Annexure I
Definitions, Statistical Terms and Methods

Cancer Registration may be defined as the process of continuing, systematic collection of data on the occurrence and characteristics of reportable neoplasms to help assess and control the impact of malignancies on the community.

Cancer Case refers to all neoplasms with a behaviour code of '3' as defined by the International Classification of Diseases - Oncology, third edition (ICD-O-3) are considered reportable and are registered in NCRP.

Cancer Registry is the office or institution that attempts to collect, store, analyse and interpret data on cancer patients.

Population Based Cancer Registries (PBCRs) systematically collect information on a reportable neoplasm from multiple sources in a geographically defined population residing in the area for one year.

Hospital Based Cancer Registries (HBCRs) are concerned with recording of information on the treatment, management and outcome of cancer patients registered in a particular hospital.

Sources of Registration will usually be hospitals or cancer centres but, depending on the local circumstances, a population-based registry will also involve private clinics, general practitioners, laboratories, health insurance systems, HBCRs, screening programmes and Vital statistics Department.

Data Processing Data Processing involves importing or downloading data from the registries into the local database at ICMR-NCDIR. Quality of the data is checked for errors that may have been committed at data collection, abstraction or entry. Identification and elimination of duplicates is done through deterministic approach and by identifying names that are phonetically the same. Multiple combination of variables is used to generate the probable list. Duplicate deletion is done without any loss of information. Mortality data is linked/matched with incidence and the unmatched mortality cases are identified as either Death Certificate Notification (DCN)/ Death Certification Only (DCO). Clarification at each step is sought from each registry and the data is finalized for further analysis.

Age-Group used for estimating populations as well as grouping cancer cases as per the WHO guidelines which is 0-4, 5-9, 10-14….75+.

Cancer Incidence denotes new cases diagnosed in a defined population in a specified time period.

Cancer Mortality denotes the number of cancer deaths occurring in a specified population during a specified time period.

Rates for cancer are always expressed per 100,000 population.

Crude Incidence Rate (CR) refers to the rate obtained by division of the total number of cancer cases by the corresponding estimated population (mid-year) and multiplying by 100,000.

$$CR = \frac{\text{New Cases of cancer of a particular year}}{\text{Estimated population of the same year}} \times 100,000$$

Age Specific Rate (ASpR) refers to the rate obtained by division of the total number of cancer cases by the corresponding estimated population in that age group and gender/ site/geographic area/time period and multiplying by 100,000.

$$ASpR = \frac{New\ Cases\ of\ cancer\ of\ a\ particular\ year\ in\ the\ given\ age\ group}{Estimated\ population\ of\ the\ same\ year\ for\ the\ give\ age\ group} \times 100,000$$

Age Adjusted or Age Standardised Rate (AAR) Cancer incidence increases as age increases.

In order to make rates of cancer comparable between countries/population, a world standard population (given below) that takes this into account is used to arrive at age adjusted or age standardised rates. This is calculated according to the direct method by obtaining the age specific rates and applying these rates to the standard population in that age group. The world standard population approximates the proportional age distribution of the world and is given below:

Age Distribution of World Standard Population (Segi.et.al)

Age Group	World Standard Population
00-04	12,000
05-09	10,000
10-14	9,000
15-19	9,000
20-24	8,000
25-29	8,000
30-34	6,000
35-39	6,000
40-44	6,000
45-49	6,000
50-54	5,000
55-59	4,000
60-64	4,000
65-69	3,000
70-74	2,000
75+	2,000
All Ages	100,000

$$AAR = \frac{\sum_{i=1}^{A}(a_i w_i)}{\sum_{i=1}^{A} w_i}$$

Where,

a_i is the age specific rate (ASpR) in age class i;

w_i is the standard population in age class i;

A represents the number of age intervals.

Or expressed in simpler terms thus:

$$AAR = \frac{\sum(ASpR) \times (No.\,of\ persons\ in\ Std.\,world\ population\ in\ that\ 5\ yr.\,age\ group)}{100,000}$$

Mortality rate Crude Mortality Rate (CMR) and Age Adjusted Mortality Rate (AAMR) were calculated using the same above formula. However, the numerator is death cases.

Cumulative Risk refers to the probability that the person will develop a particular cancer during a certain age period in the absence of any other cause of death. The Cumulative Rate (CuR) is an approximation of the cumulative risk. It is obtained by adding the annual age-specific incidence rates for each five-year age interval (up to 74 years of age or for whatever age group is to be used to calculate the cumulative risk) multiplied by 5 (representing the five-year age interval) times 100/100,000.

61

$$CuR = \frac{5 \times \sum(ASpR) \times 100}{100,000}$$

And cumulative risk is expressed as

$$Cumulative\ Risk = 100 \times [1 - \exp(-\frac{CuR}{100})$$

$$Probability\ one\ in\ number = \left(\frac{1}{Cumulative\ Risk}\right) \times 100$$

About Joinpoint Regression Program- Joinpoint Regression Program, Version 4.7.0, is a statistical software for the analysis of trends using Joinpoint models, i.e, where several different regression lines are connected together at the "Joinpoints". The software takes trend data (e.g. cancer rates) and fits the simplest Joinpoint model that the data allow. The program starts with the minimum number of Joinpoints (e.g. 0 Joinpoint, which is a straight line) and tests whether more Joinpoints are statistically significant and must be added to the model (upto that maximum number). In this report we have seen Annual Percent Change (APC) of straight line for a specified period of time.

For example, if the APC is 1%, and the rate is 50.0 per 100,000 in 2000, the rate is 50 × 1.01 = 50.50 in 2001 and 50.5 × 1.01 = 51.005 in 2002.

Rates that change at a constant percentage every year, change linearly on a log scale. For this reason, to estimate the APC for a series of data, the following regression model is used.

$$\log(R_y) = \beta_0 + \beta_1 y$$

Where, $\log(R_y)$ is the natural log of the rate in year y.

The APC from year y to year $y + 1 = \frac{(R_{y+1} - R_y)}{R_y} \times 100$