Chapter 2

OVERALL PLAN AND METHODS

Initial Registration:

An invitation was sent to the Principals of all Medical Colleges (based on the list with the Medical Council of India) with copies to the Head of the Department of Pathology during November-December 2000. Similar letters were also sent to several major hospitals all over India. In all 264 institutions were contacted. A registration form was also sent along with the letter. The form requested institutions to provide information on the following items:

1. Name and address of Institution;
2. Name of Head of Institution;
3. Name and designation of Principal Investigator, Co-Principal Investigator and Faculty-in-charge;
4. Possible method to interview and record residential address of the patient in whom a malignancy has been diagnosed;
5. Number of malignant neoplasms reported per annum by the department of pathology (comprising histo-pathology, haematology and cytology);
6. Computer facilities, internet and telephone connectivity and access to the same in the institution and in the department of pathology;
7. Budgetary requirement for collation and transmission of core items of patient information;
8. Approval of Head of Institution and Head of the Department of Pathology (or other as the case may be) for participation in the study with respective signatures.

The form also made clear that:

A. The Principal Investigator will be the main corresponding/contact person for all matters including release of funds and be overall in-charge of the project in the respective institution.

B. However, a person preferably a junior faculty member on the permanent role had to be identified. This person would be responsible for the day-to-day working of the project. The person so identified should be interested in such work in the project that involves:

   (i) completion of identifying information, especially residential status of all malignant neoplasms as and when reported in the Department of Pathology;

   (ii) completion of Topography and Morphology details including coding of such cases;

   (iii) ensuring that the same was correctly entered on to the computer;

   (iv) ensuring that the data so entered was regularly transmitted through the net to the Coordinating Unit;

   (v) replying to queries concerning the data transmitted.

Appropriate training and guidance were to be provided on all of the above.

C. Nomination of Co-Principal Investigator was Optional.

D. WHO/ICMR did not accept any responsibility for employing persons on the project.

As soon as the completed registration forms were received, centres were assigned a numerical centre code (according to the order received) and blank printed core forms (Appendix I) supplied to the Principal Investigator. Brief guidelines for completion of the items of patient information were also sent. Centres were asked to commence data collection on all malignant neoplasms reported from 1 January 2001.

Core Committee

A core committee consisting of the following members was constituted:

1. Dr P.C. Gupta, Senior Research Scientist, Tata Institute of Fundamental Research, Mumbai and member of the Steering and Monitoring Committees of NCRP;
2. Mr P. Gangadharan, Emeritus Scientist, Regional Cancer Centre, Thiruvananthapuram and Hon. Consultant, Natural Background Radiation Registry, Karunagappally, Kerala and Member, Monitoring Committee of NCRP;
3. Dr R.N. Visweswara, Professor of Pathology, Vaidehi Institute of Medical Sciences, Bangalore and Member, Monitoring Committee of NCRP.

The committee with Dr P.C. Gupta as chairperson met on 9 February 2001. The meeting reviewed the response received from the several centres and recommended that visits by the Project-in-charge and at least one of the members of the core committee be made to each of the potential centres. The meeting outlined the objectives of the visit (see below).

In addition, the meeting recommended that preference should be given to pockets where little knowledge is available about cancer patterns. A well formulated "terms and conditions of agreement" with the selected centres needed to be worked out. The scientific validity of the data using minimal incidence rates based on microscopic verification as well as incomplete coverage of the given geographical area was discussed at length.

Visits to Centres and Agreement on Terms and Conditions

The objectives of the visits were:

i) to make sure that the items of information provided by the institutions in the registration form broadly conform to what was being sought, in particular with reference to the number of malignancies reported per annum;

ii) feasibility of the set-up in obtaining accurate and complete information on residential status in "all" cases of malignant neoplasms reported;

iii) the agreement and cooperation of the Head of the Institution;

iv) the interest and commitment of the Head of the Department of Pathology and/or Faculty in Charge in taking complete responsibility of the Project;

v) assessment of existing infrastructure for successfully executing the project and existing computer, internet and telephone facilities;

vi) facilities at the institution for radiation and other oncology services;

vii) Details of Budget requirements;

viii) Continuation of the Project beyond two years;

ix) Possibility of having the cooperation of other pathologists within the area to provide information on cases reported by them.

Apart from agreeing on budgetary and computer (based on existing facilities and number of malignancies reported annually) support the terms and conditions specifically referred to the timely transmission of required core patient information and the responsible person for such activity.

Workshops

1. Four Regional/Zonal Workshops were held on the dates and places indicated below:

a)	Southern Zone	11 - 12 May 2001	Bangalore
b)	Western Zone	8 - 9 June 2001	Mumbai
c)	Eastern Zone	15 - 16 June 2001	Kolkata
d)	Northern Zone	14 - 15 September 2001	Lucknow.

These workshops essentially aimed at the following objectives (specific details of which are discussed in the later part of this chapter):

a) Provide a forum for collaborating centres to present their initial experience towards data collation;

b) Give training in basic principles and techniques of cancer registration and coding according to the International Classification of Diseases;

c) Importance of contact with patient or close relative/friend in order to obtain reliable and accurate information on permanent place of residence and other identification details;

d) Guidelines for correctly completing the various items of patient information in the core proforma;

e) Efforts required to get the exact primary anatomical site of tumour in all reported malignant neoplasms;

f) The methods of collation of patient information in different settings - government medical colleges and hospitals, private hospitals, cancer centres, pathology laboratories etc.

g) Necessity of coverage of other institutions registering and reporting malignant neoplasms in the geographic area;

h) An overview of web-site development and on-line transmission of data.

2. A meeting of existing cancer registries in India (both within NCRP and those outside the network) was held on 2 August 2001 - this was mainly to examine how these registries could contribute to this project and vice versa;

3. Meeting-cum-workshop for centres that could work towards obtaining Cancer Incidence Rates in their geographic area - 26-27 March 2002;

4. North East Regional Workshop was held on 11 May 2002;

5. All India Review Meeting cum Workshop was held from 31 July to 2 August 2002.

Collation of Data by Collaborating Centres

The overall method of data collation that is generally adopted by the centres is presented here. Some specifics could vary between and among centres.

1. Identification of a cancer case:

The first step towards collation is identification of the recording of a malignant neoplasm. The method of obtaining this varies in different settings.

Cancer Centres: Generally cancer centres in India are referral centres for diagnosed or suspected cancer patients. Therefore, the identifying information is completed for all patients who attend that centre for the first time, regardless of whether a microscopic report of malignancy exists or not. This is made at the time of initial registration

by a medical doctor, trained social worker, post-graduate medical student, nurse or any other trained person. A provisional noting of the diagnosis is made in the core proforma wherever a record/report of diagnosis of malignancy is available. The diagnostic portion is subsequently completed after reviewing the records/reports of the pathology department.

Medical College Hospitals and other General Hospitals (Government and Private): Usually, cancers constitute less than 10% of all diseases in a general hospital setting. Therefore, unlike that in the cancer centre the contact with the patient/relative/close friend is taken-up only after a diagnosis of malignancy, is made by the department of pathology. However, centres use different approaches for histo-pathology, haematology and cytology. For the latter two methods of diagnosis, normally, patients personally visit the laboratory for giving blood or bone marrow samples or present themselves for smears to be taken. The chances of the pathologist looking up the patient and the patient's records for details of suspected cancer if any are high. The identifying information in the core proforma is completed for such patients wherein a malignancy is diagnosed or suspected. Whenever a histopathology diagnosis of malignancy is made, the concerned patients are followed back to the in-patient wards and in where the patient is not admitted or discharged, through the concerned physician.

Pathology Laboratories: Histopathology specimens are often received at the pathology laboratory and the report collected by one of the close family members or friends of the patient. In these circumstances, identifying the report with a diagnosis of malignancy and contacting the patient's representative for the required identifying information, by the concerned pathologist with the help of his secretarial staff posed little difficulty. However, occasionally in some pathology laboratories, specimens are sent through courier or messengers, by surgeons practising in rural areas to the laboratory in the urban area. In such instances the collaborating pathologist has developed a rapport with the oncologists in the area and the required information is gathered.

2. Completion of the Core Proforma

a) *Identifying Information:* Besides the name of the patient certain additional details like Name of Father, Mother, Spouse and that of Son, Daughter are sought, mainly to help in checking duplicate registrations. The details of address of the permanent place of residence are of paramount importance. The key information is the location (at least at the level of district) for the past year. Therefore the information on duration of stay at the residential address is a guide to confirming that the patient is actually residing at that address and that it is not a temporary place of dwelling for the purpose of treatment. This is the same rule that is followed for cancer registries under the NCRP, in that a cancer patient is taken into the registry, provided he or she has stayed in the geographic area of cancer registry operation for a minimum period of one year.

b) *Diagnostic Information:* Once a pathological - microscopic diagnosis of cancer has been made, the details of the diagnosis including coding according to the WHO, International Classification of Diseases - Oncology, 3rd edition [ICD-O-3] (WHO, Fritz et al, 2000) is done. The concerned faculty (mostly from the department of pathology or radiotherapy) who is in charge of the project at the respective centre oversees the diagnostic information and coding. Whenever the exact primary site of tumour is unknown, efforts are made to contact the treating clinician to obtain the details of primary site.

Software Development and Functioning of web-site

The web-site with the following address was launched in January 2002: canceratlasindia.org and cancermapindia.org.

The data is being received on a day to day basis.

The advances in information technology helped the Cancer Atlas Project identify internet as one of the primary communication medium for collecting the data. Internet as a means of data collection for patient information was a

unique concept being tried for the first time in India under the project.

The 'core proforma' for collecting the patient information was hosted on the web-site www.canceratlasindia.org. It was designed to be user friendly and reduce the time taken for data entry. Internet Browser based data entry eliminated the need for software to be installed on every system, the hassles of administration and maintenance.

Care was been taken to code/encrypt the data entered so that the identity or the nature of the data cannot be deciphered by any one except those concerned with the project. Thus, security of data sent over the Internet was implemented at four levels.

A unique username and password for each participating centre ensured authorized data entry.

Once the data is sent over the Internet, encrypting the data is ensured using a 128-bit Secured Socket Layer (SSL) based algorithm. SSL is a protocol for transmitting data over the Internet by using a private key to encrypt data that's transferred over the SSL connection. SSL ensures that the information is sent unchanged only to the specific server, to which, it was planned to send.

A third level of security ensures that once any participating centre enters the data on the proforma (on the web-site) and submits the same, the centre or any user cannot recall that completed proforma to make any changes or alterations of the data so entered. The Coordinating Unit examines the data and makes changes/corrections if any, only in its data base.

The fourth level of security is implemented by storing the data on a Database Server in an encrypted form. The Coordinating Unit periodically downloads this data and the data on the database server hosted on the Internet is deleted, thereby ensuring a comprehensive security mechanism for the data.

The data that is downloaded by the Coordinating Unit is decrypted and stored on a local database on an SQL Server. This data is sanitized and used for off-line checks, analysis, preparation of reports etc.

The Coordinating Unit uses a 'Database Server' and 'analysis stations' connected over a Local Area Network. A Secure Firewall installed in the Coordinating Unit keeps away hackers from accessing any information/data.

The successful working of this concept is reflected, first, in the data that is downloaded on a regular basis for the past two and a half years. Secondly, in the feedback received from the participating centres during the All India Workshop - 95% of the respondents, felt, that the web-site was easy to use and 80% of them have a fairly stable Internet connection.

Transmission of Data:

Collaborating centres were given an individual login-ID and password with detailed instructions on entering the data. Essentially, the process of data transmission through the web-site "canceratlasindia.org" involves the following steps:

a) one set of completed forms as hard copy kept ready;

b) connecting to the internet;

c) opening the internet browser;

d) opening the URL: http://www.canceratlasindia.org

e) typing login ID (user ID) and password;

f) opening the page with the "Proforma";

g) transferring all the details from the hard copy of a particular form to the proforma on the web-site

h) submitting the form

Entry of data on to the proforma on the web-site followed by transmission, is done by the person, authorised to do so by the Principal Investigator of the respective centre. Usually it is done from its own internet connection. If internet connectivity is not available at the department or at the collaborating centre or if the desired speed of

connection is not existing, centres use the public browsing locations.

Data from a few centres is also transmitted through floppy disks or file transfer. Two centres have wholly computerised hospital databases. The information on the core proforma, except the coding according to the ICD-O-3 is readily available on their database. The concerned staff of the project, scrutinize the records for accuracy of diagnosis and code the same. Data of the required variables are copied in a 'file' and the same transmitted periodically, either through e-mail or on a floppy disk. In a couple of other centres, the number of cases reported annually is more than 4000 and the quality of internet connection at these centres is not satisfactory and therefore not cost-effective. An off-line data entry programme has been provided to these centres. These centres enter the data off-line and transmit the same either through e-mail or floppy disk.

A few rural centres do not have ready access to personal computers or internet connection. These centres photocopy the completed forms and send the set of forms at periodic intervals.

The Data Flow Diagram of the functioning of the Website is shown in Fig. 2.1.

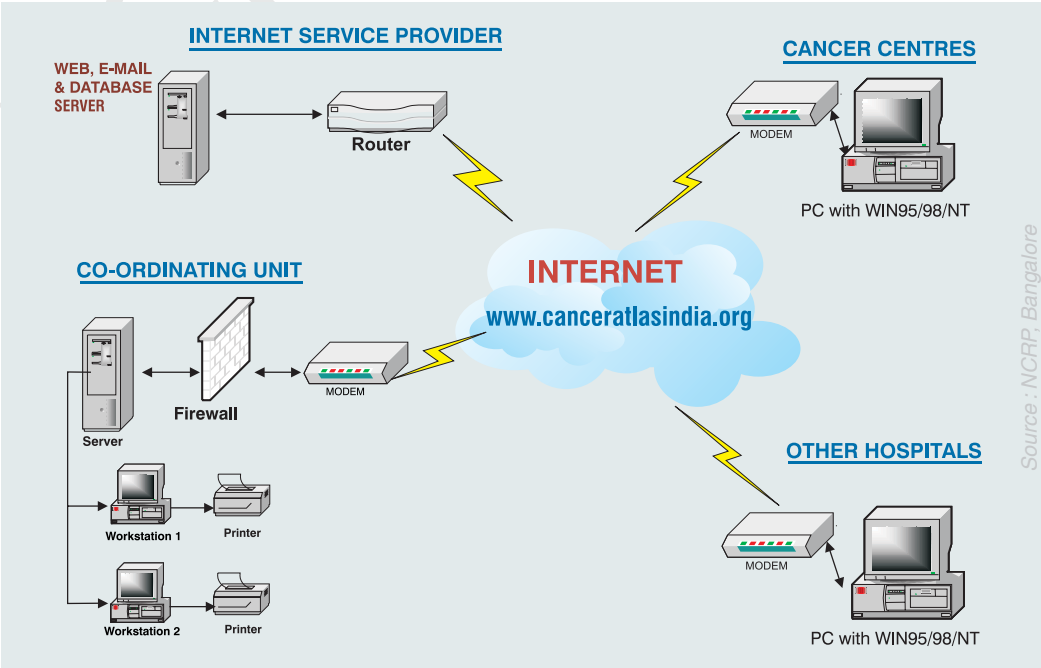
Checks and Data Processing

The data received through the web-site is downloaded on a weekly basis.

The following checks were done on the data and where needed, clarifications sought from individual centres:

1. For the data already entered on to the computer, checks were done to ensure that the data received had no formatting errors, specifically with reference to registration number, name of patient and date of report;
2. Information on completeness and correctness of information on residential status (at least the District of residence of the patient should be known);
3. Information on completeness and correctness of information on diagnosis was checked and whether coding was done according to ICD-O-3;
4. From steps 1-3 the following cases were excluded/clarifications sought from centres:
   a) District Unknown;
   b) Date of report earlier than 1 January 2001;

FIGURE 2.1 : Data Flow Diagram of the functioning of Website



c) Non-microscopic diagnosis of cancer;

d) Behaviour code of morphological diagnoses being less than /3;

5. A detailed check programme that was developed in-house was carried out on the data. These included the following:

A. On Identifying Information

a) Postal PINCODE versus District;

b) Centre-wise listing of possible duplicate registrations using patient's name, topography site and morphology, age, gender, address and other identifying information, hospital registration and pathology slide number;

c) Similar check as in b) within each district

d) Date of Report and Registration number

B. On Diagnostic Information - an in-house version based on the list of checks furnished by IARC (Parkin et al, 1994) was prepared. It included the following:

a) Range check on code - topography and morphology;

b) Impossible sex and site (topography) combinations;

c) Wrong or unlikely sex and morphology combinations;

d) Unlikely age - site combinations;

e) Unlikely age - morphology combinations;

f) Checks on 63 morphological families for unlikely site and morphology;

6. Further checks on Identifying Information:

a) Unlikely district distribution of cases in a given centre;

b) Unlikely centre distribution of cases in a given district;

c) Sorted listing of cases by district code and verification of address and district name with district code for each case;

7. Further checks on Diagnostic Information:

a) Sorted listing of cases by Topography Site code and verification of topography description and code;

b) Sorted listing of cases by Morphology code and verification of morphology description and code;

8. Specific attention to look into, clarify and minimize cases with:

a) Unknown Primary Site of tumour;

b) Ill defined sites;

9. Conversion to both ICD-9 and ICD-10 using SEER and IARC conversion programs and verifying for any inconsistencies;

Principles in Data Analysis and Presentation of Results

1. The reference manual - Cancer Incidence in Five Continents (Parkin et al, 2002) was used to group neoplasms by site (ICD-10 - WHO ISCD, 1994) and determine leading sites of cancer. There are few finer differences in the definition of some anatomical sites between ICD-9 and ICD-10. Thus oropharynx that was earlier grouped along with tonsil in ICD-9 is now separated in ICD-10. Similarly, rectum has been separated into rectum and anal canal, vagina into vagina and vulva, ovary into ovary and other female genitalia. In the other way round brain and nervous system are now grouped together as brain in ICD-10 and myeloid, monocytic and other leukaemias of specified cell types are combined as myeloid

leukaemia. Such variance may be kept in mind when relating the rates or charts with earlier NCRP reports.

2. Diagrammatic bar charts with relative proportion (%) of ten leading sites of cancer for each centre that reported at least 200 cases are presented in Chapter 7 (Individual Centre's Data);
3. Population (Census of India publications) according to five-year age group and gender is available by district. As per the 2001 census results, there are 593 districts in the country. Information on cancer cases also gives the identity of the district for each case. Therefore, the district is taken as a unit for calculation of incidence rates;
4. The most recent data from established population based cancer registries (PBCR: 1997-1999) is also included, for description and comparison.
5. All cases are microscopically confirmed; all institutions in the district have not been covered; therefore, the age adjusted incidence rates calculated here are referred as only minimum crude (MCR) and minimum age adjusted incidence rates (MAAR) for districts. (See Discussion).
6. Most of the centres have contributed data for both the years 2001 and 2002. However, a few centres have provided data only for the year 2001 and a few only for 2002. Accordingly, where the variation in MAAR of districts between each of the years is no more than 10% the average annual MAAR has been taken. Where the variation in MAAR is more than 10% between each of the years the higher MAAR is used.
7. All districts that have a higher MAAR than that calculated for the most recent PBCR at Barshi (1997-1999) are represented in the salient features on individual districts. PBCR Barshi has the least incidence rates among the registries under NCRP, it is the only rural registry and majority of the districts in the country has predominantly semi-urban or rural population.
8. The different district-wise maps of India in Chapter 4 (all sites) and Chapter 6 (specific sites) display districts with relatively higher incidence rates in darker shades and those with lower incidence rates in lighter shades of the same colour. A key has also been provided for each map. Besides, there are also areas/districts in grey. These latter areas represent places with paucity of information on cancer cases so much so that no meaningful incidence rates could be calculated.
9. The following two points were considered in choosing specific anatomical sites of cancer for detailed comparison of incidence rates nationally and internationally (Chapter 6). (i). Those sites where at least 5 districts showed a higher MAAR than the highest MAAR of that site in the PBCRs under NCRP. (ii). A site was also included for description if the MAAR of that site in any district was comparable with the highest incidence rates in the world. Sites of cancer with less than 10 cases, even if they have higher MAARs than that of the highest MAARs in the PBCRs are excluded from the bar charts so as to avoid overestimation or misinterpretation. However, in order to place the facts in right perspective the appropriate shades of colour depending on the MAAR are portrayed for districts in the map, regardless of the numbers of cancers of that site. The latter would also account for districts with small populations and identify potential hot spots of high incidence that could possibly sustain over time.
10. While presenting the profile of cancers in collaborating centres (Chapter 7), the centres are arranged in the descending order of the number of cancers (for the combined years 2001 and 2002) on which information was provided to the project, after grouping into HBCRs, PBCRs and all other centres.

Other Definitions, Statistical Terms and Methods used in Calculations of incidence rates are given in Appendix II.

The list of Topography sites with ICD-10 code with mode of grouping for determining leading sites of cancers and acronyms used in figures of bar charts is given in Appendix III.

The list of States/Union Territories with code (according to Census of India, 2001) and acronyms used in figures of bar charts against names of districts is given in Appendix IV.