SUMMARY OF THE THESIS

In India medical colleges are the main teaching and training centers in medicine. The hospitals associated with each such college provide specialized medical care services to the patients. They generate data of good quality, which could be used for training of medical students, for research of different aspects of medicine and for strengthening medical and health services in the area. These institutes are also undertaking a variety of short and long term research studies based on field, laboratory and hospital data. Because of their very nature they contribute to the different aspects of medical statistics in the country. The purpose of the present study is to place quantitative techniques in the hand of the medical statistician or epidemiologist, illustrated by application to bona fide sets from hospital or registry data. Especially notable quantitative techniques are Generalised Linear Mixed Model (GLIMIX) and logistic models developed for the cancer cases registered in Ahmedabad Urban Agglomeration area during 1988-1994.

The laboratory worker can frequently exclude in which he is not interested and confine his attention to one or more controlled factors at a time, the worker in clinical or preventive medicine is often unable to experiment. He must inevitably use records, which may be influenced by factors which he can not control but have essentially to be taken into account. We can define such statistics, therefore, as 'quantitative data affected to a marked extent by a multiplicity of causes and statistical method as ' method specially adopted to the elucidation of the quantitative data affected by a multiplicity of cause' (Yule and Kendal An Introduction to the Theory of Statistics). More broadly we can regard
statistics as the 'discipline concerned with the treatment of numerical ' data derived from groups of individuals' (P. Armitage, Statistical Methods in Medical Research). Data preparation is, by contrast, a problem of serious proportions in many large-scale investigations on the 'human ' scale. In large-scale therapeutic and prophylactic trials, in prognostic investigations, in studies in epidemiology and social medicine and in many other fields, large number of people may be included as subjects, and very many observations may be difficult to obtain in unambiguous form and the precise definition of the variables require careful thought. There is a temptation to collect more information than is clearly required, in case it turns out to be useful in either the present or some future study. While there is obviously a case for this course of action, it carries serious disadvantages.

In one sense medical statistics are merely numerical statements about medical matters: how many people die from a certain cause each year, how many hospital beds are available in a certain area, how much money is spent on a certain medical service. Such facts are clearly of administrative importance. Numerical facts also supply the basis for a great deal of medical research. Such fact may be found in in official publications of national or international health departments, in the published reports of the research investigations. Present study is concerned with the methodology rather than the factual information, with the applied quantitative techniques in medical research rather than with results of particular studies. This study focuses on the treatment of numeric data derived from group of individuals. These individuals are either patients living in a specified geographic area or healthy normal volunteers for measuring drug effect. The percentage increase or decrease in the morbidity of the particular disease is is an argument for statistical
information. The doctor needs statistical information regarding the occurrence of specific disease during a determined period of time. Quantitative techniques used in measures of disease occurrence such as incidence rate, crude rate, age-adjusted rates and cumulative rates are discussed and computed for the cancer cases just to show that these measures are equally important are used to describe the disease for different situations.

For computing all these measures of disease occurrence, the population of that particular region is more important as it being the denominator. Such data derived normally from population censuses and are carried out at varying intervals of time and on a regular or irregular basis according to county. Problems arise when estimating a population between census years and for the post projecting a population for post central years. In favorable situations the totals given by the censuses are adjusted and updated annually not only by the numbers of births and deaths, statistics generally kept up to date with precision, but also by migration, which is much more difficult to enumerate accurately. Various models are discussed for estimation of population for intercensal years and post-censal years in the present study.

The percentage increase or decrease in the incidence of disease over a number of years is of crucial importance for epidemiologist for associating factors involved in such changes. General Linear Mixed Model with Poisson error has been applied to obtain linear predictor of trend for selected sites. The age-period model is used to predict the trend. The maximum likelihood method is used to predict the linear trend.
Survival data are special and, thus, they require some special quantitative techniques for their analyses. Naturally, we think of survival data as dealing with the time until death, actually the quantitative techniques that are discussed in the present study can be used for data that deal with the time until the occurrence of any well-defined event. In addition to death, that event can be, for example,
Relapse of a patient in whom disease had been in remission.
Death from a special cause.
Development of a disease in someone at high risk.
Recovery of platelet count after bone marrow transplantation.
Relieved from headache, rash, nausea, etc,

What makes analyses of these types of data distinctive is that often there are many subjects in whom the event did not occur during the time that the patient was followed. This can happen for several reasons. Here are some examples:

The event of interest is death, but at the time of analysis the patient is still alive.
A patient is lost to follow up without having experienced the event of interest (death).
A Competing event occurs that precludes the event of interest. For example, in a study designed to compare two treatments for prostate cancer, the event of interest might be death caused by the cancer. However, a patient might die of an unrelated cause instead, such as an automobile accident.
A patient is dropped from the study, without having experienced the event of interest, because of a major protocol violation or for reasons specified by the protocol.
In all of these situations, you don’t know the time until the event occurs. Without knowledge of the quantitative techniques that are described in the present study, a researcher might simply exclude such cases. But clearly this throws out a great deal of useful information. In all of these cases, we know that the time to the event was at least some number.

Human beings are not unique in their responses to some given treatment; there is no doubt that they are likely to be variable. Far, therefore, from arguing that the statistical approach is impossible in the face of human variability, we must realize that it is because of variability that it is often essential. It does not follow, to meet another common criticism, that the statistical approach invariably demands large numbers. The efficacy and safety of medicinal products should be demonstrated by clinical trials. The role of statistical research in the clinical trial design and analysis is acknowledged as essential in International Conference on Harmonization (ICH) guidelines. The proliferation of statistical research in the drug approval process and health care in general necessitate a succinct document on statistical issues related to clinical trials. In the present study principles of quantitative techniques applied to clinical trial for marketing applications submitted to the regulatory authority are discussed. Clinical trial contributing to a marketing application, important statistical issues of its design, sample size determination, conduct of trial and statistical methods to analyze clinical trial data are discussed in detail.

The performance of the drug product is tested; keeping all the other parameters that may influence the pharmacokinetics of the drug constant. Therefore, bioavailability is tested in healthy volunteers, who do not receive any other drug; who are carefully screened for age, height, body
weight, blood chemistry and medical history; and who receive the products to be tested at the same time of day, with identical conditions of fasting, food, activity, and so forth. Statistical concepts used in the estimation of bioavailability and in evaluation of bioequivalence are discussed in the present study. The further analysis of pharmacokinetic data from a bioequivalence study involves the determination of whether the test and reference formulation differ within predefined level of statistical significance. The design used to conduct such studies is discussed as special case of equivalence clinical trials. Statistical quantitative techniques such as 90% confidence interval based on two one-sided t-test and criteria to conclude bioequivalence are summarized in the present study. In practice, because individual subjects may differ widely in their response to the drug, in addition to equivalence in average bioavailability, it is important to compare the variability of bioavailability. If variability of test formulation is much larger than that of the reference formulation, then the safety of the test formulation may be of concern and the exchangeability between two formulations is questionable. The computation of intrasubject variability using the Generalized Linear Model procedures is also defined.

The most recent census (1991) puts the population of India at 843 million. The density of the population is 267 inhabitants per square kilometer (Nanda 1991). The population, which stood at 238.4 million at the beginning of the 20th Century doubled by the time of 1971 census. A country's population is mainly affected by fertility and mortality. In India the system of vital statistics suffers from under reporting and under registrations thus complicating to study the trends of fertility and mortality. Information on such trends is available for census years. Both the birth and death rates have decreased throughout the country.
Relatively high death rate 1911-1912 and low birth rate 1941-1951 are the major deviations from linear trend. The birth and death rates for the year 1986 were 32.6 and 11.1 per thousand respectively, which were obtained thorough a sample registration scheme (C.B.H.I.,1988). The age distribution of a country is a among the best indicators for studying or planning to meet the health needs of population. The age group 0-14 constitutes 39.6 percent of the population in India. The corresponding figure for the United States of America is 21.7 percent. The persons above the 65 years account for 4.3 percent of India's Population whereas the corresponding figure for the USA is 11.9 percent. One of the accepted indicators of the health status of a country is life expectancy at birth. At the beginning of the 20th century, life expectancy at birth was very low in India: 23.6 for males and 23.9 for females. It is not only the size and age composition of the population that has consequences for health planning, but also the spatial distribution. Given that the rural population is typically the most undeserved by health facilities, it is important to study how the population is distributed between urban and rural areas. Whereas in 1921 only 10 percent of the Indian Population were residing in urban areas, this figure had increased by 2.28 times in 1981 (i.e. 23.3%). As mentioned, the Indian System of Vital Statistics is deficient and the information on cause of death is unreliable. Attempts are now being made to collect useful information by other methods. The Registrar General of India have started two programs for this purpose, one for rural and the other for urban areas. The Model Registration Scheme for rural areas obtains information through paramedical workers who are equipped with a list of signs and symptoms. According to this data senility accounts for nearly 22.4 percent of all deaths, the other major causes being 'Coughs' and 'Fever' (CBHI 1988). In urban area, information on causes of death is being collected from selected major hospitals using the 'Medical
Certified Cause of Death'. According to this data, the major causes of death are 'Infectious and Parasitic disease', 'Disease of Circulatory System' and 'Disease of Respiratory System'. Neoplasm’s accounts for 4.1 percent of all deaths (CBHI 1988). This information suffers from coverage error as the number of hospitals reporting the information is not the same each year.

MEASURES OF DISEASE OCCURRENCE

Measures of disease occurrence can describe either the pool of existing cases, or the occurrence of new cases. Measures of prevalence describe what proportion of the population has the disease in question at one specific point in time. Measure of incidence; on the other hand, describe the frequency of occurrence of new cases during a time period. It is useful to think of each individual as being in one of two "states" : diseased or disease-free. In this framework the prevalence measure describes the proportion of the population that is in the diseased state at a specific time. The incidence measure describes the rate of flow from the disease-free state to the diseased state. In epidemiology studies where the aim is to explore casual theories or to evaluate effects of preventive means, the interest is focused on the rate of flow of cases from the disease-free state to the diseased state. The relevant measure of disease is occurrence, therefore, is incidence. Measures of prevalence may be relevant in connection with the planning of health services or in assessing the need for medical care in a population.

The usual method of combining age-specific rates for comparison across different population is that of direct standardization (Fleiss, 1973). The directly standardized (adjusted) rate consists of a weighted average of the age-specific rates for each study group, where the weights are chosen to
be proportional to the age distribution of some external standard population. Hypothetical standard populations have been constructed for this purpose, which reflect approximately age structure of world populations (Waterhouse et al., 1976). An alternative and even simpler summary measure is the cumulative incidence rate, obtained by summing up the annual age-specific incidences for each year in the defined age interval (Day, 1976). Elandt-Johnson et al., 1975 modeled the instantaneous incidence rate at time to as the rate of increase in net risk, expressed relative to the proportion of the population still at risk.

In Cancer Incidence of Five Continents Vol-I, (Doll et al., 1966) three standard populations were used and the separate standardized incidence rates were calculated for each type of cancer reported by each of registry. These three standards varied from an 'African' population with a low proportion of old people through an intermediate 'World' population to an 'European' population with a high proportion of old people. In volumes II & III (Doll et al., 1970; Waterhouse et al., 1976) the same three standard populations were used and in addition the truncated standardized incidence rates were computed. In Vol IV, V and VI (Waterhouse et al., 1982, Muir et al., 1987) rates are presented which have been standardized to the world population and to the truncated population only.

**CANCER MORTALITY**

In the absence of information concerning the occurrence of cancer on a defined population mortality data may be helpful. The mortality rates provide, at best, a crude index to variation in the incidence of cancer in different geographic and social environments. In general, malignant disease has a relatively high fatality, analysis of mortality data provide information on the nature and extent of cancer. The estimated 51 million
deaths in the world non-communicable disease such as cancer and heart disease account for about 19 million or about 36% of the total. Non-communicable diseases are also emerging as a major cause of death in the developing world. Cancer accounts for 6 million or 12% of deaths globally some 58% of them in the developing world. Cancers of airways and lungs mostly caused by cigarette smoking are the leading cause with 1 million deaths followed by cancer of the stomach with over 700,000 deaths. Respiratory diseases such as chronic bronchitis Kill 3 million people a year. Cancer killed some 6 million people in 1993. The majority of the deaths occurred in the developing world, although cancer is widely perceived to be a disease of industrialized nations. In the developed world there were around 2.5 million cancer deaths, contributing some 22% of the total. By contrast the 3.5 million cancer deaths in the developing world made up just 9% of the total. There were estimated to be some 20 million-cancer sufferers in the world in 1994 with a million new cases a year.

An analysis is carried out in 1993 by the International Agency for Research on Cancer (IARC) looked at worldwide cancer mortality trends in 24 geographical areas, using 1985 data. World wide, lung cancer was the biggest single killer in men, accounting for 22% of cancer deaths. In women breast cancer was the main cause of deaths in developed world after cervical cancer. Another frequent cause of death for both sexes was stomach cancer, followed by liver cancer in men and colon cancer in women. The researchers estimated that 20% of all cancer deaths could be prevented if tobacco smoking was eliminated. Deaths from cancer of liver and of the cervix, both major problems in the developing world could be substantially reduced by vaccination against hepatitis B and the introduction of cervical smear test programme. If all countries were able
to introduce an effective community wide screening programme as in Finland, 76% of the world’s deaths from cervical cancer expected to reach around 276000 by the year 2000 could be prevented. A vaccination campaign against Hepatitis B in countries with high rates of carriers could reduce deaths from liver cancer is expected to kill some 296000 men and around 137000 women by end of the century.

PROJECTED INCREASE IN CANCER INCIDENCE AND DEATH RATES BETWEEN 1991-2026.

The life expectancy of Indian Population is rapidly increasing. During the last 50 years the life expectancy has increased from 30 to 60 years. In order to plan a cancer control programme for India it is essential to make a projection of the quantum of increase in incidence in mortality from cancer that can be expected in the coming decades. The projected increase in life expectancy of the Indian Population for incidence and death rates from cancer during this period is discussed. It is estimated that life expectancy of the Indian Population will increase from 60.60 for males and 61.70 for females during the period 1991-1996 to 69.60 males and 71.00 females during the period 2021-2025. With this increasing longevity, the proportion of the Indian Population in the cancer age will increase substantially. For example, there will be a 4.7 fold increase in the population aged 50 and above between the period 1991 and 2026. The cancer incidence and mortality figures will consequently increase with the increasing aging population. As the estimates projected, the total cancer burden in India for all sites will increase from 6,02,095 new cases per year to 14,14,115 between 1991 and 2026, an increase of 2.68 fold. Similarly during this period cancer deaths will increase by a factor of 2.78.
By the year 2000 there are expected to be about 4 million deaths annually from cancer in males worldwide, up from 3.1 million in 1990 and 3.2 million female cancer deaths up from 2.6 million of the projected male cancers around 2.3 million will occur in the developing world and 1.6 million in the developed of the female cancers almost 2 million will occur in the developing world. WHO estimates that in 25 years time new cases of cancer each year (not deaths) in developed countries will have increased form roughly 4 million today to 5.5 million. In the developing countries the figure will have doubled from 5 million to 10 million. Almost two thirds of cancers over the next 25 years will occur in the developing world. Approximately 20 million people are alive with cancer by 2015 there will probably be more than 30 million.

CLINICAL MEDICINE

The essence of an experiment in the treatment of a disease lies in comparison. To the dictum of Helmholtz that ‘all science is measurement,’ we should add, as that great experimenter Sir Henry Dale pointed out, a further clause, that ‘all true measurement is essentially comparative.’ On the other hand there is a common catch phrase that human beings are too variable to allow of the contrasts inherent in a controlled trial of remedy. Yet if each patient is ‘unique’ it is difficult to see how any basis for treatment can be sought in the previous observation of other patients-upon which clinical medicine is founded. In fact, of course, physicians must, and do, base, their ‘treatment of choice’ upon what they have been happen before – whether it be in only two or three cases or in a hundred. However, though, broadly speaking, human beings are not unique in their responses to some given treatment, there is no doubt that they are likely to be variable. Two or three observation may therefore give, merely through the customary play of chance, a favorable
picture in the hands of one doctor, an unfavorable picture in the hands of another. As a result, the medical journals become an arena for conflicting claims – each in itself, maybe perfectly true of what the doctor saw but insufficient to bear the weight of the generalization placed upon it.

In order to establish bioequivalence between two formulations in a crossover trial, it is common to assume a mixed-effect analysis of variance (ANOVA) model and perform two one-sided tests. When the analysis is done on the untransformed data, the numerators of the test statistics are not, in general treatment contrasts. Consequently, the standard errors of the numerators are difficult to compute. The usual practice is to approximate these with the standard errors of treatment contrasts. Hsuan FC 16 examined the goodness of this approximation. He derived Best Linear Unbiased Estimator (BLUE) for the treatment means as well as its covariance matrix. Due to the presence of the intersubject variability, the variances and covariances of the BLUE of treatment means are much larger than is commonly believed.

The applied statistician often encounters the need to compare two or more groups with respect to more than one outcome or response. Several options are generally available, including reducing the dimension of the problem by averaging or summarizing the outcome, using Bonferroni or other adjustments for multiple comparisons, or applying a global test based on a suitable multivariate model. For normally distributed data it is well established that global tests tend to be significantly more sensitive than other procedures. While the global tests have also been proposed for multiple binary outcomes, their properties have not been well studied nor have they been widely discussed in the context of clustered data. Lefkopoulou M and Ryan L 20 derived a class of quasi-likelihood score tests for multiple binary outcomes, and show that special cases of this
class correspond to other tests that have been proposed. They discussed extensions to allow for clustered data, and compare the results to the simple approach of collapsing the data to a single binary outcome, indicating the presence or absence of at least one response. The asymptotic relative efficiencies of the tests are shown to depend not only on the correlation between the outcomes, but also on the response probabilities. Although global tests based on a multivariate model are generally recommended, these findings suggest that a test based on the collapsed data can maintain surprisingly high efficiency, especially when the outcomes of interest are rare.

Much of the health progress made in the past century has been a result of the conquest of infectious diseases, often by environmental means or through the use of simple preventive or therapeutic measures such as vaccines and antibiotics. Interventions are generally not so simple, however, for the problems that are currently the most important. Many current killers and cripples are heavily affected by behavior. The principal risk factors associated with unnecessary deaths or potential hours of life lost are tobacco, injury, lack of prevention services, high blood pressure and improper nutrition.

Qualitative evaluation of the weight of evidence from human, animal and other studies does not make full use of this range of knowledge; it assess whether there is a hazard but no how great the risk. Many groups in society including scientists want to know more than simply whether cancer risks exist, and are pressing for information to guide them when they set priorities for the control of cancer causing agents in the environment Industrial groups want to quantify risks in order to estimate the benefits (and costs) of changing current work practices. Courts of
law seek quantity estimates the benefits (and costs) of changing current work practices. Courts of law seek quantitative estimates of cancer risks to assist decisions on liability for individual cases of cancer.

In developed countries, the intensity of formal regulation of health-endangering exposures has increased during the last 30 years. Quantitation of cancer risks assists the control of cancer hazards by providing policy-makers with information about the magnitude and gradient of risk. This information, coupled with population exposure profiles, enables an assessment to be made of the relative importance of each carcinogen in relation to other carcinogens and other hazards "competing" for regulatory attention and resources.

Quantitative estimates of risks may be subject to many sources of uncertainty. For example, experimental and sampling error as well as random and systematic measurement error in both exposure and response varieties all contribute to uncertainty in risk estimation. Risk can also vary appreciably among individuals in the population of interest, as illustrated in figure 1.1. Even though estimates of cancer risk are subject to both uncertainty and variability, quantitative analyses can often provide a clearer basis for risk management decisions than qualitative evaluations of (known, possible or probable) carcinogens.

Since the 1940s, industrialization and the proliferation of synthetic organic chemical have resulted in a myriad of actual and potential health-endangering exposures. In many industrialized countries, cancer, vascular disease, and other chronic conditions have now replaced infectious diseases as the major causes of mortality. This has led to a
Mathematical models were used to estimate the dose-related excess lifetime cancer risks (including the estimated upper-bound excess risk) for humans, based on the dose-response curve obtained in animal bioassays and taking into account differences in species sensitivity to carcinogen exposure (Crump et al. 1976). Subsequent improvements in cancer modeling have come about through awareness of the process involved in carcinogens and improved modeling of tissue dosimeter. Model for the impact of an environmental agent on the risk of disease in a population is discussed in the present study. The impact of an environmental agent on the risk of disease in a population will depend not only on the strength of its effect in the exposed sub-population but also if the agent is a very potent carcinogen, its impact on the cancer burden of the entire population will be small if only a small fraction of the population is exposed. On the other hand, if exposure to a week carcinogen is widespread, the population impact could be substantial. A measure of risk that attempts to quantify the population burden of disease due to a spetion (PAF), which is defined as the fraction of all cases in the population that can be attributed to the exposure, and is given by the expression 

\[ PAF = \frac{(I_T - I_u)}{I_T} \]

where \( I_T \) is the incidence in the total population exposed to the agent of interest.

Quantitative Estimation and Prediction (QEP) models for the analysis of epidemiological data specially focused on RR models are presented in this chapter. In epidemiological studies, modeling is carried out to estimate the risk of cancer as a function of the exposure of interest and of the host and environment factors which may modify risk.
Epidemiological studies used for quantitative estimation of risk should generally encompass a range of exposure levels to permit characterization of the relationship between exposure and risk. The two main types of studies which provide data for this purpose are: (i) Cohort studies, in which a group of persons with a range of exposure levels is followed, for mortality or morbidity, from a particular disease; and (ii) case-control studies, in which the exposure history of all cases and appropriate controls is reconstructed. The most common measures of risk used in QEP are the age- and time-specific "absolute" and "relative" risk. Both of these measures can be expressed as a function of the level of the exposure of interest as described below. Absolute risk (AR) cannot be estimated from case-control studies without supplementary information on the level of risk in unexposed individuals. Relative risk (RR) can, however, be estimated from both case-control and cohort studies. Most of the developments in empirical QEP models for the analysis of epidemiological data have focused on RR models.

Statistical models may also be used to summarize dose-response data from laboratory studies. For the most part, this chapter focuses on experiments involving a series of increasing dose levels, including an unexposed control group, the dose level being held constant throughout the duration of the experimental period (Krewski & Goddard, 1990). Most long-term animal experiments encompass the greater part of the expected lifespan of the test-species, which is typically 2-3 years for rodents. In describing such models, it will be convenient to distinguish between models used to describe the lifetime probability of cancer and those which are used to describe the temporal patterns
of tumour incidence. Note that the lifetime probability of cancer will be influenced by the survival rate of the animals in the experiment, early mortality reducing the opportunity for tumour occurrence. Dose-response models used to describe the relationship between the lifetime probability of cancer and dose are referred to as quintals response models since the response of interest (the presence or absence of a given tumour in a given animal during the course of the study) is a binary random variable. Krewaski & Van Ryzin (1981) have published a detailed review of quantal response models for toxicological data, including carcinogenicity data, and such models have also been discussed by Finney (1971), Govindarajalu (1988), and Hubert (1992).

The strength of an agent with carcinogenic potential may be expressed in terms of quantitative measures of carcinogenic potency, first mentioned in the scientific literature in the 1930s. Twort & Twort (1930, 1933) examined the times at which tumours appeared during the course of animal experiments, and used the time at which 25% of the animals developed tumours as a measure of carcinogenic potency. More recent developments in measuring carcinogenic potency are related to the TD50 index introduced by Peto et al. (1984) and Sawyer et al. (1984). Formally, TD50 is defined as the dose that will halve the proportion of tumor-free animals at a specified point in time.

The Armitage – Doll model, which has been used extensively in the last four decades, was first proposed to explain the observation that in many human carcinomas, age-specific incidence rates increase roughly with a power of age. The age-specific incidence rate is a measure of
the rate of appearance of tumors in a previously tumor free tissue. The appropriate statistical concept is that of the hazard function is discussed.

The main aim of a registry is to collect, insofar as this is possible, the total number of cases in a defined area. Information on the number of persons at risk in the total population covered by a registration area is generally obtained from official demographic sources such as the institute of statistics, regional, national or international. Such data derived normally from population censuses and are carried out at varying intervals of time and on a regular or irregular basis according to county. Problems arise when estimating a population between census years and for the post projecting a population for post central years. In favorable situations the totals given by the censuses are adjusted and updated annually not only by the numbers of births and deaths, statistics generally kept up to date with precision, but also by migration, which is much more difficult to enumerate accurately. The following population models are derived in detail:

Single - Species Non-Age-Structured Population Models
Stochastic Models
Linear Birth - death - immigration - emigration processes
Age Structured Population Model
Linear Continuous time Model

The following demographic methods for population estimates for intercensal years are clearly determined and described:
Simple Diagonal Method
The Diagonal Method With Deaths
The Diagonal Method With Mortality Indices.
Recent advanced statistical technique called Generalized Linear Mixed Model with Maximum Likelihood Method and error term Poisson of SAS System Release 8.1 (SAS Institute Inco., USA) is used to assess the trend of selected cancer site in Ahmedabad Urban Agglomeration Area. The mixed models have two features in common. First, the errors are assumed to be normally distributed. Second, the response variable is equated directly to a linear combination of fixed and random effect. However, in many particle situations the response variable of interest may not have a normal distribution. In other cases, there may be restrictions on the range of allowable values for predictable functions that direct modeling of the response variable cannot address. For this purpose the site wise cancer incidence data from the Annual reports of The Gujarat Cancer & Research Institute is used for the years 1988-1994. Various measures of disease occurrence are derived computational formula. Rates have been computed for the cancer cases. Most appropriate measure of disease occurrence is Age Standardized Rate processed under World Standard Population. The salient feature of this measure is its comparability with international registry’s data. Linear Mixed Model with Poisson error has been applied to obtain linear predictor of trend for selected sites. The age-period model is used to predict the trend. During the period under review 14,225 new cancer cases were detected among residence of Ahmedabad of whom 8,124 were males and 6,101 were females. The age standardized rates per 100,000 were 115.5 for males and 88.0 for females. The five most common cancers among males were Lung(10.5), Tongue(9.2), Oesophagus(6.39), Hypopharynx(6.12) and Larynx(5.7) and Breast(17.4), Cervix(12.3); Oesophagus(4.8), Ovary(3.6) and Lung(2.6). The single most common site of cancer in either sex is cancer of Breast constitutes over 20% of
cancers in women and account for 8.6% of all cancers in either sex. The Maximum Log Likelihood Estimates revealed that in males cancer of Tongue and Lung showed increasing trend of 2.9% and 3.7% and in females cancer of Breast and Cervix also showed increasing trend of 8.9% and 5.2% respectively over a period seven years from 1988 to 1994.

Time trends in age standardized rates, however, are strictly interpretable only when only if the effect of calendar time on cancer risk is multiplicative i.e. when the age incidence or age-mortality curves for successive calendar periods are parallel on a logarithmic scale. Trends observed in age-standardizes rates may be misleading when the time trends in different age – groups are significantly different from one another. In these circumstances, the evolution of cancer risk with time is more readily understood by examination of trends in the age-specific rates. Such an approach can be satisfactory for the detailed examination of a single data set or a small number of homogenous data sets. Examination of age-specific rates alone can suggest the existence of a cohort effect, for example with a rapid increase in risk at younger ages, and progressively smaller rates of increase, or even decline, in successively older age groups. When the age-specific trends are summarized only by rates of increase. However, much useful information is ignored: further, the estimates of age-specific trends are carried out independently of each other, using only that portion of the data corresponding to the relevant age group.

The age-period-cohort modeling is described for smoothing purpose in the present study. This choice is motivated by the desire to provide both trends in the age-standardized rates with calendar period and trends in the cumulative risk with year of birth. This approach provides the advantage
of a simple and efficient description of time trends when the data may be described by a smooth function.

Survival data are special and, thus, they require special methods for their analyses. Without knowledge of the methods that are described in the present study, a researcher might simply exclude such cases. In all of these cases, we know that the time to the event was at least some number. A subject's observed time, $t$, is right censored if, after time $t$, he or she is known to still be alive. Thus you know that this subject's survival time is at least $t$. The concept of censoring and its important is determined. Although this study focuses on right-censored data. Distribution functions, Survival function and Hazard functions are described in this Chapter. Some commonly used survival functions are discussed. The Cure Model and Mixed Model are also determined. Kaplan Meirer and the Log Rank Statistics are discussed in detail through an illustration. Estimates of the Variance are obtained using Greenwood's Formula and Peto formula. Hypothesis and confidence interval based on asymptotic normality are derived.

The broad aim of the process of clinical development is to find out whether there is a dose range and schedule at which the drug can be shown to be simultaneously safe and effective, to the extent that the risk–benefit relationship is acceptable. The concepts of the development plan, confirmatory trial and exploratory trials are discussed in detail in this study. The clinical relevance of primary variable or target variable or primary end point is described in this Chapter. Importance of secondary variables and secondary objectives with respect to the primary objective and primary endpoint are also described.
The essential aim is to prevent identification of the treatments until all such opportunities for bias have passed. For this purpose, blinding or masking to limit the occurrence of conscious and unconscious bias in the conduct and interpretation of a clinical trial arising from various influences discussed in detail. Randomization introduces a deliberate element of chance into the assignment of treatment to subjects in a clinical trial. Generation of Randomization schedule is the salient feature of this chapter and is illustrated using SAS System Release 8.1 for the parallel design. One of the distinguish advantage of his randomization schedule is its tractability.

An interim analysis in any analysis to compare treatment arms with respect to efficacy or safety at any time prior to formal completion of a trial. Under certain conditions, it is convenient and sometimes prudent to look at data resulting from a study prior to its completion in order to make a decision to abort the study early or to increase the sample size, for example.(Pharmaceutical Statisitics ,Sanfoec Bolton). Significance Levels for Two-Sided Group Sequential studies with an Overall Significance Level of 0.05.(according to O’Brien and Flemming) is discussed in detail as a special case in this chapter. Issues like adjustment of sample size, breaking of randomization procedure etc. are also addressed in the present study.

The number of subjects in a clinical trial should be large enough to provide a reliable answer to the questions addressed. This number is usually determined by the primary objective of the trial. Determination of the sample size is illustrated through systolic blood pressure (SBP) example. a SAS program is developed for the computation of sample size and power determination. Choice of the design and the statistical methods
for the analysis of data are two important aspects in planning clinical trials. In this chapter an attempt has been made to determine appropriate design for a bioavailability/bioequivalence trials. We then introduce some designs that are currently available for bioavailability. Statistical methods currently available for the assessment of average bioequivalence are also provided. Method to analyze clinical trial data from parallel design is also provided in this study.

Alongside information on incidence and mortality, survival statistics are a means of quantifying the effectiveness of these two interventions at the population level. Thus, information on survival has long been recognized as an important component in monitoring cancer control activities (WHO/IARC). Like all other health indices, survival statistics are useful primarily as comparative measures showing how survival differs between different populations over time, and between population subgroups (defined by, for example, age, sex, ethnicity or socioeconomic status). It is these comparisons that help us to suggest possible reasons for the variations and provide targets for improvement and a means of monitoring progress towards them. The types of survival, estimation of survival, relative survival, age-standardized survival are the topics considered in this chapter. Factors associated with cancer survival are also described. Geographical variations in cancer survival among some selected cancer sites are also discussed in this chapter. Techniques for survival analysis with special reference to cancer survival are provided in this chapter. This chapter discusses method for estimation and comparison of survival rates. Methods for estimation of net survival, cause - specific survival and relative survival are also described. Considerations in sources of data, Lost to follow up (LFU) and Loss adjusted survivals rates(LAR) are also discussed in the present study.
Survival estimation is one of the end points in any prospective or retrospective study for evaluation results of treatment. This is particularly important for malignant diseases to optimize treatment, to assess the effectiveness of a given treatment, to understand the biological behavior of tumors and to plan and conduct clinical research. Currently there are a number of methods to estimate survival such as the actuarial method (Berkson and Gage 1950, Cutler and Ederer 1958, Chiang 1968), the product-limit (K-M method) method (Kaplan and Meier 1958) and Cox’s proportional hazards model (Cox 1972). However, the reliability of survival estimation depends to a large extent of the degree of follow-up information available for individual patients in the study group, losses to follow-up (LFU) are one of the major problems in survival analysis. The problem is of paramount significance in developing countries due to a number of factors and reported survival results may carry lack of reliability and the solution for this is discussed in the present study.