



## DEVELOPMENT OF DATABASE AND ANALYSIS OF COHORTS ON CANCER

**Ms Rashmi K S**

Research Scholar, Department of Biochemistry, Meenakshi Academy of Higher Education &amp; Research, Chennai.

**Dr. D. Anandhi\***

Assistant Professor &amp; Research Scientist, CRL, Department of Biochemistry, Meenakshi Ammal Dental College &amp; Hospital, Chennai. \*Corresponding Author

**ABSTRACT** Cancer refers to a group of diseases that show abnormal growth of cells which can spread to more than one region of the body from the site of origin. Cohort studies are a kind of observational study that samples a particular group of experimental subjects that all share a similar characteristic. Cohort studies on cancer enable researchers to find a pattern in the illness and to make related observations for further analysis. In this project, cohort study data are collected, and a database is developed from the accrued data. Further analysis is conducted using the data collected.

**KEYWORDS :****INTRODUCTION**

Experimental studies generate a great amount of data. It is important for a researcher to arrange, understand, interpret, and present the data. To perceive a pattern from the collected data, statistical analysis is required<sup>[1]</sup>. In this process, first, the data is collected and sorted according to a previously defined parameter. Then, functions are used to interpret the data. These functions can range from simple addition and subtraction to complex calculations using certain formulae. The analysis is done according to the researcher's requirement. The analyzed data is of subjects sharing a similar trait. The shared trait can be the same birth year, then arranged and presented to perceive its pattern. Once the pattern is perceived, it can lead to further experimentation, policymaking in the field of healthcare, etc. A cohort refers to a group occurrence of a particular medical condition, or gender, etc. Cohorts are useful in epidemiological studies, as they are used to collect data of people having certain disease sharing some characteristic, like being located in the same area, risk behavior<sup>[2]</sup>, etc. Cohorts can be of two broad types:

**Birth Cohort:** It is defined as a cohort wherein the subjects are born in the same year. The birth year is, thus, the shared characteristic in the cohort, e.g., people born in the year 2000 represents a birth cohort<sup>[3]</sup>.

**Adult Cohort:** It is defined as a cohort wherein other parameters are shared, instead of the birth year, e.g., people suffering from a particular disease; people living in a particular study area, etc.

In the field of sciences, with special reference to biological sciences and medicine, cohort studies try to provide proof of causative influence on occurrence of a disease and try to show a possible Cause-Effect relationship.

**Advantages of Cohort Data**<sup>[4]</sup>

**Accuracy:** As cohort data is collected over a particular defined time, it is usually more accurate.

**Tempo Effects:** Cohort data is not affected by tempo effects, unlike period data.

**Disadvantages Of Cohort Data**

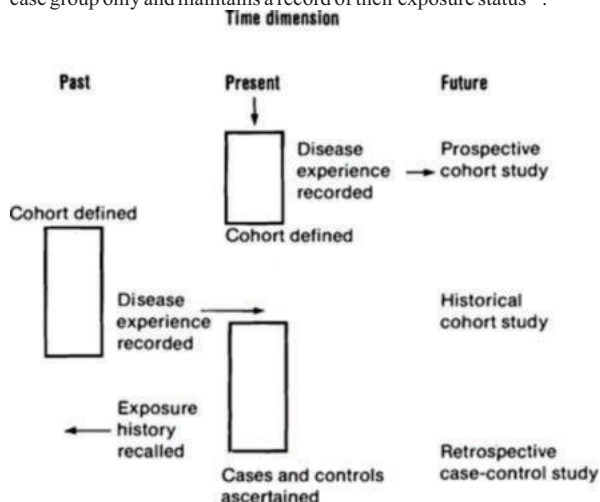
**Time-consuming:** Collection of cohort data takes a long time.

**High Cost:** As the collection of cohort data takes time, a steady source of funding is required throughout

**Cohort Study**

Cohort studies are a type of longitudinal study which samples a defined cohort and performs cross-sectional observations at particular time intervals. Cohort studies are used in epidemiological studies. They are used to understand the risk factors that result in the occurrence of a particular disease. This, in turn, helps in the formulation of a possible treatment that can be implemented to combat the disease. In a cohort study, pre-existing factors, like exposure, are identified and these have to be present in the participants. The subjects are then followed over time to observe the incidence rate of the outcome, i.e., the disease selected in the epidemiological study<sup>[5]</sup>.

As cohort studies are difficult to conduct, a case-control study is used more often. In a case-control study, the case group contains individuals suffering from the particular disease and the control group contains individuals who do not show evidence of disease. Then, their experiences of exposure are recorded. A cohort study usually has the case group only and maintains a record of their exposure status<sup>[6]</sup>.



**Fig:** Difference between cohort and case-control studies

Many studies employ both methods in the same experimental framework. In this hybrid design, the cohort contains subject group and the disease experience and exposure experience is collected via the case-control method. This in turn, makes the study more accurate due to the cohort structure and it also becomes cost-effective as the time and funding is spent only on the subjects that are most informative<sup>[7]</sup>.

**Cohort Effect** refers to the influence of a group's experience on the outcome of the experiment. The group could belong to the same age group, the same location, etc. This characteristic is shared by the entire group and it has an effect on the entire cohort. This effect is important in epidemiology. Cohort studies are conducted to test for the cohort effect<sup>[8]</sup>.

**Cohort Studies on Cancer: An Indian Perspective**

Cancer epidemiology cohorts are large observational studies of populations in which groups of subjects belonging to a particular cohort, i.e., sharing certain characteristics are followed to observe the incidence of new cancers. Cohort data has helped researchers understand the etiology of cancer and this has helped them understand the disease better and research ways to possibly prevent or combat it, by avoiding certain risk factors<sup>[9]</sup>.

Cohort studies on Cancer in India have helped to pin-point certain basic characteristics that show a significant pattern and variation of cancer within the country. Primarily, retrospective cohort studies are employed as they are easier to carry out, due to the availability of a database of previously collected information. These studies also help

in minimizing recall and selection bias.

Certain basic criteria must be categorized while preparing database for cohort studies, as follows:

- a) **Cohort name-** The name of the cohort is to be mentioned, to give it a unique identity.
- b) **Cohort description-** A description is required in database for ease in selection and future use.
- c) **Cohort nature-** The nature of cohort, whether Birth or Adult is to be mentioned.
- d) **Location of cohort-** The sampling site is mentioned, which helps in understanding the distribution of specific types of cancer and is useful for future analysis.
- e) **Year of initiation and end of study-** Information about the start and end time of study is also mentioned.
- f) **Cohort initiation and end year-** It gives an idea about the age range of the cohort in question.
- g) **Inclusion Criteria-** Certain criteria for inclusion in a cohort are previously determined. These are called the Inclusion criteria.
- h) **Exclusion Criteria-** Some criteria are not acceptable for inclusion in a particular cohort. They are thus called Exclusion criteria.
- i) **Gender Involved-** Whether male, female or both genders are included in subject, must be mentioned.
- j) **Age Defined-** Minimum and maximum age of eligibility of subjects, is mentioned.
- k) **Total Number of Subjects-** The total number of subjects included in a particular cohort is mentioned. Male and female numbers are mentioned separately.

#### 1. Objectives of the study

- Cohort studies on cancer in India, to find a pattern in the illness and to make related observations for further research work.
- Development of database on cancer.

The Steps Used In The Development Of The Database Are Mentioned As Follows:



**Fig1:** Steps Employed in Development of Database of Cohorts on Cancer

## 2. Methodology

To build the database, the following steps were performed:

#### Screening of Research Articles

Cohort research articles were screened from NCBI, using the keywords- "cancer", "cohort study", "cohort", "India", "malignant neoplasm", "breast cancer", "ovarian cancer", "oral cavity cancer", etc.

#### Data Entry-

The relevant information from the article was entered into the Indian Cohort Registry database. The information added include- cohort name, cohort description and nature, location of cohort, year of initiation and end of study, cohort initiation and end year, inclusion and exclusion criteria, gender, age and total number of subjects. The information on total 60 cancer cohort studies in India is collected from

online database.

#### Data Export-

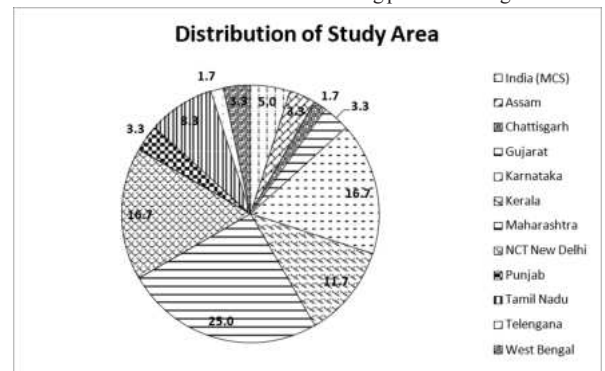
The data entered into the Indian Cohort Registry database, is exported in analytical format like Microsoft Excel. This enables ease in analysis and summarizing of the collected data.

#### Analysis and Summary-

The collected information was analyzed and summarized according to various parameters including State, Gender, Duration of Study, Study Type, Minimum and Maximum Age of Eligibility and Type of Cancer. The data is represented using pie charts for ease in visualization and to establish a pattern.

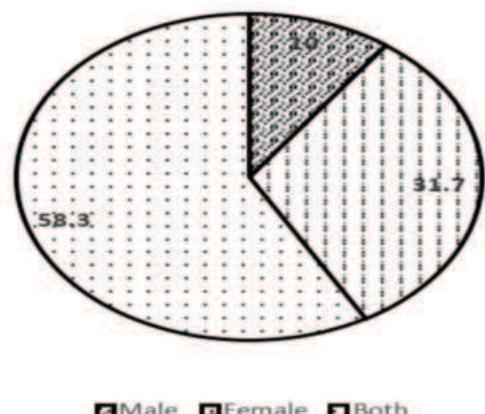
## RESULTS

After collection of data, pie charts and bar graphs were used for effective visualization of data. The following patterns emerged:



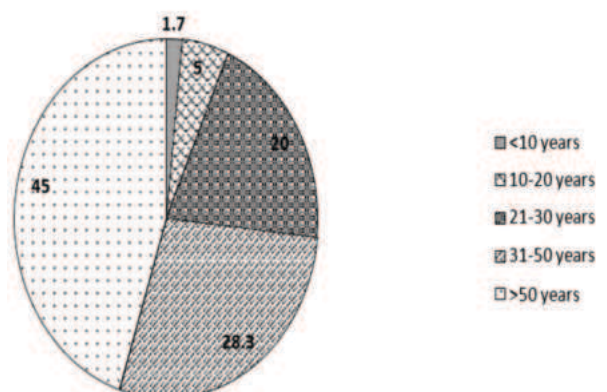
**Fig2:** Pie Chart Representing The Distribution Of Study Area, As Included In The Database

#### Distribution According to Gender



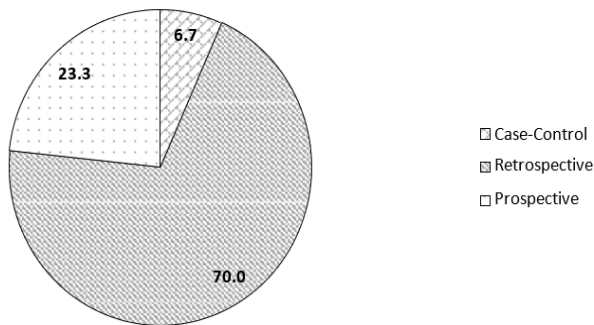
**Fig3:** Pie Chart Representing The Distribution According To Gender

#### Duration of Study Cohort



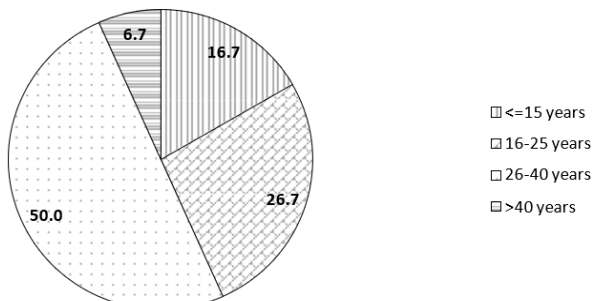
**Fig4:** Pie Chart Representing The Distribution According To Duration Of The Study Cohort

### Type of Study



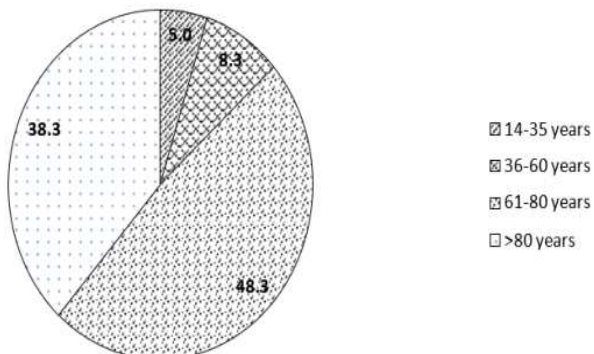
**Fig5:** Pie Chart Representing The Distribution According To Type Of Study

### Minimum Eligible Age of Cohort



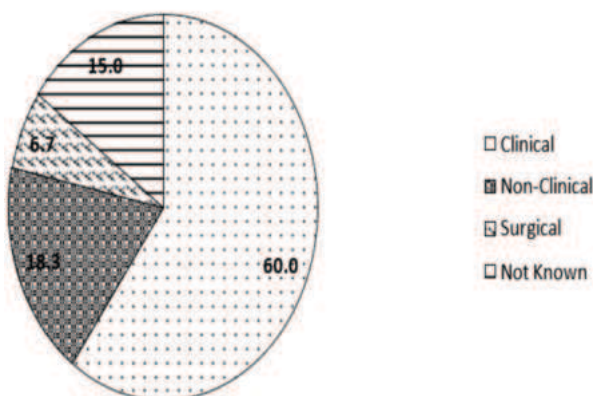
**Fig6:** Pie Chart Representing The Distribution According To Minimum Eligible Age Of Cohort

### Maximum Eligible Age of Cohort



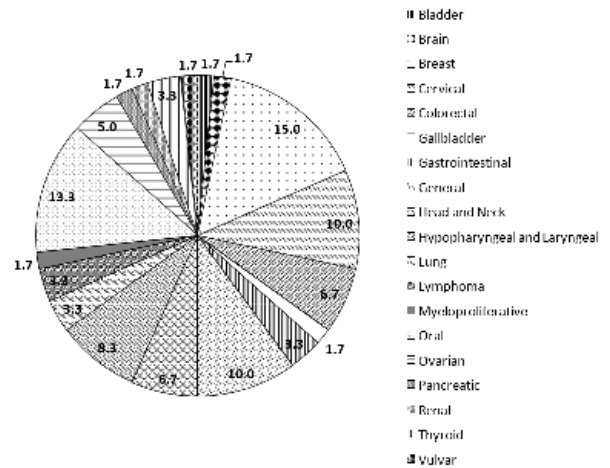
**Fig7:** Pie Chart Representing The Distribution According To Maximum Eligible Age Of Cohort

### Type of Procedure



**Fig8:** Pie Chart representing the Type of Procedure

### Type of Cancer Included



**Fig9:** Pie Chart representing the Types of Cancer Included

### DISCUSSION

#### Distribution:

Most of the Cohort studies on Cancer are conducted in South India. As seen in the data, about 65% of the data is obtained from the south. Fewer studies are conducted in North India and the most number of studies in the North are restricted to the National Capital Region (NCR).

#### Gender:

Recruitment of both genders in a particular study is predominant in cohort studies. Gender-specific cancers, like male-specific (oral cancer due to smoking) and female-specific (ovarian, breast) cancers are also represented in the collected data. Due to the incidence of more female-specific cancers, the representation of female-only cohort data is also relatively high.

**Duration of Study Cohort:** The duration of the study cohort ranges from less than 10 to more than 50 years. This means that a wide age range of subjects is included in the cohorts. This wide age range is beneficial as it helps understand the health effects in different age groups and it can be seen how age may be related to incidence of cancer, under influence of a particular risk factor.

**Type of Study:** Different types of studies are represented in the data. The most highly represented study is the Retrospective Cohort Study. As mentioned before, Retrospective cohort studies are easier to conduct as they are not highly time-consuming and are much more cost-effective. A ready database complete with previously collected patient/subject data is available in a retrospective cohort study. This is the reason why this kind of cohort study is more prevalent, as is represented here.

**Minimum Eligible Age of Cohort:** The age group most represented in the collected data, with respect to minimum eligible age of cohort, is the 26-40 years of age data group. Thus, in most studies, 25+ years of age is taken as inclusion criteria. As we know, due to ethical concerns and in accordance to experimental requirements, adult groups are included more often in experimental studies and the collected data here also represents this pattern.

**Maximum Eligible Age of Cohort:** The age group most represented in the collected data, with respect to maximum eligible age of cohort, is the 61-80 years of age data group. The second-most represented group is the above 80 years of age data group. The maximum eligible age is usually 60-65+ till about 85+. Advanced aged individuals are not usually included in experimental studies due to the possibility of complications.

**Type of Study:** The type of study that is most prevalent in the data is the Clinical type of study. These types of studies usually include biological samples taken from the patients.

**Type of Cancer Included:** The most number of cohort studies on cancer are done on Breast Cancer, Oral Cancer and Cervical Cancer. This supports the fact that breast cancer and cervical cancer are the two most

prevalent cancers among females and oral cancer is the most common cancer among males in India.

### Future Scope

The cohort data updated in the database will be linked to the Clinical Trial Regulation of India and these records will be helpful in performing large-scale cohort studies on all types of disease in India to understand existing risk factors and pattern for cancers. Additionally, this will strengthen the research and help in formulation of policies to combat Cancer.

### REFERENCES:

1. Kuh D, Ben-Shlomo Y, Lynch J, Hallqvist J, Power C. Life course epidemiology [published correction appears in J Epidemiol Community Health. 2003 Nov;57(11):914]. J Epidemiol Community Health. 2003;57(10):778-783. doi:10.1136/jech.57.10.778
2. Sugiyama K, Sugawara Y, Tomata Y, Nishino Y, Fukao A, Tsuji I. The association between coffee consumption and bladder cancer incidence in a pooled analysis of the Miyagi Cohort Study and Ohsaki Cohort Study. Eur J Cancer Prev. 2017 Mar; 26(2):125-130.
3. Noordzij M, Dekker FW, Zoccali C, Jager KJ. Study designs in clinical research. Nephron Clin Pract. 2009; 113(3):c218-21. Epub 2009 Aug 18
4. Terry MB, Flom J, Tehranifar P, Susser E. The role of birth cohorts in studies of adult health: the New York women's birth cohort. Paediatr Perinat Epidemiol. 2009; 23(5):431-445. doi:10.1111/j.1365-3016.2009.01061.x
5. Euser AM, Zoccali C, Jager KJ, Dekker FW. Cohort studies: prospective versus retrospective. Nephron Clin Pract. 2009;113(3)
6. Hood MN. A review of cohort study design for cardiovascular nursing research. J Cardiovasc Nurs. 2009 Nov-Dec;24(6):E1-9
7. Patnaik, J.L., Byers, T., DiGiuseppi, C. et al. Cardiovascular disease competes with breast cancer as the leading cause of death for older females diagnosed with breast cancer: a retrospective cohort study. Breast Cancer Res 13, R64 (2011). <https://doi.org/10.1186/bcr2901>
8. Song JW, Chung KC. Observational studies: cohort and case-control studies. Plast Reconstr Surg. 2010; 126(6):2234-2242. doi:10.1097/PRS.0b013e3181f44abc
9. Matthew B. Weinger, Jason Slagle, Sonia Jain, Nelda Ordonez, Retrospective data collection and analytical techniques for patient safety studies, Journal of Biomedical Informatics, Volume 36, Issues 1–2, 2003, Pages 106–119, ISSN 1532-0464,