Does Trends of Suicide Rate Cluster Geographically in India?

A Spatial Analysis Based on Method of Dissimilarity Coefficients for Longitudinal Dataset

Ajay Kr Bansal[*1], Anuradha Bansal[2], Narinder Kr Aggarwal[3], Anjur Tupil Kanan[4]

[1-] Department of Biostatistics & Medical Informatics, University College of Medical Sciences, University of Delhi, Dilshad Garden, Delhi 110 095, India

[2-] WZ-93, Sri Nagar, Shakur Basti, Delhi 110 034, India

[3-] Department of Forensic Medicine, University College of Medical Sciences, University of Delhi, Dilshad Garden, Delhi 110 095, India

[4-] Department of Community Medicine, University College of Medical Sciences, University of Delhi, Dilshad Garden, Delhi 110 095, India

akchbansal@yahoo.com

***Abstract*- The longitudinal data can always be represented in the form of polynomic trend on a graph which depicts the increasing or decreasing, linear or nonlinear behaviour of the data. Results on these trends without providing statistical analysis may be due to chance. In this exercise, the geographical trends of suicide rate from 1995 to 2009 are obtained for each State of India and are subjected to Average linkage between the groups method of Hierarchical Cluster Analysis using Bansal's method of dissimilarity coefficient for longitudinal dataset (Bansal & Sharma 2003, Bansal & Sharma 2009), which indicates that these geographical trends tend to cluster and suggest 13 clusters for major States and Union Territories of India. Limitation of the method is that the trends which are more towards linearity in nature form distinct clusters rather than non linear trends. In this regard, some modifications in the method are required, which remains a question of further research. These clusters based on states will definitely help in the implementation of policies specific to administrative areas to alleviate this burden on the society.**

Keywords- Geographical Time Trend; Cluster of Time Trends; Trends of Suicide Rate; Longitudinal DataTrends

I. INTRODUCTION

Suicide is the manifestation of hopeless state of mind and is the act of terminating one's own life which may occur for a number of reasons, including depression, shame, guilt, desperation, physical pain, emotional pressure, anxiety, financial difficulties, or other undesirable situations. The World Health Organization (http://en.wikipedia.org/wiki/Suicide#cite_note-1) noted that it was the thirteenth leading cause of death among teenagers and adults under 35. In India there were 127151 suicidal deaths out of total 357021 accidental deaths in 2009 which is nearly 36%(ADSI, 2009). This is really an alarming figure and is a matter of great concern. In 2009, among all the states and union territories of India, Puducherry ranked first (47.2), Sikkim ranked second (39.9), A&N Island ranked third (31.0), Punjab ranked 29 (3.2) and Manipur ranked the last (1.0)(Table 2.2, ADSI, 2009). Suicide rate ranges from 1.0 per lakh to to 47.2 per lakh. It is clear from the table that rate of suicide is more in coastal areas than in the hilly and plain areas(Table 2.2, ADSI, 2009). The studies from other countries also show the geographical variations in suicide rate (Liu 2009, Arán, Gispert, Puig, Freitas, Ribas & Puigdefàbregas, 2006, Yip, Wong, Cheung, Chan. & Beh 2009, Middleton, Sterne, Gunnell 2008, Brock, Baker, Griffiths, Jackson, Fegan & Marshall 2006). To study these vast geographical differences in time trends of suicide rate for major States of India, we have attempted, to obtain the groups of, and to classify, the major states with similar patterns over a period of 15 years from 1995 to 2009. This is done using the technique of cluster analysis by which a set of observations with similar characteristics is classified into non overlapping groups or sets. These groups are called clusters (Brock, Baker, Griffiths, Jackson, Fegan & Marshall 2006, Anderberg 1973, Copley & Lohens 1971, Devijver & Kittler 1982, Fukunaga 1972, Hartigan 1975, Jain & Dubes 1988, Zupan 1982). This technique minimizes the within group(cluster) variations and maximizes the between group variations (Garson 2012). Sometimes the cluster analysis is also called segmentation analysis, automatic classification, numerical taxonomy or typological analysis. Because of the natural tendency of human beings and heavenly bodies to form groups, this technique finds application in many fields of study, such as machine learning, data mining, pattern recognition, image analysis, bioinformatics, space sciences, earth sciences, engineering, life sciences, behavioral sciences, medicine, social sciences, etc.

In a longitudinal dataset, we can always record same type of information at different points of time in space. The main advantage of a longitudinal data set is that we can always measure a change and this can be extended in past as well as in future depending upon the trend in time on the existing values. But there is a disadvantage that a trend can never be

represented by a single value straight forwardly without doing any secondary calculations. In this exercise, we have used a method described in the next section, to obtain a single measure for each trend.

II. MATERIALS AND METHODOLOGY

The dataset of suicide rate from 1995 to 2009 for major states of India is taken from the Annual Reports of Accidental Deaths and Suicides in India, National Crime Records Bureau, Ministry of Home Affairs, Government of India (ADSI, 1995-2009).

The technique of cluster analysis requires a single measure for each trend (similarity or dissimilarity coefficient) to obtain the groups. In this exercise, for obtaining the clusters we need to calculate this single measure of dissimilarity for the time trends of suicide rate, for which the model proposed by Bansal (Bansal & Sharma 2003, Bansal & Sharma 2009) is employed. In this model each State was represented as n^{th} degree polynomial by fitting a curve with the help of curvilinear regression method. The total difference in rate of change from time t_1 to t_n (where n=2,3,4,…,N) for each State was obtained by summing the differences in velocity of trend between two adjacent time points i.e.

$$\sum_{n=2}^{N} [(f_p'(t_n) - f_p'(t_{n-1})]$$

where p=1,2,3,...,P and P are the number of states to be clustered. The distance of the trend from the base was calculated using the formula

$$sqrt[(x_p t_1 + x_p t_n + \sum_{z=1}^{Z-2} x_p t_{[1+(\frac{n-1}{z-1})*Z]})^2]$$

by dividing the trend objectively in to the optimum number of divisions Z. The Z was postulated as 3 if the (degree of the trend^2/number of time points) ≤3 and otherwise round (degree of the trend^2/number of time points). The dissimilarity coefficient (DC) is obtained by adding these two:

$$DC = \sum_{n=2}^{N} [(f_p'(t_n) - f_p'(t_{n-1})] + sqrt[(x_p t_1 + x_p t_n + \sum_{z=1}^{Z-2} x_p t_{[1+(\frac{n-1}{z-1})*Z]})^2]$$

In this exercise, the method of curvilinear regression is applied to obtain the polynomial equation of each trend of suicide rates for each state. MSExcel2007 is employed to draw the graphs of the observed and predicted trend of suicide rates. Through SPSSv17.0, the method of hierarchical cluster analysis (Average Linkage between Groups) is used to obtain the optimum number of clusters. SPSS (**S**tatistical Product and **S**ervice **S**olutions) is a Statistical Package offered by IBM to carryout all kinds of data analysis.

III. RESULTS

Observed and predicted polynomial trends were drawn in a single graph which is shown in Figure 1. For the brevity of the results, we have just shown only for the two states, so that the method followed is conveyed.

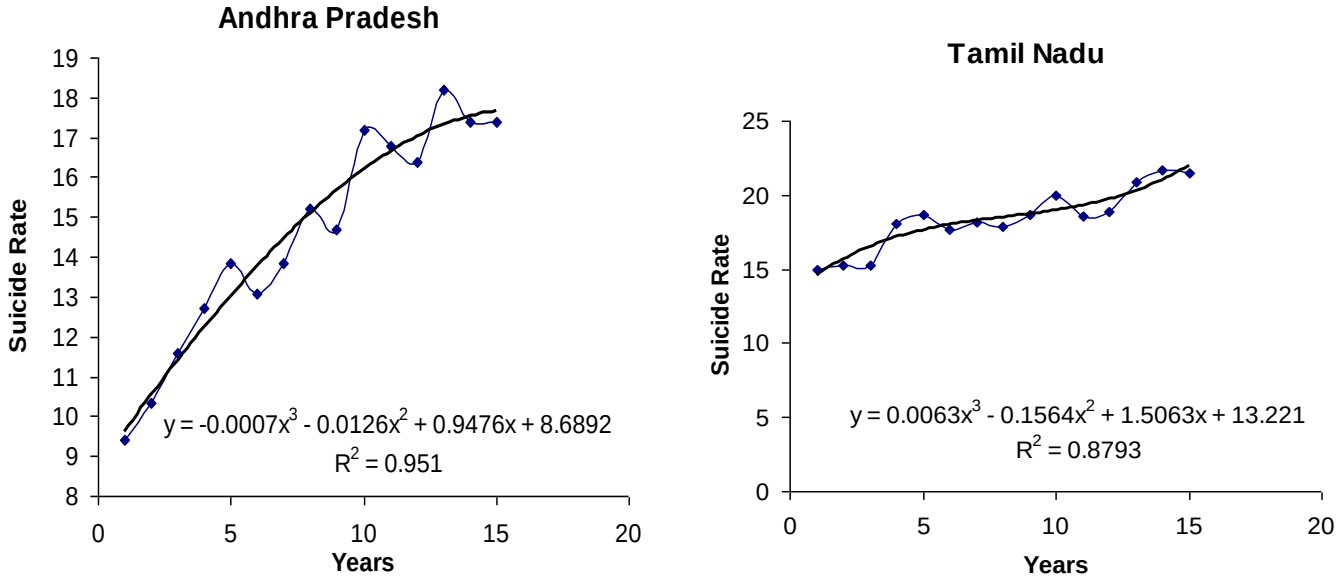


Figure 1 Showing the observed and predicted trend

Based on these polynomial trends obtained and the formula for calculating dissimilarity proposed above, the dissimilarity coefficient for each state is calculated, given in Table I.

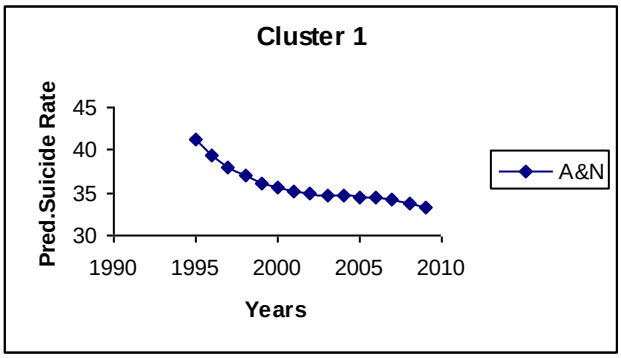
TABLE I DISSIMILARITY COEFFICIENT(DC) FOR STATES & UNION TERRITORIES OF INDIA

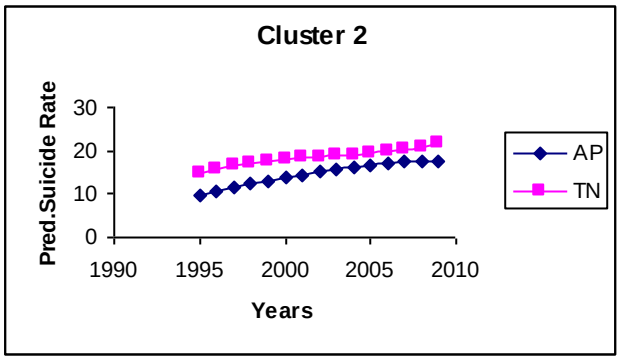
State	DC	State	DC
ANDAMAN & NICOBAR(A&N)	60.5	KERALA(KL)	52.7
ANDHRA PRADESH(AP)	40.8	MANIPUR(MN)	2.4
ARUNACHAL PRADESH(ArnP)	21.7	MEGHALAYA(MG)	6.9
ASSAM(AS)	19.2	MAHARASTRA(MH)	28.9
BIHAR(BH)	1.9	MADHYA PRADESH(MP)	28.6
CHANDIGARH(CHD)	18.4	MIZORAM(MZ)	11.0
DAMAN & DIU(D&D)	20.9	ORISSA(OR)	26.7
DADAR & NAGAR HAVELI(D&N)	50.4	PUNJAB(PJ)	6.9
DELHI(DL)	15.6	PUDDUCHERRY(PD)	80.9
GUJARAT(GJ)	21.0	RAJASTHAN(RJ)	15.1
GOA	31.9	SIKKIM(SK)	79.7
HIMACHAL PRADESH(HP)	17.7	TAMIL NADU(TN)	46.7
HARYANA(HR)	22.7	TRIPURA(TP)	47.2
JAMMU & KASHMIR(J&K)	5.9	UTTAR PRADESH(UP)	4.1
KARNATAKA(KT)	44.1	WEST BENGAL(WB)	34.5

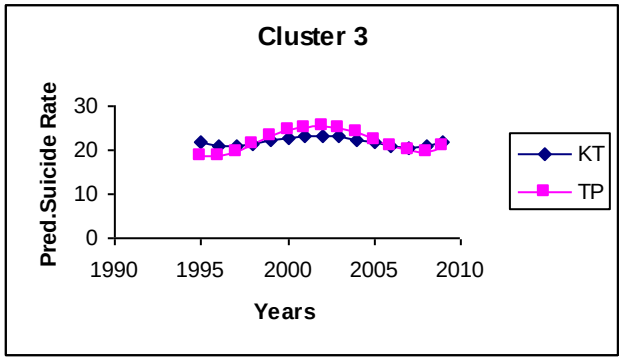
After obtaining the dissimilarity coefficient as above, with the help of SPSSv17.0, *Average Linkage between the groups* method of Hierarchical clustering technique was applied and agglomeration schedule is obtained(SPSSv17.0), given as Table II. The optimum number of clusters are obtained by minimizing the within cluster distances and maximizing the between cluster distances. In Table II it is represented under the head "coefficient", where, if there is a sudden change in this coefficient value after a stage, then that particular stage gives the optimum number of clusters(SPSSv17.0). If we see table 2 thoroughly, we find that the sudden change happens after the 18^{th} stage where the number of clusters are 13 (at Stage 1, 30 clusters are there, at Stage 2, 29 clusters, at Stage 3, 28 clusters and so on, at the last Stage there are 2 clusters left). These 13 clusters are shown in Figure 2.

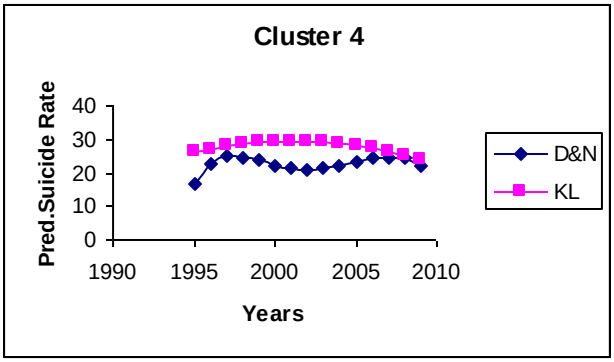
TABLE II AGGLOMERATION SCHEDULE FOR AVERAGE LINKAGE BETWEEN THE GROUPS METHOD

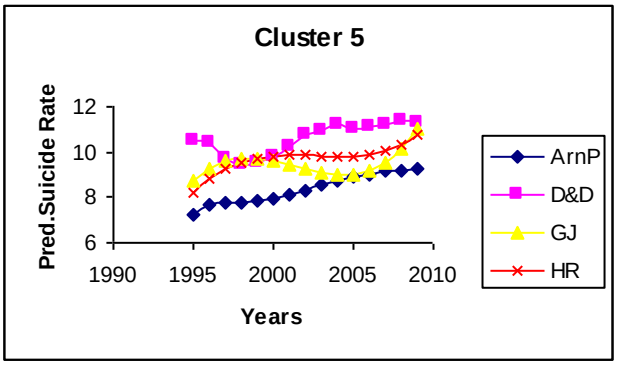
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	18	23	.000	0	0	9
2	7	10	.010	0	0	8
3	19	20	.090	0	0	14
4	27	28	.250	0	0	17
5	9	25	.250	0	0	18
6	5	17	.250	0	0	13
7	6	12	.490	0	0	11
8	3	7	.565	0	2	12
9	14	18	1.000	0	1	19
10	24	26	1.440	0	0	28
11	4	6	1.445	0	7	18
12	3	13	2.377	8	0	20
13	5	29	3.865	6	0	19
14	19	22	4.225	3	0	22
15	8	16	5.290	0	0	24
16	11	30	6.760	0	0	22
17	15	27	8.185	0	4	21
18	4	9	**9.945**	11	5	20
19	5	14	15.297	13	9	23
20	3	4	22.190	12	18	25
21	2	15	28.887	0	17	24
22	11	19	28.990	16	14	25
23	5	21	44.002	19	0	27
24	2	8	54.700	21	15	26
25	3	11	134.395	20	22	27
26	1	2	197.872	0	24	28
27	3	5	348.353	25	23	29
28	1	24	1020.799	26	10	29
29	1	3	1785.393	28	27	0

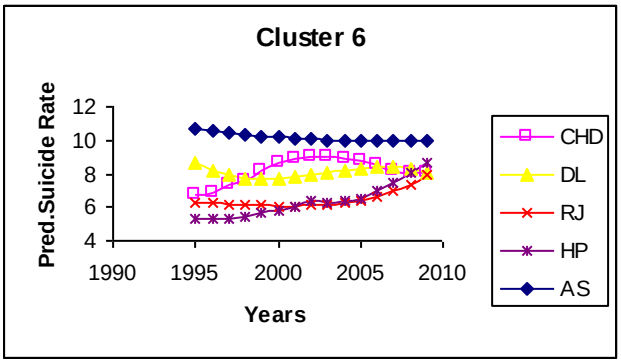


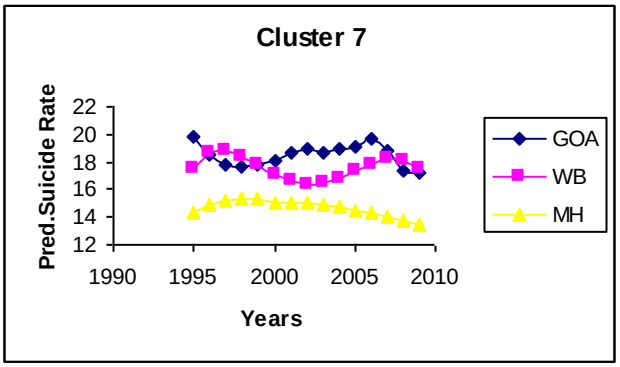


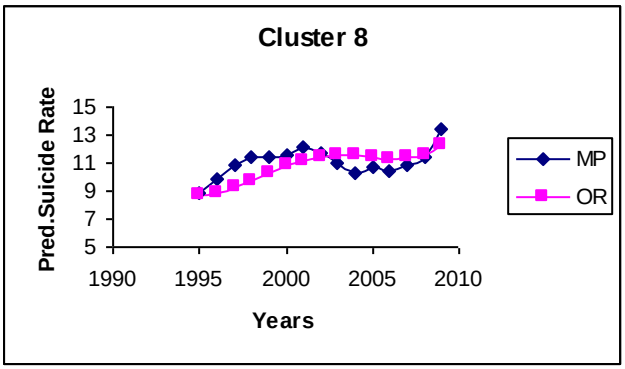

Cluster 3
Pred.Suicide Rate
30
20
10
0
1990
1995
2000
2005
2010
Years
KT
TP

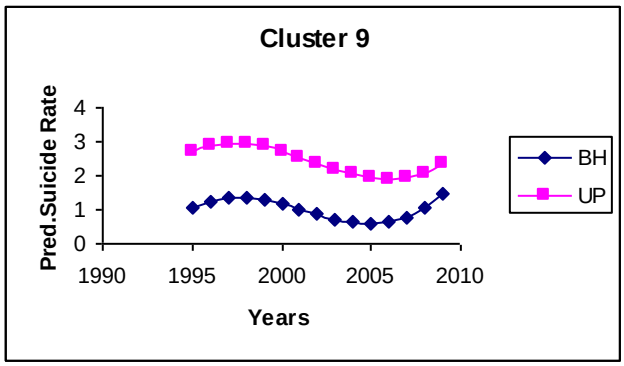

Cluster 4
Pred.Suicide Rate
40
30
20
10
0
1990
1995
2000
2005
2010
Years
D&N
KL

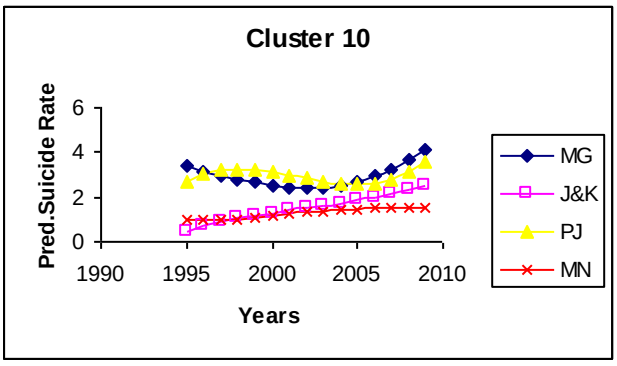

Cluster 5
Pred.Suicide Rate
12
10
8
6
1990
1995
2000
2005
2010
Years
ArnP
D&D
GJ
HR


Cluster 6
Pred.Suicide Rate
12
10
8
6
4
1990
1995
2000
2005
2010
Years
CHD
DL
RJ
HP
AS


Cluster 7
Pred.Suicide Rate
22
20
18
16
14
12
1990
1995
2000
2005
2010
Years
GOA
WB
MH


Cluster 8
Pred.Suicide Rate
15
13
11
9
7
5
1990
1995
2000
2005
2010
Years
MP
OR


Cluster 9
Pred.Suicide Rate
4
3
2
1
0
1990
1995
2000
2005
2010
Years
BH
UP


Cluster 10
Pred.Suicide Rate
6
4
2
0
1990
1995
2000
2005
2010
Years
MG
J&K
PJ
MN

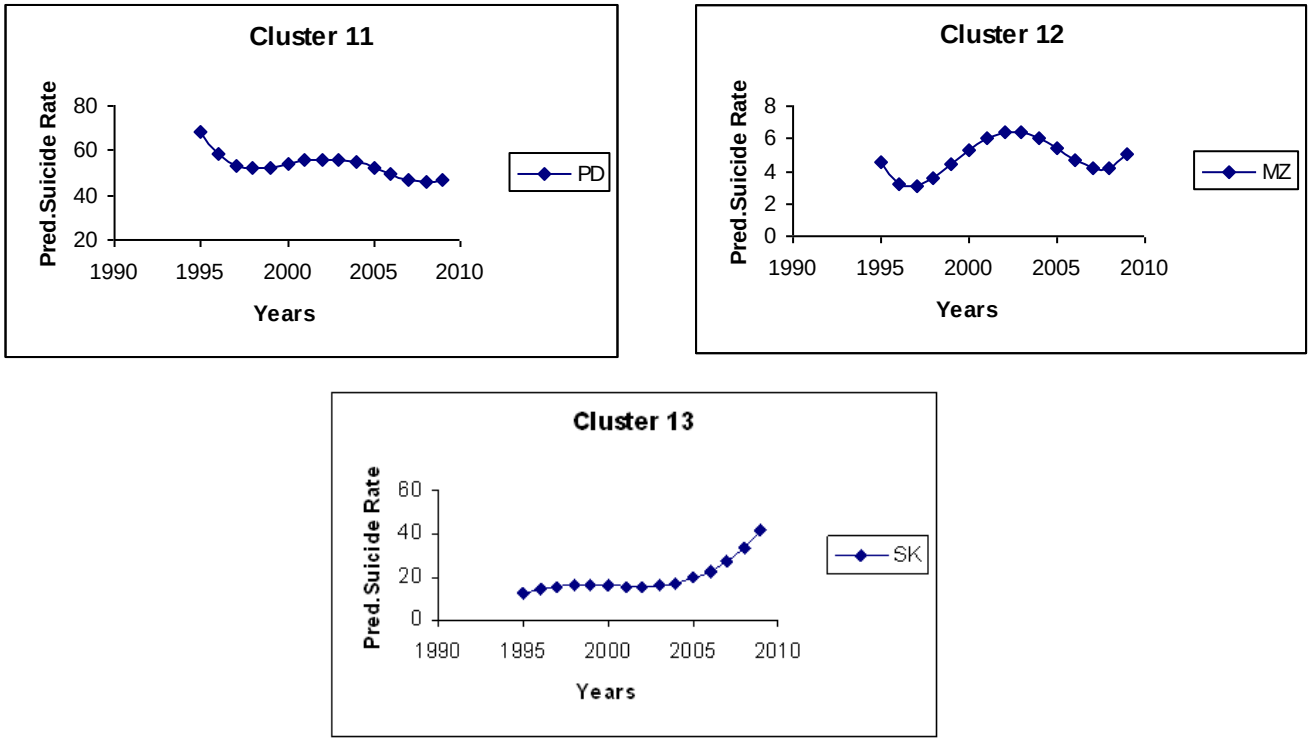


Figure 2 Showing the different clusters

IV. DISCUSSION

Variation in suicide rate may be due to many different influences. They could be personal, social, relationship or environmental. A report from 20 OECD (Organization for Economic Co-operation and Development) countries investigated the hypothesis that sun shine is highly correlated with suicidal tendency (Petridou, Papadopoulos & Frangakis, 2002). Mckenzie N et al. (Mckenzie, Landau, Kapur, Meehan, Robinson, Bickley, Parsons & Appleby, 2005) showed that imitative suicide occurs among people with mental illness and may account for about 10% of suicides. Baldassin S, et al. (Baldassin, Ferraz Alves, Andrade, Antonio & Martins, 2008) studied the characteristics of depressive symptoms among 481 medical students during medical education and training. Bossarte RM et al. (Bossarte, Simon & Swahn, 2008) developed a behavioral typology based on self-reports of suicidal behaviors, physical violence, and psychological abuse using a sample of dating adolescents from a high-risk school district. They identified five clusters of behaviors among the 1,653 students who reported being abusive or violent in the past year. Vamik A et al. (Vamik, Kolves, Van der Feltz-Cornelis, Marusic, Oskarsson, Palmer, Reisch, Scheerder, Arensman, Aromaa, Giupponi, Gusmao, et. al., 2008) obtained a gender specific difference on suicide method and found that countries divided in five main groups based on Ward's method of hierarchial clustering did not yield clear results.

Our study is based on population trends rather than individuals. It may be a limitation of this study in the sense that it may not be used to analyze individual's behavior but talk of a trend over a period for the entire administrative boundary i.e. state.

Geographical trends are always a matter of interest for the researchers. In this study we have tried to obtain the clusters of the time trends to show the geographical and temporal differences. In Figure 2, Clusters 1, 11, 12 & 13 are the single states alone, whose behaviour over the time is entirely different than the other states of India. In Cluster 2, AP & TN and KT & TP in Cluster 3 follows almost a same pattern. Cluster 4 indicates that these two states initially show an increasing trend and in the later half, a decreasing pattern and form a separate group. MP & OR in cluster 8 and BH & UP in Cluster 9 also show the same pattern and form a separate group. ArnP & D&D follow slightly a different pattern but are in close proximity to GJ & HR, which is clear from the Cluster 5. In Cluster 6, CHD, DL & AS, in Cluster 7 MH and in Cluster 10, PJ & MG are the odd one out. It may be due to the limitations of the method followed which may require few modifications as a part of further research. But our primary objective was to show that geographical clusters of these trends do occur or not using the method of dissimilarity coefficients for longitudinal dataset and this exercise indicates an evidence of spatial clustering. The results also demonstrate that if the analysis can be conducted with some modifications, which are yet to be investigated, then a different picture can be obtained.

V. CONCLUSIONS

These clusters can be used to devise the similar policies for the States in the same cluster and another policy for the States in other cluster based on the characteristics of geographical areas to improve the coordination among the policy makers, authorities and social agencies to alleviate this social problem and to reduce the suicide burden in India to have more productive lives.

ACKNOWLEDGEMENT

We are thankful to the Department of Biostatistics and Medical Informatics of Delhi University College of Medical Sciences for providing necessary software support in carrying out the data analysis and preparation of the manuscript.

REFERENCES

[1] http://en.wikipedia.org/wiki/Suicide#cite_note-1 accessed on 08/12/2011.

[2] Accidental Deaths and Suicide in India: Snapshots (2009). National Crime Records Bureau, Ministry of Home Affairs, Government of India. pvi-viii.

[3] Accidental Deaths and Suicide in India Table 2.2 (2009). National Crime Records Bureau, Ministry of Home Affairs, Government of India.

[4] Liu H.L.(2009). Epidemiologic characteristics and trends of fatal suicides among the elderly in Taiwan. Suicide Life Threat Behav, 39(1), 103-13.

[5] Arán Barés M., Gispert R., Puig X., Freitas A., Ribas G., Puigdefàbregas A. (2006). Geographical distribution and time trends of suicide mortality in Catalonia and Spain 1986-2002. Gac Sanit, 20(6), 473-80.

[6] Yip P.S., Wong P.W., Cheung Y.T., Chan K.S., Beh S.L. (2009). An empirical study of characteristics and types of homicide-suicides in Hong Kong, 1989-2005. J Affect Disord, 112(1-3), 184-92.

[7] Middleton N., Sterne J.A., Gunnell D.J. (2008). An atlas of suicide mortality: England and Wales, 1988-1994. Health Place, 14(3), 492-506.

[8] Brock A., Baker A., Griffiths C., Jackson G., Fegan G., Marshall D. (2006). Suicide trends and geographical variations in the United Kingdom, 1991-2004. Health Stat Q, 31, 6-22.

[9] Anderberg M.R. (1973). Cluster analysis for applications. Academic Press, New York.

[10] Copley W.W., Lohens P.R. (1971). Multivariate Data Analysis. Wiley.

[11] Devijver P.A., Kittler J. (1982).Pattern Recognition: A Statistical Approach. Prentice Hall, Englewood cliffs.

[12] Fukunaga K. (1972). Introduction to Statistical Pattern Recognition. Academic Press.

[13] Hartigan J. (1975). Clustering Algorithms. Wiley.

[14] Jain A.K., Dubes R.C. (1988). Algoritm for Clustering Data. Prentice Hall Eaglewood Cliffs.

[15] Zupan J. (1982). Clustering of large data sets. New York, John Wiley and Sons.

[16] Accidental Deaths and Suicide in India (1995-2009). National Crime Records Bureau, Ministry of Home Affairs, Government of India.

[17] Bansal A.K., Sharma S.D. (2003). A model for clustering of longitudinal data sets of infant mortality rates in India. Med Sci Monit, 9(4), 1-6.

[18] Bansal A.K., Sharma S.D. (2009). Reliability of the model for clustering of longitudinal data sets of infant mortality rate in India, COBRA Preprint Series. Article 57.
http://biostats.bepress.com/cobra/ps/art57.

[19] SPSS Base Ver 17.0, Application Guide. Statistical Software for Social Sciences, SPSS Inc.

[20] Petridou E., Papadopoulos F.C., Frangakis E. (2002). A role of sunshine in the triggering of suicide. Epidemiology, 13(1), 106-9.

[21] Mckenzie N., Landau S., Kapur N., Meehan J., Robinson J., Bickley H., Parsons R. and Appleby L. (2005). Clustering of suicide among people with mental illness, British Journal of Psychiatry, 187, 476-480.

[22] Baldassin S., Ferraz Alves T.C.T., Andrade A.G., Antonio L., Martins N. (2008). The characteristics of depressive symptoms in medical students during medical education and training: A cross-sectional study. BMC Medical Education, 8:60.

[23] Bossarte R.M., Simon T.R., Swahn M.H. (2008). Clustering of adolescent dating violence, peer violence, and suicidal behavior. J Interpers Violence, 23(6), 815-33.

[24] Vamik A. , Kolves K., Van der Feltz-Cornelis C.M., Marusic A., Oskarsson H., Palmer A., Reisch T., Scheerder G., Arensman E., Aromaa E., Giupponi G., Gusmao R. et. al. (2008). Suicide methods in Europe: a gender specific analysis of countries participating in the "European Alliance Against Depression". J Epidemiol Community Health, 62:545-551.