Demography India (2017) ISSN: 0970-454X
Vol. 46, Issue: 2, pp: 44-52

Research Article

Estimation of Population for Population Based Cancer Registry Areas in Developing Countries: A Comparative Appraisal

Jang Bahadur Prasad[1] & Murali Dhar[2,*]

Abstract

PBCRs indispensably require population of the PBCR area by calendar year, age and sex. In India, Census provides population figures once in every ten years and traditionally, exponential growth model (EGM) was used for estimating population for inter-census years and projecting beyond latest Census. Later on however, Takiar and Shobana (2009) criticized the existing method and published Difference Distribution Method (DDM) claiming it to provide better estimate of age-structure. Although DDM has been accepted by the Indian PBCRs, its comprehensive evaluation has been missing. Therefore, objective of this study was to conceptually and empirically evaluate the superiority of DDM method over existing exponential method. Population for the years 1991 and 2001, were obtained from Indian Census. Using these, age and sex wise population for 2011 was estimated using EGM and DDM methods and compared with that obtained from 2011 Census. Different rates using three sets of population were also compared. Sum of absolute errors in projected populations was very close to each other. Age specific cancer incidence rates according to both methods were similar to each other and also close to the true value. Comparison of age standardized rate also did not show considerable difference between the two methods. EGM has sound scientific base whereas DDM possesses many potential fallacies. Hence, PBCRs in India should continue exponential growth model till another model is found to fit better. However, in view of rapid demographic transition in developing countries, a regular review of population growth is required.

Introduction

Cancer is universally perceived as a frightening disease, its risk is higher in high income countries than low and middle income countries (WHO, 2004). However, it has undoubtedly emerged as an important public health problem in developing world also. Estimated global cancer burden for 2012 was about 14 million new cases, 8 million cancer deaths per year and five-year prevalence of 33 million among individuals above the age of fifteen years (Ferlay et al., 2013). About half of this prevalence, more than half of incidence and about two third of deaths are from the developing world.

Population Based Cancer Registries (PBCRs) have been accomplished in many developed and developing countries throughout the world. The primary objective of such registries is to improve efforts towards cancer control by providing information not only on magnitude and patterns of cancer in various segments of population but also to study the various etiological and epidemiological factors as well as for evaluation and planning of health services. India has a long tradition of cancer registration and research in cancer epidemiology. It was in 1964, when Indian Cancer Society, Mumbai established a population based cancer registry covering the population of Mumbai. Since then cancer registration in India has expanded to about 30 PBCRs besides few hospitals based cancer registries and other cancer control/research related activities.

One of the essential requirements of PBCRs is calendar year wise population of the PBCR area by quinquennial age groups and sex. In India, Census has been providing population figures once in every ten years starting from the year 1881, the latest in the year 2011. Therefore, to provide

[1]Ph. D. Fellow, International Institute for Population Sciences (IIPS), Mumbai.
[2]Associate Professor, Department of Population Policies and Programmes, IIPS, Mumbai.
*Corresponding author: Dr. Murali Dhar Email: m.dhar@iips.net.

various rates for given year for cancer incidence data, lying between any two Census years or beyond the latest Census, the estimation/projection of population is required. For estimation and projection of population, there are many methods available in the literature ranging from simplest linear inter/extrapolation to the fitting of complex models. In principle, none of the methods can be considered as standard method applicable for all the populations and for all the time periods. Suitability of a particular method can be only population and time specific. For estimation and projection of population for PBCR areas in India, exponential growth model has been in from beginning. Subsequently, Takiar and Shobana (2009) criticized the existing method and proposed a new method, namely, Difference Distribution Method (DDM) claiming it to provide a better estimate of age structure of the population (Takiar and Shobana, 2009). Although the method has been accepted by the Indian cancer registration authorities, its comprehensive evaluation has been missing. Therefore, the objective of this study was to conceptually and empirically evaluate the superiority of DDM method over existing exponential method.

Methods and Materials

Exponential growth method (EGM)

Let us assume that P_0 and P_t are known population at time 0 and t respectively, with time duration between being t years. The exponential growth rate r is estimated as following.

$$r = \frac{1}{t} ln(\frac{P_t}{P_0})$$

Population at time x (P_x) can be interpolated (if $x < t$)) or extrapolated (if $x > t$) as following.

$$P_x = P_0 e^{rx}$$

Projection of population using EGM involves following three steps:

i) Projection of overall (total for all ages) population using above formula
ii) Projection of population using above formula within each of five-year age groups.
iii) Proration of difference between total projected and some of all age specific projected population.

Difference distribution method (DDM)

i) Calculate proportional age distribution of difference between P_0 and P_t (p_{0t}).
ii) Project overall (total for all ages) population at time x (P_x) using EGM method.
iii) Calculate difference in total population at time 0 and x (D_{0x}) as following

$$D_{0x} = P_x - P_0$$

iv) Calculate projected population at time x in the age group i (Px_i) as following

$$P_{xi} = P_{0i} + D_{0x} * p_{0t}$$

Conceptual difference between the two methods

As far as the projection of total population is concerned, DDM uses EGM method, so the two are same. The difference lies in the way age distribution of the total population is obtained. EGM is based on the assumption of exponential growth model within each age group whereas DDM assumes that the proportional age distribution of the difference between total populations at two time points is constant over time. Conceptually, DDM seems to be similar to ratio method (Newell, 1990; Agarwal, 2000) and replica of proportionate method (Shryock and Siegel, 1971; Agarwala, 2000) available in demographic literature for many decades now and briefly described below. However, these methods were developed and proposed for estimating the population for smaller area within a larger area with population for the given year being available.

Ratio method

The ratio method is based on the ratio of the smaller area to that of the larger area. This method is based on the assumption that the observed trends in the ratios of a smaller area population to that of

the larger population will continue in future. It is highly practical in projection of sub-national population in relation to projected national population. The formula used for projection of future population of area '*i*' at any time'*t*' (P_i^t) is as

$$P_i^t = R_i^t * P^t$$

Where:

R_i^t =ratio of future date population of i^{th} area to the population of larger area of which it is a part (this is calculated through the ratio of population of smaller area to the larger area at various points of time in the past)

P^t = future population of the larger area

Proportionate method

This method is similar to ratio method proposed by Shroyock and Siegal (1971) to estimate sub-national or subgroups population when larger area population is already projected (Shroyock and Siegal, 1971). Registrar General of India used this procedure (named as water II method) to project the district wise population in India (Registrar General, 1987). This also assumes that distribution of change at larger level (state) to small level (different district) for recent period would continue in the future. Registrar general had used the following equation for the projection:

$$P(i,81) = P(i,71) + \frac{P(i,71)-P(i,61)}{P(71)-P(61)} \times [P(81) - P(71)]$$

Where, P(i, 81), P(i, 71) and P(i, 61) are the population of i^{th} district in 1981, 1971 and 1961 census. P(81), P(71) P(61) are the state population from respective census to which district belongs.

Empirical validity

To test the validity of two methods under consideration, namely, EGM and DDM, we used Census population for the years 1991 and 2001 and projected age and sex wise population for the year 2011 and compared the same with that provided by Census. For the objective comparison of overall age structure, we summarized the differences into sum of absolute over/under estimate. In addition, to assess the impact of the differences on the incidence rates, we calculated crude, age specific and age adjusted cancer incidence rates using the population projected by EGM and DDM and that provided by Census and compared the three.

Data on number of incident cases by five-year age group and sex were obtained from consolidated report of population based cancer registries (PBCR) covering data for different time periods during 2008-2010 (National Cancer Registry Programme, 2013). For the standardization of rates, we collected age wise distribution from world standard population (Doll et al., 1966). The analysis was carried out for five long standing urban PBCRs in India, namely, Chennai, Mumbai, Bangalore, Bhopal and Delhi, however, detailed age structure comparison was limited to PBCR Chennai.

Results

Comparison of age structure (Tables 1a & 1b)

Population of Chennai PBCR area was projected to be 2,480,193 in 2011 against a true value of 2,335,844 according to 2011 Census, an over projection of 6%. However, this over projection does not have any role in the comparison of two methods, as, these methods differ only in the age distribution of the population. Regarding age distribution, the over/under projection of the population in different five-year age groups was within 2%. Populations in the age groups, within 10 years, 40-44 years and 55-64 years were under projected and in other age groups, it was over projected by both methods. Barring rare exceptions, the over/under projection of population in different age groups by two methods was similar. The sum of absolute errors in projected populations was 8.8 percent units by EGM and 8.9 percent units by DDM, very close to each other.

Overall absolute error sums varied between 8 to 10 percent units in all the PBCRs under study in both sexes with the exception of Delhi, Females, where it was 11%. Comparison between the two methods showed that EGM gave better assessment of age structure than DDM for Chennai, Delhi and Bhopal, whereas, for Mumbai and Bangalore, DDM seem to have slight edge. The differences between the two methods however appeared to be negligible.

Table 1a: Estimation of population by five-year age group using exponential and different distribution methods and their comparison, Chennai, Male-2011.

Age Groups	1991	2001	Proportion (%) by age 2011			Over/under estimate		Absolute over/under estimate	
			Census	EGM	DDM	Exp.	DDM	EGM	DDM
0-4	167407	156443	7.23	5.81	5.81	-1.42	-1.42	1.42	1.42
5-9	191025	174686	7.07	6.35	6.31	-0.72	-0.76	0.72	0.76
10-14	195524	200575	7.63	8.17	8.31	0.54	0.68	0.54	0.68
15-19	198581	216803	7.99	9.40	9.56	1.41	1.57	1.41	1.57
20-24	216844	231618	9.01	9.83	10.00	0.82	0.99	0.82	0.99
25-29	200616	226999	9.83	10.20	10.34	0.37	0.51	0.37	0.51
30-34	168146	197627	9.11	9.23	9.30	0.12	0.19	0.12	0.19
35-39	151786	181515	8.57	8.62	8.66	0.06	0.09	0.06	0.09
40-44	125427	148448	7.58	6.98	7.02	-0.60	-0.56	0.60	0.56
45-49	98541	130115	6.65	6.82	6.67	0.17	0.02	0.17	0.02
50-54	81152	104987	5.39	5.39	5.31	0.01	-0.08	0.01	0.08
55-59	62384	70861	4.42	3.20	3.24	-1.22	-1.18	1.22	1.18
60-64	52094	64215	3.59	3.14	3.14	-0.45	-0.46	0.45	0.46
65-69	31846	44587	2.20	2.48	2.37	0.27	0.17	0.27	0.17
70-74	22924	32622	1.73	1.84	1.75	0.12	0.02	0.12	0.02
75-79	10515	18073	1.03	1.23	1.07	0.20	0.04	0.20	0.04
80+	11467	19365	0.96	1.30	1.14	0.34	0.17	0.34	0.17
All ages	1986278	2219539	100.00	100.00	100.00	-	-	-	-

Note: EGM: Exponential Growth Model, DDM: Difference distribution method

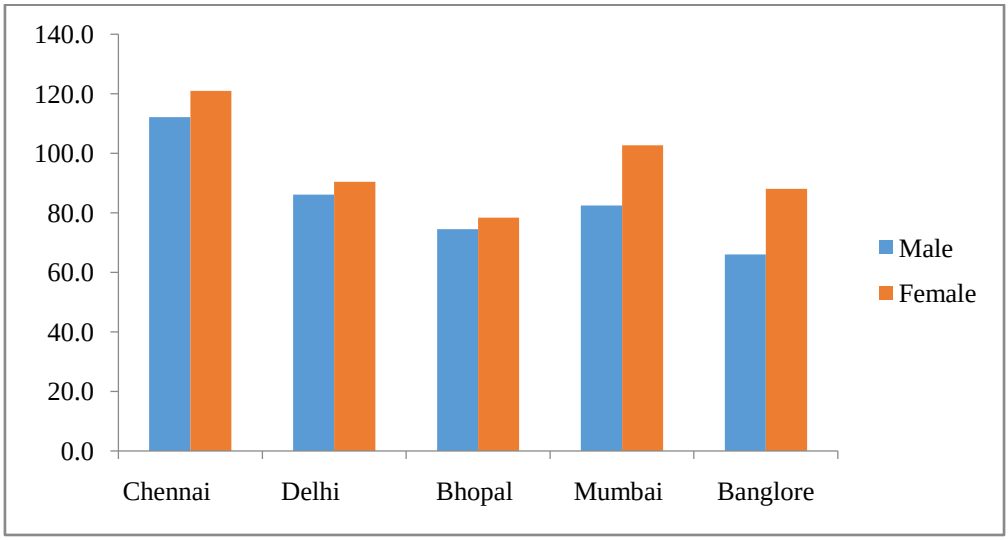
Table 1b: Absolute over/under estimates in age specific population by Exponential method and Difference distribution methods

Registry	Sum of absolute over/under estimate (%)	
	Exponential Method	DDM
Male		
Chennai (2009)	8.83	8.90
Delhi (2008-09)	8.25	8.32
Bhopal (2009-10)	9.57	9.69
Mumbai (2009-10)	9.51	9.23
Bangalore (2008-09)	9.57	9.51
Female		
Chennai (2009)	7.99	8.53
Delhi (2008-09)	11.18	11.14
Bhopal (2009-10)	9.47	9.34
Mumbai (2009-10)	7.13	6.94
Bangalore (2008-09)	9.65	8.58

Crude cancer incidence rate in PBCRs under study (Figure 1)

Cancer incidence rate per hundred thousand person years varied between about 65 and 120 with a highest rate in Chennai in both sexes and lowest in Bangalore in males and in Bhopal in females. Cancer occurrence was more in females then in males with the differences being larger in Mumbai and Bangalore. As far as the comparison of crude rate according to the two methods is concerned, both are same as crude rate is based on total population only, independent of age structure.

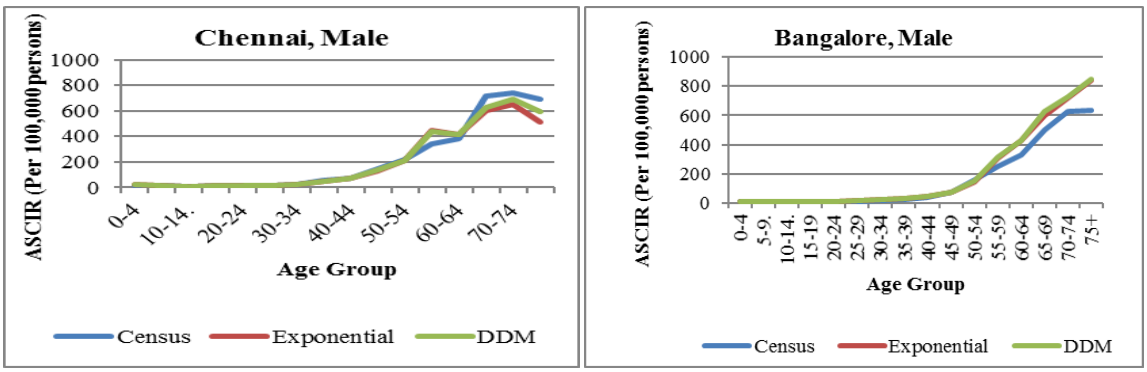
Figure 1: Cancer incidence rate (per 100,000 people) among selected PBCRs under study, 2011



Comparison of age specific cancer incidence rates

Figures 2a and 2b present respectively for males and females, age specific cancer incidence rates (ASCIR) calculated using population provided by census and that estimated by EGM and DDM methods for five selected Indian PBCRs. Age specific risk of occurrence of cancer is very small in the age groups within 30 years. Therefore, the issue of difference in ASCIRs according to different populations does not arise. In the quinquennial age groups between 30 and 55, ASCIRs according to both methods were similar among each other and also close to the true value. Differences in ASCIRs according to different populations become noticeable in the age groups beyond 55 years. In age groups 55-59 and 60-64, ASCIR based on populations of two methods were by and large close to each other but away from true value. There were appearing to be really considerable differences in ASCIRs according to different populations in the age groups beyond the age of 65 years. However, none of the two methods appeared to give consistently better estimate than the other. In some cases, EGM seems to result in ASCIRs closer to true value and in other cases DDM.

Figure 2a: Age specific cancer incidence rates per 100,000 person years using population according to census, exponential and difference distribution methods, Males



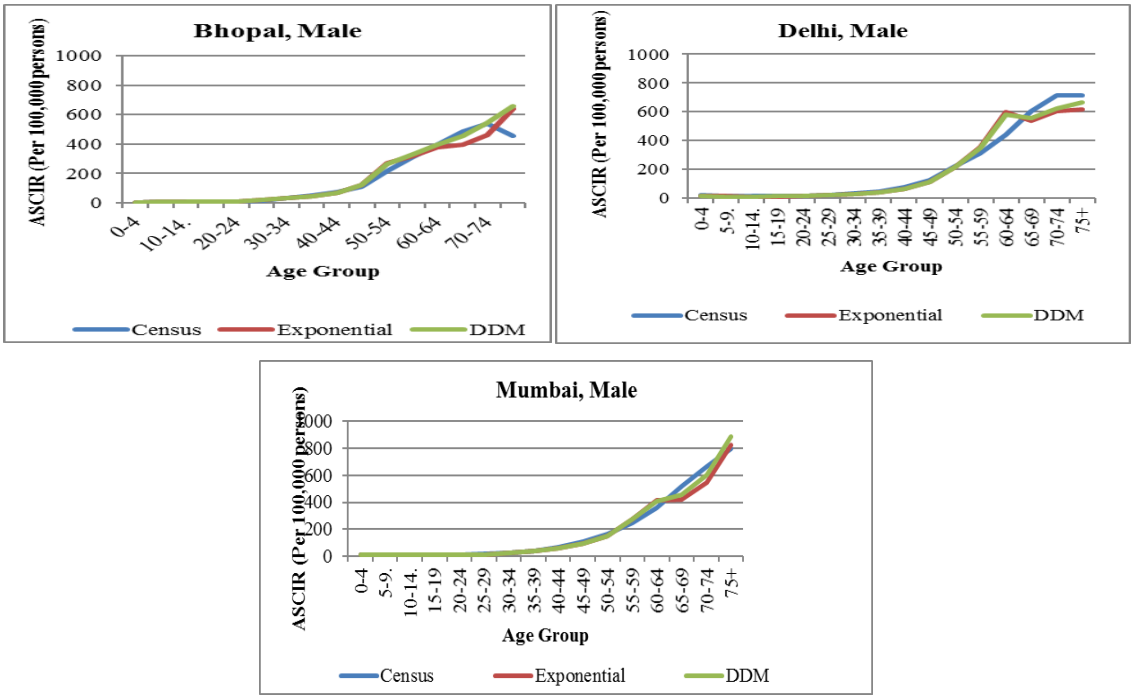
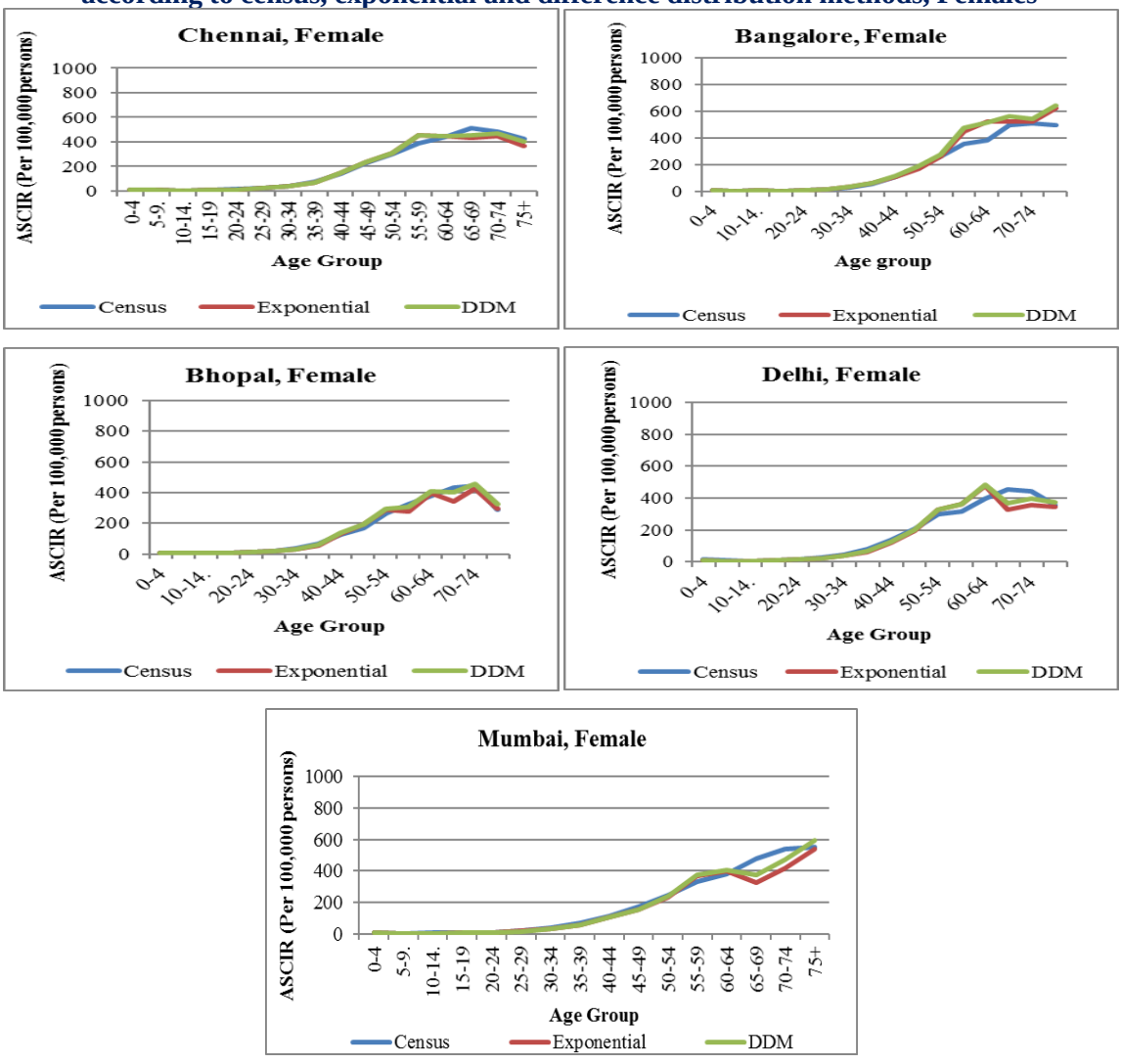


Figure 2b: Age specific cancer incidence rates per 100,000 person years using population according to census, exponential and difference distribution methods, Females



Comparison of age adjusted cancer incidence rates (AACIR)

Table 2 provides age adjusted cancer incidence rates for the year 2011 for five PBCRs using the true population and the ASCIRs arrived at using age specific populations according to two methods. World Standard Population age structure was used for adjusting the rates. Over/under estimation of AACIR was negligible (< 4%) for all the registries except Bangalore. For Bangalore, it was 10 to 12% in males and 6 to 8% in females. AACIR based on EGM was closure to the true AACIR compared to that of based on DDM in case of Bhopal and Bangalore. In case of other PBCRs, DDM had a slight edge over EGM.

Table 2: Age adjusted cancer incidence rates (AACIR) per 100,000 person years using the population according to census, exponential and difference distribution methods and over/under estimation according to two methods.

Registries	AACIR with different populations			Over/Under estimation percentage	
	Census	Exponential	DDM	Exponential	DDM
Males					
Chennai (2009)	116.20	110.96	114.05	-4.51	-1.85
Delhi (2008-09)	116.43	112.67	113.88	-3.24	-2.20
Bhopal (2009-10)	95.93	96.67	101.43	0.78	5.74
Mumbai (2009-10)	97.23	92.30	95.45	-5.08	-1.84
Bangalore (2008-09)	85.90	101.60	103.68	18.27	20.70
Females					
Chennai (2009)	118.62	116.97	119.27	-1.39	0.55
Delhi (2008-09)	109.04	106.04	109.18	-2.75	0.13
Bhopal (2009-10)	97.45	95.27	100.72	-2.23	3.36
Mumbai (2009-10)	106.20	97.38	101.99	-8.31	-3.96
Bangalore (2008-09)	105.85	119.79	124.14	13.16	17.28

Discussion

Cancer incidence and prevalence rates are required not only by epidemiologist for research purposes but also by policy planners and administrators for prioritizing the resources as well as planning and effective implementation of policies. Population Based Cancer Registries (PBCRs) are the only source of the cancer incidence rates. However, they indispensably require year wise midyear population of the registry areas. Technically, a rate is derived from two numbers; number of cases (numerator) and the exposed population (denominator) and both are equally important in fixing the accuracy of a rate. There is always possibility of under counting of numerator due to missing some of the cases or over counting due to duplicates and this has direct proportional impact on the rate. Similarly, under/over assessment of exposed population has inversely proportional impact on the rate. Thus the numerator and the denominator both have exactly same quantitative potential to affect the accuracy of a rate. There are a lot of efforts on arriving at the numerator as close to the true value as possible. Assessment of denominator however, unfortunately, does not get its due attention and emphasis with the exception of some special efforts in 1990's, when International Institute for Population Sciences, Mumbai, India, an institution specialized in demographic research was involved. An erroneous assessment of the denominator has long lasting implications in the assessment of pattern and trend of cancer in a country.

By and large, census is the only source of population in developing countries including India; however, the same is conducted at a gap of 10 years. Under the circumstances, for obtaining midyear population for the PBCR areas, traditionally it was established and therefore perceived that these populations follow exponential rate of growth within each of the quinquennial age groups. Accordingly, PBCR personnel in India used to apply exponential rate of growth to estimate the 5-year age group and sex wise annual population. Somewhat recently, Takiar and Shobana (2009) come out with DDM claiming it to provide better estimate of the age wise distribution of the population (Takiar

and Shobana, 2009). Although the new method has been accepted by NCRP, its comprehensive evaluation for superiority has been lacking. Therefore, present study was undertaken to examine the validity and superiority of DDM method.

The article by Takiar and Shobana (2009) seem to possess many technical fallacies related to the method for projection of population.

1. Authors seem to have messed up between geometric and exponential models. On one hand, they talk about exponential growth method and on the other, they used geometric growth method formula as if both are same.
2. Authors have themselves imagined individual and overall exponential methods which do not exist. Standard methods for the estimation of population using any growth model (linear, geometric, and exponential) consist of the three steps enumerated in the methods and materials section. Moreover, while comparing so called individual and overall EGM, they have exaggerated the differences of less than half percent as big differences.
3. Although third step is part of the estimation of population applying a particular growth model; authors of the article have referred this step as adjustment.
4. Another scope for investigator bias may potentially be in the selection of registry for the application of the two methods. Authors have not given the rationale for selecting Chennai.
5. Third objective of their study was to show that their method scores over the existing methods. This indicates that they may not have done an unbiased empirical comparison of two methods. Their thrust towards fulfilling this objective may have resulted in a biased selection of registry for empirical application.

Present study has shown that both the methods give very similar age distribution. Comparison of age standardized rate also does not show any considerable difference between the two methods. DDM is not a population estimation method. It is just a method for arriving at the age distribution of the total population. This method lacks scientific ground and it doesn't give the age distribution better in accuracy than the one obtained using standard (EGM) method. Authors of DDM method failed to use a gold standard like we did in the present study. They advocated the superiority of DDM method based on false imaginations and inadequate evidence.

As for as the applicability of the methods highlighted here is concerned, it is applicable in almost all the developing countries where information on population is not accurate enough. Most of the developing countries however are in different phases of demographic transition. Mortality rates have already declined and still under declining trend. Fertility rates also are in different levels of decline. As a result, proportion of children is on decline and that of aged is on increase and populations are far from attaining stability. Consequently, population growth rate of developing countries is highly vulnerable to unpredictable fluctuations making it difficult to suggest a particular model of inter-census estimation or post-census projection of population.

To conclude exponential method has sound scientific base whereas DDM is based on imaginary scenario. Therefore, PBCRs in India should continue exponential growth model till another model is found to fit better. PBCRs in other developing countries should have a close examination of the growth pattern to select closest fitting model. Moreover, in all the developing countries including India, there should be a periodic review of suitability of the method currently in use.

References

Agarwala, U.D., (2000) 'Population projection and its accuracy', New Delhi: *Asia Publishing House*, pp; 3-20.

Doll, R. Payne P. and Waterhouse J. (eds), (1966) 'Cancer Incidence in Five Continents: A Technical Report', Berlin, Springer-Verlag (for UICC).

Ferlay J, Soerjomataram I., Ervik M., Dikshit R., Eser S., Mathers C., (2013) GLOBOCAN 2012 v1.0, Cancer incidence and mortality worldwide: IARC Cancer Base No. 11 [online]. Lyon, France: *International Agency for Research on Cancer*, http://globocan.iarc.fr, (accessed 24.03.2014).

National Cancer Registry Programme (NCRP), (2013) 'Three-years Report of Population Based Cancer Registries 2009-2011', *Indian Council Medical Research*, New Delhi.

Newell, C., (1990), 'Methods and models in demography', *Guilford Press*; 1990 Jun 1.

Registrar General, (1987) 'District Level Population Projections 1981-1991', Occasional paper No.2 of 1987, Census of India. *Office of the R.G. of India*, New Delhi.

Shryock, H.S. and Siegel, J.S., (1971) 'The methods and materials of demography', *US Bureau of the Census*.

Takiar R and Shobana,B., (2009) 'Cancer incidence rates and the problem of denominators- a new approach in Indian Cancer Registries', *Asian Pacific J Cancer Prev,10*:123-126.

WHO (World Health Organization) (2008) 'The global burden of disease update report: 2004 Update', *World Health Organization*, Switzerland.