RESEARCH COMMUNICATION

Problem of Small Numbers in Reporting of Cancer Incidence and Mortality Rates in Indian Cancer Registries

Ramnath Takiar*, Deenu Nadayil, A Nandakumar

Abstract

The present paper examines the problem of small numbers (< 20 cases) associated with many sites of cancers in Indian cancer registries. The cancer incidence data of 14 Population Based Cancer Registries for the period of 2001-03 and 2004-05 were utilized for the analysis. Nine out of 14 registries had more than 50% of their sites being associated with small numbers while seven registries had 50% of their sites having as low as 5 cases. Sites associated with small numbers showed a lot of variation and significant differences in their incidence rates within two years duration which are not possible. The percentage age distribution was also found to vary with different periods. The paper has effectively shown the effect of population size on incidence rates. For a registry of population size 300,000, the incidence rate of 6 can very well be unstable. There are many registries in the world with their population size less than 200,000. Even in the case of registries with high population (≥500,000) the practice is to report the cancer incidence by different ethnic groups with populations less than 200,000 and thereby introduce the problem of small numbers in reporting the incidences of various cancer sites. To overcome this problem, pooling of data over broad age groups or ten years age groups or 3 to 5 years periods is one of an immediate solution.

Key Words: Cancer - AAR - small numbers - least significant incidence rate - data pool - broad age groups

Asian Pacific J Cancer Prev, **10**, 657-660

Introduction

The Indian Council of Medical Research (ICMR) started a National Cancer Registry Programme (NCRP) in the year 1982 with the main objective of generating reliable data on the magnitude and pattern of cancer in India. There are 23 Population Based Cancer Registries (PBCR) which are currently functioning under the network of NCRP. The cancer incidence data so collected is analyzed and reported in standard format from time to time in the form of one year/two years/three years report. One of the regular features of these reports is to report number of incidence and mortality cases, Crude Rate (CR), Age Adjusted Rate (AAR) as well as rates by five yearly age groups for selected sites of cancer. The National Centre for Health Statistics, USA does not publish or release rates based on fewer than 20 observations, because they feel these data do not meet their requirement for a minimum degree of accuracy and such rates are termed as unstable rates (Chronic Disease Teaching Tools, USA - 1999). Accordingly, it is expected that whenever the cancer incidence rates are based on small numbers (<=20 cases) should not be reported and if reported, at least should be highlighted. However, no such practice is evident in Indian Cancer Registries. Thus, the present paper examines the problem of small numbers associated with many sites of cancers in reporting from Indian cancer registries. The objectives of the paper are:

1. To provide the distribution of sites according to their number of cases to show the extent of problem of small numbers in Indian Cancer Registries.
2. To highlight the variation occurring in AAR with period due to small numbers associated with them.
3. To show the variation in percentage distribution of Age of cancer cases by selected site and different periods, occurring due to small numbers associated with them.
4. To construct a table providing the least significant incidence rate associated with the given population size.
5. To provide the possible solutions to deal with the problem of small numbers in Indian Cancer Registries.

Materials and Methods

For the analysis purposes, the cancer incidence data of year 2004-05 (NCRP–2008), for the Population Based Cancer Registries of Bangalore, Barshi, Bhopal, Chennai, Delhi, Mumbai, Ahmedabad and Kolkata was utilized. In addition, the cancer incidence data of the year 2005-06 (NCRP–2008), for the North East Population Based Cancer Registries of Dibrugarh, Kamrup Urban, Silchar town, Imphal West District, Mizoram State and Sikkim State was utilized. For males, there are 52 cancer sites and for females there are 56 cancer sites (C00-C95) for which five yearly age group distribution of cases and incidence rates are reported routinely.

In addition, for each site, the total number of cases,

*National Cancer Registry Programme, Indian Council of Medical Research, Bangalore, India *For Correspondence: ramnath_takiar@yahoomail.co.in*

Crude Rate (CR) and Age Adjusted Rate (AAR) are also reported routinely.

Each site, based on the number of cases associated with them, is categorized into following five categories: ≤5; 6-10; 11-15; 16-19 and ≥20. Sites having cases more than 19 were considered as having stable rates otherwise sites with cases below 20 were termed as having unstable rates and being associated with small numbers. In order to show how the sites with small number of cases (<20) can show variation in their rates from one year to another, five selected sites, known for having small numbers, were considered from the incidence data of Bhopal PBCR from the report of the year 2001-03 (NCRP – 2006) and 2004-05 (NCRP – 2008). The sites, with their ICD10 Codes, selected were: Bone (C40-41), Corpus uteri (C54), Thyroid (C73), Hodgkins disease (C81). The S.E. of these rates were calculated using the formula S.E= rate/√cases. The difference between the rates for two periods were tested using the formula; $(\text{Rate1-Rate2})/((\text{S.E. of R1})^2 + (\text{S.E. of R2})^2)^{0.5}$. For females, the percentage age distribution for the site of Larynx (C17) by five different periods for the registry of Bhopal was obtained. This was mainly done to show that if the site is associated with small numbers and then if we are comparing the percentage age distribution, thereby incidence rates, between different periods, it can be misleading.

In order to show how even a high incidence rate of certain site, in the presence of less populated registries can be associated with small numbers, a table was constructed. In table, in rows, different hypothetical populations ranging from 100,000 to 2000,000 were displayed. In column an attempt was made to translate the given incidences from one to ten (1-10) for each population to an equivalent number of cases. For example, an incidence rate of 10, for the registry population of 100,000 may give rise to only 10 cases (Cases = Incidence x Population)/100,000). Thus, an incidence rate when seen in relation to population of its registry, may give the actual status of the incidence so as to say whether it is stable or unstable. In addition, in the last column of the table, the Least Significance Incidence Rate (LSIR) was calculated and shown. An Incidence rate is termed as “Least significant Incidence Rate” for given registry of any population size if all the incidence rates below it are often associated with less than 20 cases.

IARC is reporting from time to time, the various rates related to cancer incidence from different parts of the world. However, the problem of small numbers in reporting is even not highlighted in their publication (Parkin etal., 2002). The problem of a small number is very often closely associated with the population size of the registry. Invariably, registries with lower total populations (say, 200,000) give rise to incidence rates of various cancer sites being inevitably associated with small numbers. Therefore, it was thought interesting to ascertain the number of cancer registries whose cancer incidence rates have been reported in Cancer Incidence in Five Continents Vol. VIII (Parkin et al., 2002) and which have relatively low populations (<200,000). The LSIR values were also calculated for such registries and are shown in the Table.

Table 1. Percentage Distribution of Cancer Sites According to Numbers of Cases

PBCR area	≤5	6-10	11-15	16-19	≥20	Total
Males						
Barshi	63.5	25.0	7.7	1.9	1.9	52
Silchar town	86.5	7.7	1.9	0.0	3.8	52
Imphal West	61.5	9.6	15.4	7.7	5.8	52
Sikkim	63.5	11.5	11.5	1.9	11.5	52
Aizawl	55.8	19.2	9.6	3.8	11.5	52
Ahmedabad	34.6	21.2	11.5	11.5	21.2	52
Dibrugarh	50.0	11.5	9.6	3.8	25.0	52
Kamrup Urban	38.5	17.3	11.5	3.8	28.8	52
Bhopal	30.8	17.3	5.8	9.6	36.5	52
Kolkata	26.9	5.8	3.8	9.6	53.8	52
Chennai	11.5	5.8	5.8	5.8	71.2	52
Bangalore	15.4	3.8	-	5.8	75.0	52
Mumbai	9.6	1.9	1.9	3.8	82.7	52
Delhi	7.7	5.8	-	1.9	84.6	52
Females						
Silchar town	91.1	3.6	0.0	3.6	1.8	56
Barshi	89.3	3.6	0.0	3.6	3.6	56
Aizawl	71.4	12.5	7.1	0.0	8.9	56
Ahmedabad	57.1	26.8	5.4	0.0	10.7	56
Imphal West	67.9	10.7	5.4	5.4	10.7	56
Sikkim	64.3	16.1	3.6	3.6	12.5	56
Dibrugarh	60.7	7.1	8.9	7.1	16.1	56
Kamrup Urban	53.6	17.9	7.1	5.4	16.1	56
Bhopal	46.4	17.9	8.9	3.6	23.2	56
Kolkata	30.4	14.3	14.3	3.6	37.5	56
Chennai	16.1	8.9	7.1	1.8	66.1	56
Bangalore	19.6	5.4	0.0	7.1	67.9	56
Mumbai	12.5	3.6	3.6	5.4	75.0	56
Delhi	10.7	5.4	1.8	3.6	78.6	56

Results

The percentage distribution of sites according to their number of cases by different Population Based Cancer Registries for both males and females are shown in Table 1. The registries were arranged according to ascending order of percentage of sites with adequate numbers. The registry of Barshi to Aizawl had less 15% of their sites with adequate number of cases implying more than 85% of the sites had number of cases less than 20 thereby giving mainly the incidence estimates as unstable rates. The registry of Ahmedabad to Bhopal had between 20-40% of their sites while the registry of Kolkata had around 50% of their sites having adequate number of cases. The registries of Chennai to Delhi had 70% to 85% of their sites having adequate number of cases. Similarly, for males the registry of Barshi to Aizawl including Dibrugarh, more than 50% of their sites had number of cases as low as 5. Even the registry of Kolkata, Bhopal and Kamrup urban had 25-40% of their sites with numbers of cases below 5.

As in the case of males, the registries were arranged according to ascending order of percentage of sites with adequate numbers. The registry of Silchar town to Dibrugarh, less than 20% of the sites only had adequate number of cases associated with them. Bhopal and Kolkata registries had 20-40% of their sites having adequate numbers. In the case of other registries, 65-75% had adequate number of cases associated with them. Silchar town to Dibrugarh registries had more than 55% of their

Table 2. Variation in AAR by Selected Sites Associated with Small Numbers and Years - Bhopal PBCR

Site		Bone M	Thyroid M	Hodgkins disease M	F	Corpus uteri
ICD10		C40-41	C73	C81		C54
Cases	2001-03	19	6	19	4	19
	2004-05	23	12	6	6	25
AAR	2001-03	0.7	0.2	0.8	0.2	0.9
	2004-05	1.5	0.7	0.3	0.4	1.6
% change		214	350	37.5	200	178
SE of	2001-03	0.161	0.082	0.184	0.100	0.206
AAR*	2004-05	0.313	0.202	0.122	0.163	0.320
Significance P		<0.05	<0.05	<0.05	NS	NS

M, males; F, Females; *AAR/√cases

Table 3. Percentage Distribution of Cancer Cases of Larynx (C32) by Different Periods - Females

Age group	04-05	01-03	99-00	97-98	90-96
<= 30	-	-	-	-	-
31-40	-	20.0	-	33.3	-
41-50	-	20.0	-	-	18.8
51-60	33.3	20.0	-	50.0	37.5
61-70	-	20.0	100.0	-	25.0
>=71	66.7	20.0	-	16.7	18.8
Total cases	6	5	1	6	16

Source: National Cancer Registry Programme reports

Table 4. Calculation of Cases and LSIRs with Populations in 1,000s and Incidence Rates per 100,000

	1	2	3	4	5	6	7	8	9	10	
100	1	2	3	4	5	6	7	8	9	10	20.0
150	2	3	5	6	8	9	11	12	14	15	13.3
200	2	4	6	8	10	12	14	16	18	20	10.0
250	3	5	8	10	13	15	18	20	23	25	8.0
300	3	6	9	12	15	18	21	24	27	30	6.7
400	4	8	12	16	20	24	28	32	36	40	5.0
500	5	10	15	20	25	30	35	40	45	50	4.0
750	8	15	23	30	38	45	53	60	68	75	2.7
1000	10	20	30	40	50	60	70	80	90	100	2.0
1500	15	30	45	60	75	90	105	120	135	150	1.3
2000	20	40	60	80	100	120	140	160	180	200	1.0

sites having as low as 5 number of cases associated with them. Bhopal and Kolkata registries had 30-50% of their sites associated with less than 5 cases.

The variation in AAR by period of reporting for selected sites, associated with small numbers, is shown in Table 2. It can be seen that for the cancer sites of Bone and Thyroid among males, the relative changes in AAR for the period of 2004-05 as compared to 2001-03 were more than 200% while for other cancer sites in females, the relative changes were more than 180% of that seen in the year 2001-03. So, all the selected sites have either increased to more than doubled or reduced to one-third in 2004-05 as compared to that seen in the period of 2001-03. Among males, all the three selected cancer sites showed significant variation in 2004-05 as compared to that seen in the year 2001-03. However, in females though the variation in incidence rates was also observed but it was not found significant.

The percentage distribution of cancer cases of Larynx (C32) by selected broad age groups categories and periods is shown in Table 3. It can be easily seen that going by the data of 1997-98 and 2001-03, there is no case below the age of 30 years while by the data of 1990-96, there is no case below the age of 40 years. However, the data of 2004-05 suggests that there is no case below the age of 50 years. So, different periods suggest different minimum ages below which the cases are not seen. The percentage age distribution obviously also seems to be varying.

For individual populations, the Least Significant Incidence Rates can be obtained easily from Table 4. For populations up to 150,000, the Least Significant Incidence Rate was observed to be more than 10. For 200,000 populations, it was 10 while for 250,000 populations it was 8. The Least Significant Incidence Rate decreases as we go from low population to high populations. After 750,000 populations, the Least Significant Incidence Rate will be below 2 per 100,000 populations.

Cancer Registries from different parts of the world with population below 200,000 along with their Least Significant Incidence Rate are shown in Table 5. There were many registries with population below 200.000 whose results are reported in Cancer Incidence in five

Table 5. Cancer Registries with Populations below 200,000 and their Least Significant Incidence Rates (LSIR)

Country	Registry name	Males	LSIR*	Females	LSIR*
Canada	Yukon	14,472	138	13,184	151
Argentina	Concordia	71,071	28	74,621	26
Switzerland	Neuchatel	79,059	25	84,820	23
Italy	Biella Province	90,765	22	99,607	20
Spain	Cuenca	100,618	19	101,493	19
Switzerland	Graubunden & Glarus	110,407	18	112,872	17
Portugal	Vila Nova de Gaia	126,975	15	133,868	14
Switzerland	Valais	132,288	15	137,316	14
Iceland	-	134,062	14	133,340	14
Italy	Macerata Province	142,019	14	150,893	13
Italy	Ragusa Province	145,692	13	151,980	13
France	Tarn	166,692	11	175,364	11
Austria	Vorarlberg	170,225	11	172,264	11
France	Martinique	177,763	11	194,584	10
Argentina	Bahia Blanca	187,325	10	198,933	10
China	Jiashan	191,005	10	186,350	10

*Parkin et al., Cancer Incidence in Five Continent Vol VIII; IARC Scientific Publication No. 155

Continents (Parkin etal. 2002), However, only few selected registries are listed in the Table 6. When the registry population is less then the LSIR can be very high like in the case of registry from Canada, the LSIR is 138 for males and 151 for females. When the population is 191,005 in the case of Jiashan registry of China, the LSIR is 10.

Discussion

It becomes clear from the Table 1 that for Barshi to Dibrugarh, all registries listed had a majority of their sites having small numbers associated with them. From Barshi to Sikkim, all registries had more than 50% of their sites being associated with as low as 5 cases. Thus, it can be said that small numbers associated with various cancer sites can be in general a problem with majority of the Indian registries. Further, it was shown clearly that when the sites are associated with small numbers, they can show as much as 200% variation in their incidence rates. In few cases, the incidence can become almost 50% with the period. Thus, cancer sites when they are associated with small numbers can show a lot of variation in their Age Adjusted Rates making it difficult to interpret the changes occurring in them (Table 2).

The data provided in Table 3, make it clear how the percentage distribution when seen in relation to small numbers can be misleading. The minimum age above which the larynx cases are reported are shown to vary from around 30 years to 60 years among different registries. This variation can be attributed to small numbers. If the incidences of sites are based on adequate numbers, such a variation in minimum age should not have been there. While giving most of the results by registries in percentages like providing the relative percentages of cancers based on different microscopic diagnosis (Primary histology, Secondary histology etc.) can be less meaningful when the numbers associated with them are particularly very small.

The effect of population on incidence rate is also brought out clearly in Table 5. It can be argued that when the population of the registry is around 150,000 then the incidence rate of even 10 should be viewed with reservations. Similarly, when the population of the registry is around 300,000 then even the incidence rate of 6 can be termed as least significant and should be interpreted with care. The problem of small numbers in relation to various cancer sites, particularly, in Indian cancer registries is never highlighted. While considering the rare cancers which are often associated with small numbers should be viewed carefully for assessing the time trend and yearly fluctuations occurring in them. In addition, the basis of diagnosis according to different categories (Microscopic, X-ray etc.) along with their percentages is provided routinely in each report for each site of cancer. When the sites are associated with small numbers, the percentage distribution may not give correct picture and can be misleading.

As many registries in the world are functioning with small populations (Table 5), the problem of small numbers in reporting of cancer incidences becomes inevitable. LSIR of 138 in the case of males and from the registry of Canada signifies that all incidences below it should be viewed with reservations. Similarly, all the incidences above 10 only should be viewed with reliability in the case of incidences from Jiashan registry of China. It is desirable that for all the registries of the world, the LSIR should be close to one so that all incidences observed can be viewed as stable rates otherwise an incidence rate of 138 also should be viewed as an unstable rate and in presence of small numbers such rates can show abnormal fluctuations from one year to another. Even in the case of registries with high population (≥ 500,000) the practice is to report the cancer incidence by different ethnic groups like reported from the registries of USA. In such situations very often the ethnic groups reported end up with small populations, less than 200,000 and thereby again introducing the problem of small numbers in reporting the incidences of various cancer sites. So the problem of small numbers in reporting can be found to be associated with various registries of the world from developed as well as from developing countries.

In view of the above, it is suggested that all such sites which are associated with small numbers should be marked and should be highlighted in the report. The problem of small numbers in cancer registries can be addressed in more than one ways. First, if the registry population is less than 300,000, it should be expanded so as to cover at least 500,000 populations so that meaningful and stable incidence rates for the given registry can be provided. Second, the cancer incidence data can be pooled for 2-5 years and reported. Reporting the incidence data for broad age groups say every ten years age group can also be an immediate solution to the problem of small numbers. In general, It is needed that all registries should view the problem of small numbers and should make an attempt to either highlight or should go for any suggested solution.

References

National Cancer Registry Programme (ICMR). 2006. Consolidate Report of the Population Based Cancer Registries: 2001-2004. Bangalore, India.

National Cancer Registry Programme (ICMR). 2008. NE-PBCR Report: 2005-06. Bangalore, India (http://www.pbcrindia.org/)

National Cancer Registry Programme (ICMR). 2008. PBCR Report 2005-06, Bangalore. India (http://www.pbcrindia.org).

Parkin DM, Whelan Sl, Ferlay J, Teppo and Thomas DB. 2002: Cancer Incidence in Five Continents Vol. VIII – IARC Press, Lyon, France.

Rates Based on Small Numbers - Statistics Teaching Tools – USA -1999. (http://www.health.state.ny.us/diseases/chronic/ratesmall.htm)