financial or administrative incentives for hospitals to report MCCD, and public awareness campaigns highlighting the importance of accurate death certification. High-performing states and UTs, such as Goa and Lakshadweep, can serve as models for implementing best practices in reporting systems.

While this study provides a comprehensive analysis of MCCD-reporting trends, certain limitations, such as missing data for some states and reliance on available administrative records, must be acknowledged. Although the findings highlight active MCCD reporting as a key factor influencing MCCD rates, it is essential to recognize that several other determinants also play a critical role. These include healthcare infrastructure, state-specific health policies, cultural and religious beliefs, socioeconomic and demographic factors, economic constraints, and administrative capacity. Systemic challenges—such as the shortage of trained personnel, infrastructure deficits, administrative burden, poor data quality, and the high prevalence of non-institutional deaths—further affect the completeness and accuracy of MCCD data. Additionally, governance-related aspects, including the implementation and enforcement of the Registration of Births and Deaths Act (1969), and variations in national and state-level policy execution, significantly shape reporting practices and MCCD rates. Future research should explore the qualitative aspects of MCCD reporting, with a particular focus on systemic barriers, non-institutional death certification, and the broader socio-cultural context influencing mortality data quality across states.

Conclusion

In conclusion, our analysis demonstrates substantial inter-cluster disparities in hospital registration and active MCCD reporting, revealing that institutional participation is a key determinant of MCCD coverage. Cluster 1, with the lowest proportion of registered (58.5%) and actively reporting hospitals (52%), showed limited MCCD performance ($\beta = 15.91$, $p < 0.0001$), while high-performing Clusters 2 and 3 demonstrated $> 88\%$ registration and $> 90\%$ reporting. Moreover, non-reporting hospitals—though registered under MCCD—significantly undermine MCCD rates, especially in Cluster 1 ($\beta = -8.10$), underscoring the critical importance of strengthening MCCD-reporting practices across India. Although this negative coefficient (β) for hospital density does not mean that having more hospitals harms or reduces MCCD certification, it suggests that simply having more hospitals will not improve MCCD rates unless they are actively engaged in the reporting process and consistently reporting. Addressing these disparities and leveraging insights from clustering and regression analyses can pave the way for a more efficient and equitable mortality surveillance system.

Methods

Study design

This is a comprehensive, retrospective cross-sectional study analyzing disparities in MCCD-rates across Indian states and UTs from 2006 to 2020, using data from the Office of the Registrar General & Census Commissioner, India (ORGI). All the 28 states were categorized in five regions—North, East, West, South, and Northeast—with UTs (eight) grouped as a single region for regional and state-level comparisons. Further the K-means clustering was used to group all states and UTs into three clusters based on MCCD-patterns. Followed by the Pearson-correlation analysis to examine relationships among states within the clusters. To identify key healthcare variables / factors affecting MCCD rates, states with five or more consecutive years of missing data were excluded from the analysis. For states with less than five years of missing data, the missing values were imputed. Subsequently, multiple linear regression was used to evaluate the impact of four healthcare variables: the number of doctors per 1,000 population, the number of hospitals registered in the MCCD, hospitals reporting MCCD, and hospitals registered but not reporting MCCD per 10,000 population. For this, two separate multiple linear regression models were tested on complete data (without clustering)—one using the raw MCCD rate data and the other using weighted MCCD data—followed by intra-cluster multiple linear regression analysis of healthcare variables using the weighted MCCD approach/model.

Data source

The MCCD-data for Indian states and UTs from 2006 to 2020 were retrieved from the “Report on MCCD” available on the ORGI website (<https://censusindia.gov.in/nada/index.php/catalog>)⁹. Additional data on the number of doctors and state populations for various years (2008–2020) were obtained from the National Health Profile reports published by the Central Bureau of Health Intelligence, Ministry of Health and Family Welfare, Government of India (<https://cbhidghs.mohfw.gov.in/index1.php?lang=1&level=1&sublinkid=75&lid=1135>)¹⁰.

Data analysis

To ensure data integrity and minimize bias from missing or incomplete reports, we applied a two-stage imputation strategy to estimate missing values. First, to impute values for single-year gaps (i.e., one missing year flanked by valid observations), missing values were substituted using the lower of the two adjacent observations. Second, in cases of multi-year gaps spanning two to five consecutive years (i.e., states with fewer than five years of missing data), the missing values were estimated and imputed based on surrounding trends via linear interpolation using OriginPro 24b software. This interpolation approach was applied to Arunachal Pradesh (2008–2010), Chandigarh (2008–2010), Dadra and Nagar Haveli (2008–2010), Daman and Diu (2008–2010, 2020), Lakshadweep (2008–2012), Tripura (2008–2010), and Haryana (2011–2015). Furthermore, states with five or more consecutive years of missing data (Jammu & Kashmir, Uttar Pradesh, Jharkhand, and Madhya

Pradesh) were excluded from the analysis, followed by modeling using multiple linear regression. To examine how healthcare infrastructure indicators influence MCCD rates across states, the crude MCCD rate or the weighted MCCD rate was considered as the dependent variable (y). Four key healthcare variables were included as independent variables: doctors per 1000 population (x_1), hospitals per 1000 population (x_2), hospitals reporting MCCD per 1000 population (x_3), and hospitals not reporting MCCD per 1000 population (x_4). A multiple linear regression model was developed using the following equation (A_0 is the intercept and A_1, A_2, A_3, A_4 are the regression coefficients (β) representing the effect size of each independent variable):

$$y \text{ (dependent variable MCCD rate)} = A_0 + A_1x_1 + A_2x_2 + A_3x_3 + A_4x_4$$

Additionally, a weighted approach was used to adjust for variations in hospital reporting MCCD data across states, as substantial differences exist in both MCCD rates and the gap between the number of hospitals registered for MCCD and those actually reporting data. The weight for each state was calculated as the proportion of hospitals actively reporting MCCD data to the total number of hospitals registered to report MCCD. The weighted MCCD rate for each state was then determined using the following equation:

$$\text{WeightedMCCD Rate(State)} = \left[\text{MCCD rate (\%)} \times \left(\frac{\text{Number of Hospitals Reporting MCCD (State)}}{\text{Number of Hospitals Registered to Report MCCD (State)}} \right) \right]$$

In this equation, the MCCD Rate denotes the crude MCCD rate (%) for a given year (e.g., 55.98% for Delhi or 6.59% for Bihar in 2015); the Number of Hospitals Reporting MCCD refers to the count of hospitals in that state (e.g., 656 for Delhi or 6 for Bihar) that actually submitted MCCD data that year; and the Number of Hospitals Registered to Report MCCD indicates the total number of hospitals in that state registered to report MCCD data (e.g., 656 for Delhi or 42 for Bihar) for 2015. Using this approach, the weighted MCCD rate for Delhi comes to 55.98% [i.e. $55.98\% \times (656/656$ or 1)] and for Bihar as 0.94% [i.e. $6.59\% \times (6/42$ or 0.14)] for the year 2015. Similarly, the weighted for all states and UTs was calculated for different years. This weight calculation reflected the completeness of MCCD data reporting, with higher weights indicating better participation. All statistical analyses—including descriptive statistics, K-means clustering, Pearson-correlation, and multiple linear regression—was performed using OriginPro 2024b. A significance level of 0.05 and a 95% confidence interval (CI) was applied to all analyses.

Data availability

The datasets used and/or analysed during the current study is available from the corresponding author on reasonable request.

Received: 17 January 2025; Accepted: 5 November 2025

Published online: 03 January 2026

References

- Human Development Report 2023–24. *UNDP* (accessed 15 December 2024); <https://www.undp.org/india/publications/human-development-report-2023-24-0>.
- Rechel, B., Jagger, C., McKee, M., et al. *Living longer, but in better or worse health? In: Living longer, but in better or worse health?* [Internet] (accessed 15 December 2024). (European Observatory on Health Systems and Policies, 2020); <https://www.ncbi.nlm.nih.gov/books/NBK559813/>.
- Association between socioeconomic status and the development of mental and physical health conditions in adulthood: A multi-cohort study—*The Lancet Public Health* (accessed Dec 15, 2024); [https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667\(19\)30248-8/fulltext](https://www.thelancet.com/journals/lanpub/article/PIIS2468-2667(19)30248-8/fulltext).
- Nandita, S., Kumar, K. & Das, B. Death registration coverage 2019–2021, India. *Bull. World Health Organ.* **101**, 102–110 (2023).
- Civil Registration System, Government of India (accessed Dec 15, 2024); <https://dc.crsorgi.gov.in/crs/about>.
- India Population (2024)—Worldometer (accessed Dec 15, 2024) <https://www.worldometers.info/world-population/india-population/>.
- India—SAMPLE REGISTRATION SYSTEM (SRS)—STATISTICAL REPORT 2020 (accessed Dec 15, 2024); <https://censusindia.gov.in/nada/index.php/catalog/44376>.
- CRS - REPORT ON VITAL STATISTICS OF INDIA (accessed Dec 15, 2024); <https://censusindia.gov.in/census.website/data/VSR-REPORT>.
- Government of India | Office of the Registrar General & Census Commissioner, India. (accessed Dec 15, 2024); <https://censusindia.gov.in/census.website/data/mccdrep>.
- National Health Profile: *Central Bureau of Health Intelligence* (accessed Dec 15, 2024); <https://cbhidghs.mohfw.gov.in/index1.php?lang=1&level=1&sublinkid=75&lid=1135>.
- Kaur, M., Prinja, S., Singh, P. K. & Kumar, R. Decentralization of health services in India: Barriers and facilitating factors. *WHO South-East Asia J. Public Health* **1**, 94–104 (2012).
- Healthcare Schemes (accessed Dec 15, 2024); <https://pib.gov.in/pressreleaseshare.aspx?prid=1576128>.
- Ministry of Health and Family Welfare, Government of India (accessed Dec 15, 2024); <https://mohfw.gov.in/>.
- Update on ratio of patients and doctors nurses (accessed Dec 15, 2024); <https://pib.gov.in/pressreleaseshare.aspx?PRID=1985423>.
- The global health observatory, explore a world of health data, indicators countries. *Medical doctors (per 10 000 population)* (accessed Dec 15, 2024); [https://www.who.int/data/gho/data/indicators/indicator-details/GHO/medical-doctors-\(per-10-000-population\)](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/medical-doctors-(per-10-000-population)).
- Census tables | *Government of India*. *Censusindia.gov.in*. (accessed Dec 15, 2024); <https://censusindia.gov.in/census.website/data/census-tables> (2023)
- Rao, K. D. et al. Improving urban health through primary healthcare in south Asia. *Lancet Glob. Health* **12**, e1720–e1729 (2024).

Acknowledgements

The authors would like to express their gratitude to the Indian Council of Medical Research (ICMR), New Delhi, India, for providing research facilities for this study; to the Civil Registration System (CRS) and the Office of the Registrar General & Census Commissioner, India (ORGI) for providing MCCD reports; and to the National Health Profile, Ministry of Health and Family Welfare, Government of India, for providing population and doctor statistics for India for different years.

Author contributions

The study design was conducted by Dr. Khushwant Singh and Dr. Ashoo Grover. Manuscript writing and data analysis was performed by Dr. Khushwant Singh. Manuscript corrections were made by Dr. Khushwant Singh, Dr. Ashoo Grover, and Dr. Sanghamitra Pati. Dr. Ashoo Grover and Dr. Sanghamitra Pati also provided valuable suggestions during the study. All authors have reviewed and approved the final version of the manuscript.

Funding

No funding received.

Declarations

Competing interests

Authors do not have any competing interests.

Ethical approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

All authors have read and approved the final version of the manuscript.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-27634-1>.

Correspondence and requests for materials should be addressed to A.G.

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

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nd/4.0/>.

© The Author(s) 2025