STATISTICAL PERSPECTIVES IN CLINICAL AND HEALTH SCIENCES:
THE BROADWAY OF MODERN APPLIED STATISTICS *

by

Pranab K. Sen

Department of Biostatistics, University of North Carolina at Chapel Hill, NC.

Institute of Statistics Mimeo Series No. 2107

October 1992
Summary. Clinical and (public) health sciences constitute one of the most important areas of fruitful adoptions of statistical modelling and analysis. Along with a review of some of the salient features of statistical perspectives in this broadway of applied statistics, some challenging problems are presented.

1. INTRODUCTION

Statistical sciences, albeit of relatively recent origin (as compared to most other basic sciences), have made a remarkable stride during the past five decades. Most of these developments have taken place in the sector of basic theory and methodology pertinent to the traditional areas of applications, such as the agricultural sciences, biometry, medical studies (clinical trials), physical sciences, reliability theory and quality control, demography and population studies, econometrics, social sciences and some other allied fields. Statistical quantum physics paved the way for a sound foundation of some relevant probability theory and related stochastic processes with a variety of fruitful applications. Not only biological sciences have been encompassed in this domain but also the epidemiological and environmental sciences have been annexed to this fertile field of applications. (Civil and electrical/electronics) engineering, industrial engineering, operations research and systems analysis all have added new


Key Words and Phrases: AIDS; Antibiotics; Artificial Intelligence; bio-assays; Bio-medical sciences; Biotechnology; Cancer; Cardiovascular diseases; clinical trials; control; data management; data monitoring; design of experiments; Ecology; Environmental sciences; Epidemiology; Genotoxicity; Health education; Health Policy and Administration; Image analysis; Medical diagnostics; Neural models; Neurophysiology; Occupational hazards; Psychiatry; sampling units; sampling techniques; survey methodology; Statistics; statistical analysis; statistical methodology; statistical modelling; stochastic processes; Tomography; Virology.

* Dedicated to the memory of the late Professor P. C. Mahalanobis on the occasion of his Birth Centenary.
dimensions to statistical theory and methodology as well as wings to computational
statistics. Scores of statistical packages have cropped up to meet the growing
need of statistical analysis in virtually every walk of science and society.

This genesis of applied statistics followed the foresights of the forefathers
who, rightfully, refused to allow statistics to be dwindled into the alleys of
probability theory and mathematical analysis alone. The past two decades have
witnessed a phenomenal growth of research methodology in a variety of experimental,
epidemiological and environmental sciences. In this setup, not only statistical
principles occupy a focal position with respect to model buildings, but also they
have led to genuine adoptations and developments of statistical methodology (and
theory too) for a variety of nonstandard and novel setups. A few years ago (Jan­
uary, 1989 ), at the Indian Science Congress Association Meeting at Madurai (Platinum
Jubilee Lecture in Statistics), the present author emphasized on the
enormous scope of statistics in practically every sphere of science and life:
There are, indeed, many challenging problems and (applied !) statisticians have
the responsibility to resolve these in a way satisfactory to the pertinent fields
as well as statistics. Statistical planning (design) of experiments, statistical
modelling (of the plausible outcome variables) and statistical analysis all play
fundamental roles in this complex perspective.

In order to pose some of these challenging problems in a proper perspective,
perhaps, it will be quite appropriate to have a critical appraisal of our current
standing, to focus on some new areas of applications and then to pose some of these
challenging problems. Afterall, statistical sciences have the genesis in real
applications, and therefore, the strive for real success also rests on the scope
for further accomodation of novel areas of fruitful applications. As such, we
proceed by examining first the traditional frontiers of statistical sciences with
a view to focusing on the main fortress of applied statistics. A composite picture
of this is presented below.
Although the above picture gives us a broad view of applied statistics with its link to other departments, it does not convey the full breadth of this vast area. The following nomenclatures pertain to different aspects of applied statistics:

- Actuarial statistics
- Agricultural statistics
- Applied statistics
- Biometry
- Biostatistics
- Data management
- Economic statistics
- Ecology
- Environmental statistics
- Epidemiological statistics
- Eugenics
- Experimental statistics
- Industrial engineering
- Operations research
- Systems analysis
- Medicinal statistics
- Statistical chemometrics
- Quality control

Moreover, applied statistics incorporates theory and methodology from its mathematical counterpart, and hence, probability theory, mathematical statistics and stochastic processes have all become an integral part of this scenario. In the
sequel, with due emphasis on (public) health and clinical sciences aspects, I will confine myself to a corner of the Fortress in Figure 1 with occasional trespassing to other corners. In Section 2, various sectors within the focused domain of applied statistics are introduced. Some of these are elaborately presented in Section 3 where the challenging problems are discussed too. The last section deals with some general observations and calls for further developments in statistical methodology to suit these applications.

2. SOME ASPECTS OF APPLIED STATISTICS

The developments in various areas of applied statistics have taken place according to the specific needs of the field of applications. However, there has also been general attempts to cover theory and methodology which would be applicable in a broad class of problems. Thus, it may be convenient to enlist (at least a partial) list of areas where such statistical methodology remains potentially useful. This list is by no means complete. Rather, it portrays a comprehensive picture of the vast domain of applicability of statistical reasonings.

A. ANTHROPOMETRY, AGRICULTURAL SCIENCES, ARTIFICIAL INTELLIGENCE
B. BIOLOGICAL SCIENCES, BIOMETRY
C. COMPUTER SCIENCE, CLINICAL TRIALS, CRIMINOLOGY
D. DEMOGRAPHY, DATA COLLECTION AND ANALYSIS, DAIRY SCIENCES, DRUGS
E. ECOLOGY, EPIDEMIOLOGY, ENVIRONMENTAL SCIENCES, EPILEPSY, ECONOMICS, ENERGY
F. FAMILY PLANNING, FOLK MEDICINE, FORESTRY, FOOD TECHNOLOGY
G. GAMBLING, GENETICS, GEOGRAPHY, GEOSCIENCES, GENETIC TOXICITY (MUTAGENICITY)
H. HEALTH SCIENCES, HYGIENE, HYDROLOGY, HUMANITY
I. IMMUNOLOGY (AIDS), IMAGE PROCESSING (DNA), INDUSTRIAL RESEARCH
J. JURISPRUDENCE, JUVENILITY
K. KINEMATICS, KINETICS
L. LEPROSY, LINGUISTICS, LANDSCAPING, LEGISLATURE
M. MANAGEMENT SC., MARINE SC., MATHEMATICS, MEDICAL SC., METEOROLOGY
N. NATURAL SCIENCES, NEURO-BIOLOGY, NUTRITION
O. ONCOLOGY, OPERATIONS RESEARCH, OCEANOGRAPHY, OCCUPATIONAL HAZARDS
P. PHYSICAL SCIENCES, PUBLIC HEALTH, PSYCHOLOGY
Q. QUALITY CONTROL, QUANTITATIVE SCIENCES, QUARRYING, QUARANTINE
R. RADIATION EFFECTS, RELIABILITY (SYSTEMS ANALYSIS)
S. SOCIAL SCIENCES, SEMIOLOGY, SEXOLOGY, SURGERY
T. TOXICOLOGY, TELEPATHY, TRAFFIC ENGINEERING, THROMBOSIS
U. UNCERTAINTY, UNEMPLOYMENT, UNESCO, UNIVERSE
V. VETERINARY SCIENCE, VENEREAL DISEASES, VIROLOGY
W. WAR SCIENCES, WATER, WRESTLING
X. XEROGRAPHY, XENODERMA
Y. YOGA, YACHTING
Z. ZOOGEOGRAPHY, ZOOLOGY, ZYMOLGY
Since space and time are a premium, let me pick up a handful of these areas and illustrate some of them in some detail. Towards this, we have the following:

A :  AIR, AQUA (water), ARTERIOSCLEROSIS, ARTIFICIAL INTELLIGENCE
E :  ENERGY, ECOLOGY, ENVIRONMENT, EPIDEMIOLOGY
I :  IMMUNOLOGY (AIDS), IMAGE PROCESSING (DNA)
O :  ONCOLOGY, OCCUPATIONAL HAZARDS
U :  UNCERTAINTY, UNIVERSE (Ozone / Greening Effect)

We start with a glossary of ENVIRONMENTAL Problems which encompass a greater amount of these areas. Since most of these problems have become quite noticeable in the current time, we avoid any further description of the actual details.

I. ENVIRONMENTAL POLLUTION:

* Taj Mahal, Agra, India; Venice, Italy; .......... Air-borne Sulphur deposition; Oil Refinaries.
* Love Canal, USA, Marina in India, New Jersey, USA; Berlin Waterfront; Northern Germany Rivers; .......... Chemical dumpings; Water Contaminations.
* Chernobyl, Russia; Three miles Island, USA; Savannah River Nuclear Plant, S.C., USA. Radiation Effects: Water and Air Contaminations.
* Hiroshima & Nagasaki, Japan Nuclear Bomb (Radiation) Casualty.
* Nuclear Power Plants' Waste Deposites Sites: All over the Globe. Subsoil Water Contamination due to Radiation effects.
* Thermal Power Plants in India and elsewhere. Air-borne Carbon Deposition: Black Lungs!
* Chemical Warfare in Iraq-Iran War; Agent Orange in Vietnam Spontaneous Casualty and long term War; Burning Oil-Fields in Kuwait. Epidemiological Hazards.
* Bhopal Episode (India) and others.... Catastrophic Mortality & Morbidity due to Chemical actions.
* Automobiles, Airplanes, Factories, ... Inhalation Toxicology: Respiratory Diseases.
* Industrialized Nations Greening Effect: OZONE Layer Problem
* Deforestation: Natural or Deliberate Actions. Cooling Effects; Erosion.
* Acid-rain, ... Atmospheric Pollution Recycles.
* Solid Waste Landfill Problem in any Urban Area.
* RADON in the Basement It is no longer at Random!
It is quite clear from the above list that Environmental and Epidemiological aspects are often inseparable. There is a genuine need to have a comprehensive look into the integral aspects of such problems, and in this context, (Bio-)statistics is indispensable. Public Health has its main focus on environmental and epidemiological impacts, albeit its foundation rests on statistical considerations as well as clinical aspects. Health Administration, Health Education, Health Services, Nutrition, Maternal and Child Health all rely heavily on clinical aspects, but Biostatistics is the binding force which relegates them to the proper domain of Public Health. In this setup, Demography (Vital Statistics/Population Studies) has become an integral part of Biostatistics. We shall have the occasion to bring in these issues at a later stage. We present next some of the Clinical Sciences problems which have significant statistical components, and these are as important as the Environmental ones presented before.

II. Clinical Sciences Perspectives

Cardio-vascular Diseases
What are the Hereditary and Environmental Factors? Can they be controlled?

Cancer Etiology
What are the medical and statistical interpretations?

AIDS
Can Statistics and Medical Sciences work together?

Chronic Diseases
Can Statistics be neglected?

Medical Studies
Whither Clinical Trials? Whither Statistics in Clinical Trials?

Biotechnology
In our Dream of Super Humans can Statistics be totally ignored?

Genotoxicity
Assessment of Risks: Statistical?

Clinical Neurosciences
Deterministic vs. stochastic factors!

Empirical Evidence in Clinical Studies
How far to rely on Statistics in this context?

Medical Diagnostics
Whither Statistics?

Image Processing (DNA)
Can Statistics be denied?

Drug Information & Commercialization.
Can Statistics be neglected?

Clinical Trials
An hybrid of Statistics & Medicine?

Antibiotics
Losing Potency!

Psychiatry
Tip of the mental Iceberg!
It is possible to continue with an uncountable number of such major issues in health and clinical sciences. However, in our endeavor to assert the basic statistical perspectives, we would prefer to treat some of these problems in isolation and then to draw some general statistical conclusions. This will be pursued in the next section.

3. STATISTICAL PERSPECTIVES IN SOME SPECIFIC PROBLEMS

3.1. Acid Rain: Genesis, Appraisal and Statistical Considerations. In the entire south and eastern part of Asia, the rain, from time immemorial, has been the harbinger of civilization, the lyric of the harvesters and the melody of the commons.

Let the peacock in thy heart dance with the tune of monsoon and thy emotion in a riot of cheerful colors be in full bloom.
Let the rivers flood their banks in sheer generosity, and golden harvest echo thy thanks in total entity!
But, in an era of immensurable atmospheric pollution, can monsoon bestow you with any hygienic resolution?
With the seeded clouds, what are our expectations?
It's a down to earth invasion by lots of pollutions:
Rain drops soaked with sulphur, acids and smogs, streets flooded with human waste and garbage in tons.
Normal living utterly disrupted by the dreadful diseases; from eyes and lungs to the intestine, the aging processes.
Carrier of enormous environmental pollutions, but, alas, whom to blame for our own creations!

Combustion of fuel and gasoline, thermal and nuclear power plants emissions, absorption of ultra-violet rays (due to the thinning of the Ozone layer), and what not? All of these factors account for the acid rain phenomenon. In the industrialized western countries, the snow in the winter has a similar effect: it leads to the absorption of toxic materials in the soil and contaminates the water resources too!

While the environmental scientists can identify plausible sources of pollutions and explain their impacts in terms of physical effects, the problem is much more complex, and a statistician (with an indepth knowledge in the environmental sciences) can not only identify the demographic and epidemiologic aspects much more clearly, but also suggest suitable models for such a phenomenon to be describable in a simple and estimable quantitative form. We have a long way to go along this path!
3.2. MOTHER EARTH: Vegetation Baldness (Vanishing Tree Lines).

Although we are lamenting over the natural calamities occurring in our highlands and the planes, have we paid due attention to the conservation of plants and ecology in this planet? Deforestation not only affect the living style in the foothills, it has serious impacts on the flattening of the riverbeds due to erosion at the foothills and also contribute to the disasterous flooding in the planes. As a largely oversimplified version of this complex problem, we may turn towards Nepal or other territories in the Himalayan foothills where ecological disasters are quite imminent!

![Fig. 2. Impacts of Deforestation]

Let us look into the major reasons for such deforestations:

(i) Agricultural needs;
(ii) Energy needs: cooking and heating purposes;
(iii) Forest-fires and avalanches.

Also, we may note that the major impacts are

(i) Less vegetation and less absorption of CO₂;
(ii) More erosion due to rain/snow/avalanches.;
(iii) Landslides leading to less fertile lands for agricultural uses.

On the contrary, there are some genuine needs for Afforestation:

(i) Demographic considerations;
(ii) Ecological considerations,
(iii) Exploration of alternative sources of energy.

In each of these sectors, statistical considerations play an important role. Statistical modelling, planning (design) of studies and statistical analysis all dominate the scenario.
3.3. SOLID WASTE PROBLEM: The Garbage Dilemma

It is a major problem for every urban area in every industrialized nation! The dimensions are no less for big cities like Bombay, Calcutta or even Madurai!

Against the traditional procedure of LANDFILLing (or dumping), there are some other alternatives:

(i) RECYCLE a greater part of the Waste,
(ii) LANDFILL the remainder in modern engineered safe sites, and
(iii) Use a RESOURCE Recovery Technology to reduce non-recycled waste to some extent.

The conventional dumping of garbage into Sea or Underdeveloped Landmites is Hazardous in many respects: Toxic material in the garbage may produce far reaching hazardous effects, cause serious pollution problems and contaminations in undesirable manners. A safe and proper Landfill may require a generous amount of land and many safety precautions, and may therefore be hard to locate. Newspapers, card boards, aluminum cans, glass bottles and jars and some other materials in the garbage, if properly sorted at the collection stage, may be recycled upto 50%. Massburn Resource Recovery Plants may reduce the landfill disposal volume of non-recycled solid waste by about 80%. However, the air pollution impacts are needed to be monitored closely. Composting relates to a natural conversion of the organic portion of solid waste to humus by micro-organism activity. However, in this process, removal of hazardous and non-biodegradable materials is required. Refuse Derived Fuel: Garbage into Fuel, can be used as supplemental Fuel with about 50% heating value of coal and having low sulphur content.

It is a complex problem, and plausible recommendations include the following:

i) We must preserve the capacity of existing landfills and begin immediately to acquire and permit new facilities or reserve sites for future use.

ii) Need to develop long-range strategies to significantly reduce reliance on conventional landfills

iii) A comprehensive Recycling plan needs to be developed, preferably on a regional basis

iv) Composting plants are desirable on a small area basis: Regional plants may not be feasible.

v) Regional Refuse Derived Fuel plants are preferable

vi) Public education and statistical evaluations require utmost attention.

Any master plan in this context must take into account (i) epidemiological factors, (ii) environmental issues and (iii) awareness to recycle and other alternatives. It is clear that STATISTICAL MODELLING plays a vital role in this respect, and the standard setup of simpler statistical models may not be that appropriate here. The anatomy of garbage may have a lot of geographical variation and similar considerations dominate the adoptibility of alternative means: Statistical considerations are thus very important.
3.4. OFFSHORE OIL EXPLORATION: Marine Biology & Ecological Impacts.

We start with a crude picture of this 'crude oil' problem.

![Diagram of Offshore Oil Exploration Impacts]

Fig. 3. Impacts of Offshore Oil Exploration

From the technological side, there is a genuine need to incorporate novel statistical tools in the assessment of the dimension of the oil and/or gas reserves. There are definite stochastic factors which should be taken into account in the planning, modelling and analysis schemes. On the other hand, a detailed study of the marine biology and ecological considerations should be obligatory in such an exploration venture. Statistics and (applied) probability theory are being recognized as valuable tools in the study of such biological and ecological undercurrents. There is another important reason why statistics should be incorporated in this venture too. Statistical risk analysis of oil/gas leaks, explosions and other forms of disasters should be deemed as a necessary step before any operation is initiated. There is also a risk of the deformation of coastline and/or sub-soil strata due to energy extraction. In this respect, statistical modelling (in collaboration with researchers in Geosciences) plays a vital role.

3.5. ECOLOGY AND STATISTICS

As an illustration, we consider a simple Eco-cycle relating to marine biology. Although our illustration concerns mostly to inhabitants of seas and oceans, there is a natural loop which extends this cycle to rivers and lakes.
There is an intricate network of Ecology: Pulling any string may affect the whole Ecosystem. Each Species may depend on the ecological balance of others. For example, consider the dilemma of Commercial Whale Hunting in the Pacific/Atlantic Waters. The clearly visible impacts are the following:

(i) Likely to lead to the Extinction of a species
(ii) Disastrous effects on Marine Biology/Marine Geography
(iii) Ecological Imbalance.

The ecological chain of marine animals is indeed very complex. Not only marine biology and marine geography are pertinent in this context, but also there is a genuine need to study the environmental impacts in a critical manner. Statistical modelling is indeed a very important task in this context. Sampling techniques may also be quite different from the conventional ones, and there is an enormous scope for development of more appropriate sampling procedures to suit this purpose. In this respect, weighted or length-biased sampling procedures have been considered by some researchers, and more practical ones are needed to be developed.

To stress the ecological imbalance possibilities, we consider a second example. For Laboratory research in USA/Canada and elsewhere, in the Seventies, a large number of monkeys were imported from the Indian subcontinent. The Government of India had the feeling that this would check the growth of the overpopulated monkeys in India and would also contribute to the much needed Foreign Exchange. Alas, it did not occur to them that the insects/bugs which are the natural prey for monkeys would have an abundance, and that might create some other serious problems! Fortunately, the situation did not go out of control, and with international awareness the Monkey Exodus has been resolved to a satisfactory extent.
3.6. GENOTOXICITY: Environmental Impacts.

To start with we consider the basic Genetic setup in human beings. The explanation of the technical terms follows the presentation of the following:

**Fig. 5. The Cell.**

Chromatin: The portion of the nucleus of a cell that is readily stained by dyes. It forms a fibrillar network of granules of various shapes and sizes, composed of DNA combined with a basic protein and is the carrier of the GENE.

Chromosome: A body in the cell nucleus that is the bearer of the GENE. It has the form of a delicate Chromatin filament during interphase, contracts to form a compact cylinder segmented into two arms by the centrosome during metaphase and anaphase stages of cell division, and is capable of reproducing its physical and chemical structures through successive cell divisions.

DNA (Deoxyribonucleic acid): Type of nucleic acid containing deoxyribose as the Sugar component and found principally in the nuclei (Chromatin/Chromosomes) of living cells, usually loosely bound to Proteins. Considered to be the auto-reproducing component of Chromosomes and of many VIRUSES and the Repository of Hereditary characteristics.

RNA (Ribonucleic acid): Found in all cells, in both nuclei and cytoplasm, and in particulate/nonparticulate form, also in many Viruses.

GENE: A complex, self-reproducing molecule that constitutes the unit of heridity. As a rule, a full complement of genes is present in every (somatic/germ) cell of the body, arranged in linear order, like beads on a string, at definite points (or loci) in chromosomes. There are 23 pairs of chromosomes in humans and since these chromosomes occur in pairs, the genes are also paired. There are some thousands of different kinds of genes in human beings. For any cell, the genes govern the functions of the cell and are primarily responsible for its properties and behavior. There are literally thousands of traits (both qualitative and quantitative) and genes are associated with these.
GENETIC INFORMATION is contained in the relatively simple CHAIN MOLECULE that constitute DNA and RNA. Each of these chains consists of four different kinds of units called NUCLEOTIDES, so each of them can join chemically end to end to form a long string. Each nucleotide consists of a sugar molecule (Ribose in RNA and Deoxyribose in DNA), a phosphate group and a nitrogenous base. In DNA four nitrogenous bases are involved: the purines adenine and guanine, and the pyrimidines thymine and cytosine; RNA the pyrimidine base uracil replaces thymine.

Watson and Crick, and later on, Khorana and his associates have all developed some THEORETICAL MODELS for the DNA Structure which account well for the four essential properties of the Gene: Specificity, Replication, Mutation and Function. The Genes usually produce (through MUTOSIS or Cell Division) EXACT COPIES of themselves for hundreds of Cell Generations. But, sometimes a sudden and permanent gene change takes place - this is termed a MUTATION.

The genes of a cell thus carry a set of messages. They give the cells Chemical Instructions about how to manufacture certain kinds of molecules. These instructions are passed from the DNA to the RNA. RNA transfers the instructions out of the Nucleus to certain structures in the cell called Ribosomes - and the ribosomes determine the way PROTEIN MOLECULES are formed. Thus the DNA instruct the arrangements of Amino Acids. A small change in a DNA message can cause one amino acid to replace another - results in a different kind of protein, and this in turn can bring a change in the structure.

VIRUS : Genetic material in a Virus particle is constituted in two closely paired helical chains, coiled around a common axis - each chain consists of four kinds of nucleotides.

Virology in the light of DNA/RNA! How about CANCER?


The last decade has witnessed a lot of emphasis on the Environmental impacts on the DNA activities. In Love Canal area (in New York State), large amount of chemical dumping has been liked to a significantly higher cancer incidence rate, and similar effects have also been noticed in the Appalachian region (in USA) with respect to the Black Lungs disease. The Epidemiological emphasis has been shifted more towards the Genotoxicity aspects which can be described more conveniently in terms of the mutations of the genes caused primarily by the environmental as well as occupational hazards. Inhalation Toxicology has become a household word in the past few years; no one can deny the influence of the air pollutants, and currently they are being examined more at the genetical mutation level than in the conventional chemical reaction setups. This is by far one of the most challenging areas for the statisticians to contribute to the research methodology, data collection and analysis.
"The presence of TOXIC CHEMICALS in our Environment is one of the grimmest discoveries of the industrial era. Rather than coping with these hazards after they have escaped into our environment, our primary objective must be to prevent them from entering in the first place." [President Carter, 1977]

A MUTAGEN is an agent, e.g., X-rays or a chemical, that is capable of effecting heritable change in genetic material, be it at one specific DNA base-pair or at the Chromosomal level.

(i) Mutagens may exert their effects on Germ Cells, thereby altering the human gene pool and possibly inducing genetic defects in future generations

(ii) Mutagens that attack Somatic Cells may lead to Heart Disease, Aging and Developmental Birth Defects.

(iii) Mutagens play a significant role in the CARCINOGENESIS Process.

MUTATION: A genetically based change, generally expressed as a change in drug resistance, a new metabolic requirement, or an altered gene product.

BASE SUBSTITUTION: A change in the genetic 4-base code involving the replacement of one DNA base for another.

FRAME SHIFT: Insertion or deletion of bases such that there is an offset of the triplet reading frame of DNA, resulting in a string of incorrect Amino Acids as the gene is being translated into protein.

REARRANGEMENT: Deletion, insertion, inversion or translocation of DNA sequences.

CHROMOSOME ABERRATION: Microscopically visible change in chromosome structure.

ANEUPLOIDY: Change in chromosome number.

SISTER CHROMATID EXCHANGE: An easily scored exchange of chromosome strands within replicating chromosomes.

MICRONUCLEI: Small secondary nuclei within cells, indicating breakage of chromosomes.

DNA REPAIR: Evidence that one or more types of repair of DNA damage has taken place.

DNA ADDUCT: A binding of exogenous chemical to DNA

DNA STRUCTURE DAMAGE: Molecular weight change, strand breakage or other gross alterations in the structure of the DNA

[ Tests for these are referred to as GENOTOXICITY TESTS. ]

MARKERS of Genotoxicity are going through an evolutionary change with a matching support from the BIOTECHNOLOGICAL Developments. Some markers of genotoxicity are already being incorporated into epidemiological studies. There is a vast amount of opportunities for more sophisticated scientific studies in this direction. It should come as no SURPRISE that the process of VALIDATION and then Application to Epidemiology will present many interesting and important STATISTICAL CHALLENGES.

Fortunately, public awareness of such GENOTOXICITIES has become quite noticable in the recent past, and given this factor, further developments are bound to occur in near future. Statisticians have a great role to play in this context too. The picture is, however, not one sided. This awareness of genetic mutation has
raised the level of expectations of geneticists and biological scientists at large in regard to controlling Human Traits through forced Mutation. Biotechnology, as is currently referred to, may be of vast utility in eliminating certain hereditary diseases, but is also of grave concern as regards Ethics and Risks. For example, if our goal is to have an enormous amount of super-athelets and we take shelter to biotechnology for possible genetical mutations, are we sure that it won't induce more serious effects with respect to some other traits (viz, natural intelligence)? What are the basic criterion on which to decide whether one should be eligible for a bonafide biotechnological magic touch? Will this lead to an undeclared war against the Nature? In all these respects, statistics can be of great help in the proper assessment of the problems and their useful resolutions.

3.7. ARTERIOSCLEROSIS: A Look into the CARDIOVASCULAR PROBLEM far Beyond the Blood Chemistry.

CHOLESTEROL has become a household word; HYPERTENSION is having a new look. In the Groceries stores and Family Clinics, HIGH FIBRE Diets are the top scorer and Animal Fats are being pushed aside in dinner tables. More and more, homes and working places are converted into 'NO SMOKING' zones. To appreciate these, we start with the human heart:

The Arteries may become thick and heard and lose their supple elastic quality (a process termed the Atherosclerosis and the disorder is termed Arteriosclerosis). This is a normal Aging Process. But, hypertension, overweight, smoking, diabetes.
irregular or insufficient physical exercise, disability and some other factors may all contribute significantly towards the increase of cholesterol level in the blood. Effects: Angina Pectoris / Heart Attacks, Strokes (in the ascending Aorta), Parkinson's Disease etc.

Let us by-pass the heart and enter into an Arterial Channel: The LIPIDS are there.

Normal

[Diagram of Arterial Channel: Normal]

Arteriosclerosis

[Diagram of Arterial Channel: Arteriosclerosis]

With increasing fatty deposits, the interior channel of an artery becomes more constricted, making it more difficult to have the normal level of blood flow through it, and thereby increasing the risk of heart related diseases. Among these Lipids, we have the LDL (low density lipoproteins) which are the bad elements and the HDL (high density lipoproteins) which are supposed to be the good ones. Lowering of the LDL and increasing of the HDL levels is the Ideal way to combating this disorder. Whole milk, Animal Fats (including Crabs and Shrimps), butter, Coconut Oil, Palm Oil, Egg, Cheese, Beef, Pork and other Red Meats, etc are all suspected as liberal contributors towards the increase of LDL level. Moderate amount of physical exercise, low animal protein diet, high fibre meals, less tension (at work and home) and regular habits are supposed to be beneficial. Recent Clinical Trials have all supported that lowering of LDL and increasing of HDL improve the situation considerably. Smoking has been singled out as one of the most significant risk factors in this context. Until about 20 years ago, statistics used to reside at the end of published articles in medical journals with an explicit mention of the standard error of the estimates which might not have much statistical relevance. The medical researchers, however, have started realizing the potentialities of statistical
methodology in every sector of such medical studies. Typically, examination of arterial channel cross-sections requires the sacrifice of life, and hence, such a study was generally used to be planned for subhuman primates; there arose the basic question on the validity of the conclusion derived from Monkeys with probably abnormal doses for human beings with a normal range of risk factors. In fact, the medical ethics constitute the most important factor in the administration of such a medical study. Clinical Trials were conceived as the first bold step towards an Epidemiological resolution under a statistical framework for involving human subjects in actual medical studies. In the context of the Arteriosclerosis problem, there are potentially scores of statistical issues deserving careful studies; to mention just a few important ones, we consider the following:

(i) What Response Variables would you choose in a quantitative assessment of arteriosclerosis?

(ii) If one has to use arterial cross sections, what are the most strategic points in this respect?

(iii) To what extent modern bio-technology can be adapted to improve the recording of observations and increasing the sensitivity of the statistical procedures?

(iv) In view of (iii), what should be an overall Objective of the study?

(v) Is there a better way to define the risk ratio and/or the classification of the state according to the cholesterol level?

(vi) With respect to the classifications of diet affecting cholesterol, what is a sound statistical design to draw statistical conclusions in a valid and efficient manner?

(vii) Since cross sectional study of arteries are made following the termination of life, there is a greater need to develop a statistical plan on a sound epidemiological and environmental basis. To what extent Accelerated Life Testing Techniques are appropriate in such a study?

(viii) To what extent standard statistical methodology can be used in this study? The usual assumptions are not found to be appropriate in the current context.

3.8. CLINICAL TRIALS: Role of Randomization and Statistical Evidence.

The last twenty years have witnessed a steady growth of research literature on Clinical Trials; for the Lipids Study mentioned before, the National Institutes of Health initiated (through the National Heart, Lung and Blood Institute) a 12-year long clinical trial, and the University of North Carolina at Chapel Hill was given the task of data monitoring as well as statistical analysis of the experimental data outcomes. Structurally, the Clinical Trials are a lot different from the conventional laboratory experiment or a cohort study. Clinical Trials are usually planned for a definite end-point in mind and are generally posed to compare a Placebo with one or
more treatment groups. The most important feature is the incorporation of the usual Principle of Randomization which allows statistical methodology to play a basic role in the data monitoring and in drawing valid conclusions from the experimental outcome. For illustration, we consider the Lipids Clinical Trial. More than a million of males (between the ages 35 and 50 years) were screened, and a set of 3952 males having no previous heart attack but a high cholesterol level in their blood were taken as subjects in this trial. In contrast to the placebo group, the treatment group received a drug to lower the cholesterol level to about 160 (in about 6 months). The two groups were then put on a follow-up scheme where the failure (heart-attack) points or the dropout points were to be compared to draw conclusions about the effectiveness of lowering the cholesterol level in reducing the risk of heart attack. Clinical Constraints: Based on cost and time limitations, a maximum period of 12 years was set in advance for the trial. The trial was conducted in seven different hospitals across the Nation, so that geographical variations were to be expected in these multi-clinic studies. On the top of that the clinics might have different levels of sophistications in the data recording procedures, so that the need for standardization was also felt. Medical Ethics: If the treatment is really better than the placebo, then it should be detected as early as possible and all the surviving subjects should be switched to the treatment protocol for better prospects of living. Also, since human lives were to be involved, the loss of life should be a minimum. Further, side-effects or other forms of toxicity may develop in the actual administration of the drug, and hence, close monitoring of the trial would be needed to ensure that the subjects in the treatment group were not to be exposed to extra risks. This calls for a continuous or at least a periodic monitoring of the trial. Epidemiologists' Objectives: To collect all the data you could in 12 years and then to draw conclusions from the same (at the termination of the trial). Should it matter how many times you look into the data in the tenure of the study? Statisticians' Dilemma: This would be essentially a follow-up study with accumulating data. As such, there would be an element of time-sequential in nature, so that standard statistical analysis may not be generally valid in such a case. Moreover, valid statistical analysis of such accumulating data needs to specify the number of looks as well as the nature of looks one would like to make.

It is not surprising that these complexities led to statisticians to develop appropriate tools for proper planning (design) of such clinical trials, operational data monitoring as well as valid statistical analysis. We shall comment more on these later on. However, it may be worth recording the following observations here:

(i) How was the actual screening done? What were the basic considerations and rationality in this respect?
(ii) What were the Primary Objectives of the study and secondary ones, if any?  
(iii) Recruitment Protocol: How was Randomization implemented in the process?  
(iv) Compliance & Dropouts: How much they depend on treatment vs. Placebo?  
(v) Proposed Duration (12 years): What was the rationality of this choice?  
(vi) Primary variates & Concomitant variates: Too many covariates (blood chemistry, age, diet, smoking/not, p.e., etc.). How many of these were really relevant to the study?  
(vii) Determination of the sample size needed for the trial to be conclusive.  
(viii) Statistical Hypotheses and Power considerations?  
(ix) Interim Analysis: Medical Ethics, Clinical Constraints and Statistical Basis.  
(x) Some Important Statistical Factors:  
 a) Time-dependent Covariates (viz., the blood chemistry levels may vary considerably over time).  
 b) Staggered Entry Plan: It took nearly 2 years to recruit the 3952 subjects, so that they would have different exposure times during the tenure of the study!  
 c) Random Censoring: How realistic it would be in the given context?  
 d) Proportional Hazards Model: How much reliance one would have in this model especially when there are a number of covariates?  
 e) Competing Risks Models: Quite reasonable to be adopted in the given context?  
 f) Interim Analysis and Stopping Time: The intricate relationship between the type of interim analysis one may like to adopt and the related decision to curtail the study is by far the most important statistical issue.  

In a typical Clinical Trial for the study of the relative performance of a new drug (treatment), the other significant factor is the documentation of it as a Phase I, Phase II or Phase III Trial. The medical and Clinical implications may be quite different for these three phases, and hence, the Statistical Issues are to be addressed accordingly.

3.9. Biological Assays: Relevance of Statistical Reasonings!

Typically, a bioassay involves a stimulus applied to a subject and some measurable characteristic of the subject (whose magnitude may depend on the stimulus) constitutes the response variable. From time immemorial, in medical practice, a new drug or toxic preparation has been tried first on some sub-human primates and depending on the nature of response, safe dose levels have been prescribed for human use. The past 100 years have witnessed a phenomenal growth of literature on serious use of biological assays in medical and clinical studies. Within this broad
framework, if the dose required to produce a given response is measurable (and is typically a random variable), we have a direct assay, while, if for various dose levels, the response can be observed (and is a random variable), we have an indirect assay. The assay is termed a quantal one if the response is all or nothing. In either case, we may characterize the presence of random elements either in the dose levels to be administered or in the response levels to be obtained through the assay. This brings in the relevance of statistics in this fruitful area of medical and clinical research.

Empirical Evidence in Clinical Studies has been an issue of great concern to clinicians as well as to various regulatory agencies. While it is undeniable that there is considerable variation in the tolerance of the doses from subject to subject, it is necessary to determine plausible decision rules to administer new drugs in the market, and in this respect, biological assays provide a satisfactory platform for performing valid and efficient statistical analysis. In fact, at the present time, biological assays are generally prescribed before a clinical trial is planned (Phase I). Fortunately, during the past fifty years, significant development has taken place in the arena of statistical methods in biological assays, and these, in turn, provide a valuable service to researchers in clinical and health sciences.

When the Etiology of a disease or disorder is not precisely known, researchers are naturally led by intuitions to a greater extent and past experience with other related ones. For example, with the AIDS problem or even the Cancer problem to a greater extent, inspite of some successes in some specific corners, there is an open query: To what extent one would rely on the Epidemiological aspects and what would be a proper mechanism to channelize molecular biological approaches? Since human subjects are involved in the primary end-points, one would naturally be inclined to start with suitable biological assays with cats/dogs/monkeys and/or other sub-human primates. Such biological assays are immensely useful in collecting valuable information on which further medical studies can be made. But, there are darker sides too. For example, Saccharin is suspected to have Carcinogenicity, but the normal human consumption dose is usually small, and if any effect is there it has to be observed only over a long period of use. To have some idea about this carcinogenicity, there were some attempts to have biological assays with monkeys (instead of human beings) as subjects, but to cope with time and cost limitations, 800 times the normal human dose was prescribed: The end-product was a higher rate of bladder-cancer, and it was quite tempting to quote this evidence to ban Saccharin from the market! But this scenario is typical of what is known as an Accelerated Life Testing Model. If one assumes suitable parametric model (or even a proportional hazard model) for the carcinogenicity (as a function of the dose), one would be naturally inclined to use interpolation on such monkey-data to draw a parallel on human subject.
To make such bioassays looking more impressive, one may as well go into the molecular biological level and bring in the mutagenesis to stress on the cancer etiology! However, the question remains open: Is the carcinogenicity mechanism for saccharin on human beings similar to that on the monkeys? Further, whether the adoption of such a high dose level is valid from the statistical extra-polation point of view? Nevertheless, the positive aspects of such bioassays provide enough incentives to carry out similar experimentations with more moderate dose levels and under more general experimental setups which would resemble human conditions to a greater extent. Actually, in many developed countries, the drug regulatory agencies place a lot of weights on such bioassays in determining a safe dose level on which a clinical trial can be administered on human beings. As such, statisticians are now charged with the task of developing statistical planning and analysis procedures which would make such transitions smooth and effective. We have a long way to go along this line too.

Biological assays and clinical trials are also intricately related to another health related activity: Drug Information and Commercialization, and statistics play a major role in this sector too. Pharmaceutical research groups and the Drug regulatory agencies are on the opposite sides of this tug of war. It is undeniable that the pharmaceutical research groups have contributed most significantly to human health and welfare by developing more and more fundamental drugs which have made life more bearable. It is also true that pharmaceutical research, because of its vast user market, has a natural tendency for commercialization through patents and sale by various pharmaceutical industries (many of whom are truly multinational). Thus, in most of the countries, there are the drug regulatory agencies which monitor such drugs before they are released into the market. As is the case with USA (where FDA is the Agency), there is a natural tendency to attach more weight to clinical evidence, and in such a screening procedure, often, statistical evidence is not given the due importance. However, the past two decades have seen remarkable changes in this scenario too. There are, by now, established groups of (bio-)statisticians who regularly work closely with the Pharmaceutical research groups and others with the Agencies, and through their mutual understanding and respect for each others point of views, the situation is bound to turn for the better. In this respect, statistical planning of biological assays/clinical trials has been recognized as one of the most essential features, and no wonder more and more novel statistical tools are being developed for effective analysis of such trials which can be more convincing to both the sides. In this respect, it is, however, essential that the involved statisticians look into the depth of the medical issues and the clinicians explain their objectives very clearly so that statistical formulations are feasible.
Epidemiological informations in such studies constitute a vital part of planning, but without adequate statistical as well as clinical evidence, valid conclusions or proper decisions can not be made. The important point is that there is an increasing degree of awareness of the need for statisticians and epidemiologists working together, and with this, the scenario is bound to be bright in near future.

3.10. A Tale of the ANTIBIOTICS.

I am intrigued by a recent TIME magazine (August 31, 1992) Feature article "Attack of the Superbugs": This features the losing battle of the magnificent ANTIBIOTICS: The invisible legions of drug-resistant Microbs are fighting back! The Penecillin group of drugs, introduced nearly 50 years ago, offered an arsenal of magic bullets to knock out the germs that cause everything from Typhoid, Tuberculosis, Pneumonia to Gonorrhea. These Antibiotics have been widely used by the medical professionals, and quite often disregarding the excessive doses (running into millions) with which they are to be administered on successive occasions. Almost five decades of continual use has brought some striking news too. First, these may have serious side effects, and hence, indiscriminate use of these drugs should not be encouraged. Secondly, these Antibiotics are apparently becoming less and potent for the treatment of the specific diseases for which they were highly effective. In fact, they are now on their reverse gears in their combat with the drug-resistant microbs. It is undeniable that repeated use of such Antibiotics requires an increasing sequence of doses, setting limitations on the number of times they can be administered on the same subject.

Genetical mutation of these organisms provides good explanation of the fighting back of these microbs. When a microb replicates itself over many generations, mutations in the DNA that forms the organism's genetic chart can sometimes make it safe from an Antibiotic attack. Moreover, once a bacterium has a protective combination of genes, they are duplicated every time the bacterium reproduces itself. Further, through a process called Conjugation, the microb can pass its genetic shield to a different strain. In addition to exchanging DNA, conjugating bacteria can swap smaller snippets of DNA called Plasmids: Like Virus, these Plasmids make exceedingly effective shuttles for carrying drug-resistant traits from one bacterium to another. In this respect, the theory of Epidemics (as is usually adopted for the study of spread of some common diseases) may as well be incorporated to conceive suitable growth-curve models for such drug-resistant traits in a deterministic manner (yet allowing stochastic elements to a desired degree). Can Statistics be far away at this phase of research endeavor? We are in the era of bio-technological evolutions, and what needs to be done is the following:
(i) Better understanding of the molecular biology of the organisms;
(ii) Reviving the art of vaccine development;
(iii) Effective scrutinizing of the use of Antibiotics;
(iv) Curtailment of overuse of such Antibiotics,
(v) Developing newer and more clever drugs to combat with these drug-resistant microbes in a more effective way.

To accomplish these tasks, the clinical research setups may need to conduct a lot of bioassays at the sub-human primate level first and then on to the clinical trials at the human level. In either case, since living matters are to be identified as subjects, one has to take into account the usual sources of variations (with respect to the response variables), and has to plan the study in a sound statistical way so that valid and efficient statistical conclusions can be made.

3.11. MEDICAL DIAGNOSTICS and Statistics.

Not too long ago, a family doctor would look into the general appearance of a patient and felt for the pulse-rate to conclude on the nature and extent of the disease. Now a days, there is a necessity of a battery of preliminary tests on various aspects before a medical personnel would even start looking at the prevailing disease; there may be both medical and legal issues involved in such a precautionary measure. On the positive side, such preliminary tests and observations are generally helpful for the medical practitioner to diagnose more precisely the problem. Sometimes, it may also reveal some other problems for which no alarms were recorded by the patient. On the negative side, this screening procedure is generally costly, and the patients (directly or indirectly) provide that. Moreover, there are generally a large number of variables on which information is sought through such preliminary medical tests, and there are definite amount of stochastic elements associated with each of these measurements. The body weight, pulse rate, cholesterol level and many other blood chemistry levels depend on the time of the day and nature of foods taken prior to such tests. Fortunately, through statistical awareness, medical practitioners now a days generally agree to have more uniformity in this preliminary records by imposing other side conditions (such as no breakfast before these samples are to be taken). Even so there are literally thousands of other unseen causes which may have good influence in this matter. The medical practitioner then compares the Chart with some of the Standard Ones for various disease groups, and then decides on which one matches the current one in his (her) folder, and then draws the conclusion accordingly. In a sense, this is largely a multivariate classification problem when some of the variates may be categorical (qualitative or ordinal). May statistics be permitted to enter into such decision making arena?
The fundamental role of medical diagnostics in modern clinical practices should not be deemphasized. For example, for a patient arriving at a clinic with a vague complain of 'heart-burn' symptom may actually have a much more serious problem. It can simply indicate some disorder in the digestive system, some related problem with the gall-bladder or some other organs in the abdominal area, or more likely, it may be due to Angina Pectoris, an arterial constriction/blockage or even some problems with the heart ventricles. As such, a medical practitioner would naturally like to have a thorough medical diagnostics before the actual treatment can be initiated. Generally, if the initial symptoms are not that discouraging, as a temporary relief some standard medicine is given prior to a complete checkup of the relevant problems. However, there are also situations where there may not be enough time for postponing the medical action until a full report is available, and an immediate medical treatment has to be resumed. It is not surprising that in such a case, there is a high probability of a surgery (contrary to the claim that in USA more than 50% of such surgeries are really not that essential). The medical community prefers to attach more importance to even somewhat imminent cases than overlooking some which may result in a more serious problem in near future. There are many statistical considerations relating to medical diagnostics and some of these are yet to receive proper attention from the medical practitioners:

(i) Based on the available information (which is generally very partial in nature), the first task is to search for more relevant information through appropriate medical diagnostics, so that an appropriate decision regarding medication/surgery etc. can be made with minimum risk. However, the interpretation of information may be too vague to permit any such precise decision making. Secondly, unlike the case in mathematical statistics, the loss function is not so simple. The interpretation of risk from socio-economic, medical and statistical points of view may not be that concordant. Nevertheless, statistical interpretations may be of immense help in such a medical diagnostic too. The advent of modern computers has impressed people in every walk of life, and medical profession is no exception. But, that should not be an excuse for running standard statistical packages for such decision makings (without checking the actual relevance of such packages in the situations under consideration). Simple statistical senses can often be very rewarding in this setup too.

(ii) Most of these medical diagnostic tools are not totally safe for indiscriminate uses. For example, X-ray plates are now a days used for almost everything: Chest, abdomen to other intricate parts of our internal system. For pregnant women and young children, the radiation problem may not be totally ignorable. What factors should be taken into account, so that a more judicious use of such fancy tools
can be made in practice? As in the case with the Antibiotics (discussed earlier), excessive use can induce some other serious side-effects or some other unwanted medical problems. To stress this point further, we consider the usual practice of having mammograms for screening of potential cancer in female breasts. Although it is highly desirable to have this checkup regularly once a person is over 35 years of age, it is undeniable that repeated use of X-rays in this process may often be hazardous. Hence, there should be a general guideline as to the minimum risk mammogram checkup schemes. Again, the task can be made much more manageable if a more comprehensive picture can be drawn based on a sound statistical assessment of the relative risk of breast cancer at various age-groups and under various medical/health and socio-economic setups, so that sole emphasis need not be given on the mammograms at successive checkup periods. The general observations pertaining to this specific problem remain relevant to a general class of problems.

(iii) Medical diagnostics and modern medical technology are indeed very intimately associated. In modern medical technology modern electronics and computer science both have fundamental contributions. We can virtually monitor any body related activity through most sophisticated instruments, and the days may not be too far away when we may want to go through 'Gene Repair' shops at our own discretion for the betterment of our health and life. Mother Nature with all her generosity to mankind may not be that tolerant to such 'robotic' activities on our part, and natural or man-made calamities may surface on Earth in retaliation to our imprudent aggression! Perhaps, we should look on both sides of the fence, and indeed, sound statistical evaluation of such blue-prints may render a most valuable service to mankind. Medical as well as humanitarian ethics should govern such considerations. I hope that the majority of readers would agree with me that in academic achievement brilliance is not sufficient and dexterity often dominate the scenario; likewise in health, apart from a good medical treatment, a healthy environment is also essential. The creation of such a livable environment is our responsibility. The entire field of Public Health and Environmental Sciences is motivated by this outlook, and needless to say that Statistics is a vital component in this context too.

(iv) The form and standard of Medical diagnostics vary considerably from one country to another, even within a country from one region to another. This is not surprising, as there is an intricate relationship between the 'health care' system prevailing in a place and its adopted medical diagnostics. A socialistic system may have greater governmental control on the hospitals and medical care systems and may thereby have different protocols than in a capitalistic country where the health system enjoys the private enterprises, and where the health insurance premiums are now sky-rocketing! Although 'cost and benefit' analysis should govern the arbitration of health insurance premiums, there are many other factors, such as medical diagnostics
norms which can swing the cost more in favor of the enterprisers. Although, it is not truly scientific, statistics with its actuarial heritage may also serve as a true moderator in the formulation of such a cost benefit approach.

(v) The practice of medical diagnostics in a country or region is also related to its 'health services' facilities. We intend to discuss this in more details, and we shall emphasize on the scope of statistics in that context too.


In Clinical sciences, Image Processing is becoming an important tool in various respects. Let me mention two of the most useful areas; the general description will remain pertinent to a more general setup.

(i) **Tomography.** Modern Electronics have endowed the medical professionals the ability to scan practically everything from the tip of the hair to the nerves in the toes. Let us consider a simple model: Suppose that a patient is suspected to have a tumor in a very sensitive part of the brain. The size and the shape of this tumor (there may be even more than one) are not known. Some idea of this tumor can of course be obtained by electronic devices. For example, X-ray plates can reveal the projections of such a three dimensional object on two-dimensional plates. If the object is very regularly shaped and if the axes of the projections are in agreement with some distinct axes of the object, then a good picture can be drawn from its projection. On the other hand, depending on the orientation, the projected picture may be quite noninformative. For example, if the tumor has a cylindrical structure with a very small cross section and comparatively large length, a projection along the direction of the length will reduce it to a small circle while a projection along the perpendicular direction will yield a larger line-segment or a thin parallelogram. If the tumor has more or less the shape of an ellipsoid, a projection into a plane may reveal all sorts of pictures depending on the orientation. The situation may become quite cumbersome if the object (tumor) is of quite irregular shape. On the other hand, on medical grounds (radiation effects!) it may not be advisable to have many X-ray plates covering all possible directions. Generally, two plates in perpendicular directions are prescribed. As such, the basic problem is to construct an image of the object (tumor) based on a few two-dimensional projections of it. Tomography refers to this branch of diagnostic techniques. Obviously, it has a great impact on medical diagnostics and effective treatments. Stones in the Gall-bladder, tumor in the Lungs, obstructions in the intestinal tract (due to tissue-sands/stones), unwanted tissue growth on the stomach lining or between the stomach and pancreas, and more complicated ones, such as the arterial blockages due to Lipids or possible blood clotting are all known cases where such tomography can render a very useful service. **Projection Pursuit** is becoming a familiar word in statistics, and I invite...
the statisticians to have a look into this tomographical heritage of this fruitful area of statistical research. Let us have an alternative look into this tomography problem when we have a (fairly) large number of two-dimensional projections of a three-dimensional object (resulting from the basic consideration that taking a large number of X-ray plates at possibly different orientations covering the entire domain is economically and ethically feasible). The images obtained from the entire set of plates relate to a common object and hence they are stochastically dependent. On the other hand, the projections along different orientations generally have different forms, so that even in a stochastic setup, it may not be very reasonable to assume that they have the same marginal distribution. Moreover, two plates with a small angular difference would be expected to be more (stochastically) dependent (or associated) than in the case where they are more separated by their angular deviations. As such, it would not be generally reasonable to assume that they are all symmetrically dependent or exchangeable random variables. Assuming that the object under examination has a smooth regular form subject to minor perturbations (in all possible directions), the projections can be described by a systematic component plus an error on which various assumptions can be made to justify the adoption of the classical directional data models or more general spatial models. In passing, it may be emphasized that in such a statistical formulation, a very fundamental role is played by the regularity assumptions made concerning both the deterministic and stochastic components of the object under examination. As such, the more information it can be extracted from the actual medical context (like a plausible shape of a tumor and its probable angular position etc.), the better will be the prospects for a valid and efficient statistical analysis. Fortunately, modern medical technology has endowed the clinical scientists with an wealth of information on such plausibilities, and granted that the statisticians are willing to share this feeling with the clinicians, a satisfactory resolution of such problems can be made in a majority of cases. However, statistical packages should not be used indiscriminately without first checking the suitability of the regularity conditions underlying the specific packages to be used in a particular context.

(ii) Image Processing in Molecular Genetics. Molecular genetics, as briefly introduced in Section 3.6, has opened up a broad avenue of opportunities for scientific as well as objective investigations in the area of clinical sciences as well as in many other fields (Jurisprudence has also been annexed to this domain)! The vast and accurate wealth of information contained in the DNA strands enables us to conduct many studies which were previously not possible. On the other hand, the sheer high dimensionality of such information codes has posed some genuine statistically challenging problems. Information processing for high dimensional
data in the context of Image Analysis is indeed an active area of research in statistics and computer science, and is often classified under Artificial Intelligence. Image analysis in a statistical sense covers a broad spectrum of studies. An excellent source of such studies is Serra, J. (1982): Image Analysis and Mathematical Morphology, Academic Press, New York. Popularly known as Serra's Calculus, this set theoretic calculus is useful for statistical and probabilistic analysis of the highly structured data of images arising in fields using remote sensing and microscopy. In this setup, it covers areas beyond clinical sciences, and includes methods for the analysis of complex patterns, such as microscopic sections of tissue or rock, remotely sensed images as well as patterns of communities of vegetation. The last topic mentioned is also a member of the general class labeled as Pattern Recognition Problems, which have also witnessed significant statistical developments during the past fifteen years. In the fields of application of Image Processing, mentioned above, generally, considerable information about the probable contents of images is available. However, a full use of this knowledge is often not possible due to the vastness and/or the complexity of such information. Hence, often structural methods are used to extract images under some preset structural models. Alternatively, in some other models, individual pixels are viewed as primitives of an image, and they are then incorporated in a synthetic assessment of the whole image. Spatial sub-patterns are also sometimes used in this context. For some of these details, we may refer to Chapters 16, 17 and 18 of Handbook of Statistics, Volume 2: Classification, Pattern Recognition and Reduction of Dimensionality (1982), edited by P.R. Krishnaiah and L.N. Kanal, North Holland Pub., Amsterdam. Although the past ten years have seen some further developments, there remains the scope for enormous statistical developments in this area. The sheer high dimensionality of the data sets arising in such a study and the bulk of replicates (although not necessarily identically distributed copies) have also led such problems to be related in a broad sense to Meta Analysis in statistics. We shall refer to Meta Analysis in the concluding section, and hence, we skip the details at this stage.

3.13. NEUROPHYSIOLOGY/NEUROBIOLOGY and Statistics.

In human being, the Central Nervous System (CNS) includes the brain and the spinal cord. It has been recognized as a highly complex communication system which receives, codes and processes a staggering amount of information. A basic unit of the nervous system is the neuron (nerve cell) and it receives and transmits information. It is accepted widely that higher brain functions are performed by some large neural networks. There is some degree of uncertainty inherent in the behavior of such a system, and this stochastic variability is reflected in the information relayed between the neurons in the form of Spikes or action potentials. Because of
the established evidence that the action potentials are nearly identical in shape. It is believed that they contain a limited amount of information in their shape of wave form. Therefore, the general feeling is that the temporal patterns of spike trains are the information carriers in many areas in the CNS. In this context, we may refer to the classical text by E.R. Kandel and J.H. Schwartz: Principles of Neural Science, North Holland, Amsterdam, 1981. As a result of intensive research in neurophysiology, conducted during the past twenty years, many significant evidences have surfaced. It is well acceptable now that the CNS consists of a number of regions (areas) each one of which has a special portfolio of activities and also conveys inter-regional information in a convenient way. Because of the vastness of the number of neuron in each of these areas as well as in the entire complex, it is of natural interest to identify the nature of association or dependence of the spike trains for neurons belonging to the same area or to different ones. This picture may relate to intra-regional and inter-regional measures of association. Statistical methodology in quantitative neurophysiological studies has already been identified as a vital area of research.

Spike train data relate to times of occurrence of the spikes associated with which is a counting process. However, in view of the fact that there are many neurons we are led to conceive of a multidimensional counting process; the sheer high dimension of such counting processes generally poses challenging problems in the formulation of suitable models for which statistical analysis can be carried out in a convenient way. For a single neuron, a very convenient way of modelling such a counting process is to adopt the so called doubly stochastic Poisson processes adapted to a suitable history process leading to a (possibly stochastic) intensity process; it is such an intensity process which needs to be estimated from an experimental outcome, and it contains the interpretable information which accounts for experimental outcomes—statistical pattern and neuro-physiological interpretations too. Disguised mathematicians have not hesitated to take this opportunity to introduce suitable stochastic differential equations in the setup of an infinite-dimensional model in providing explanations for neural activities. Although on the surface, this attempt looks superb, there are a number of important statistical issues which may render them as unusable in a practical context. The Ornstein-Uhlenbeck process is a convenient continuous approximation to the activity of a single neuron, while a more natural and realistic model is the Tuckwell-Cope discontinuous model. In either setup, there are some stringent assumptions needed to justify the model, and in actual practice, as of now, there is no overwhelming evidence that such a priori assumptions are that realistic. But, as has been mentioned before, the actual setup is much more complex due to the presence of a vast number of neurons in the CNS.
The first and foremost problem is basically related to: How many neurons can be simultaneously put to recording of spike trains? Can the conventional method of recording spike trains for a single neuron be extended to an indefinite number of such neurons in the CNS? What is an identifiable basis of marking such a large number of neurons in the CNS? Even if one stratifies these into a number of small areas, the question remains open: On what basis such areas are to be demarcated? What are the consequences of misclassification on the subsequent analysis to be made? The advent of modern electronics has probably led to some satisfactory resolutions to some of these problems. Nevertheless, there remains much to be accomplished in order that valid and efficient statistical modelling and analysis can be incorporated into such quantitative neurophysiological studies. From the statistical point of view too, there are some challenging problems. Consider the sector of multivariate counting processes. It may be quite tempting to consider a bunch of independent (possibly nonhomogeneous) Poisson processes and express a multivariate process by a matrix premultiplier of this vector process (this has actually been considered by some researchers). However, one has to keep in mind that each of these counting processes is marginally a jump-process (and is therefore nondecreasing in nature). This would in turn induce the restriction that the elements in each row of the premultiplier matrix is nonnegative. This, by itself, may not appear to be that restrictive. But, if one decides to consider only such matrices with nonnegative elements, then the cross-correlations (or other measures of association) of the spike trains for different neurons may only be of nonnegative forms. This will exclude the case of negative association and thereby may not reflect on the 'habituation' patterns. A related question: How would you justify such a premultiplier matrix from the basic neurophysiological considerations? Further, given that only finitely many electrodes can be inserted into the CNS for simultaneous recording of spike trains, the more pertinent question is: How many of these are to be done, and is there an optimal design in which it should be done? The more we know about the Brain Structure from neurophysiological point of view, the more sophisticated statistical methodology can be adopted to pursue such studies. Actually, if such neurophysiological factors can justify the adoption of the basic regularity conditions for a continuous or discontinuous diffusion process approximation then much can be accomplished by incorporating modern probability theory for stochastic processes in this potential area of statistical research.


This has been one of the sectors where statistics has been quoted (or misquoted) from time immemorial. For example, during the last century and the first two decades of the current one, Plague, Cholera and other forms of epidemics have devastating
effects on our 'growing human family'. Other diseases like the Dengue, Cholera, Malaria, Pox, Typhoid, Tuberculosis, gonorrhea and many others have been a threat from time to time (if not, even now), and at the present era, AIDS dominate the scenario. Demographic or Epidemiological studies have indeed cast light on their brutality as well as the extent of damage caused by their devastating reins. Nevertheless, it did not account for the etiology of these diseases nor the possible remedies which might have been possible if enough care in public health could have been instituted. All these diseases have one main element in common: They are infected by Virus and are communicable. As has been mentioned in Section 3.10, for some of these diseases suitable Antibiotics did a superb job, but the final outcome may signal a victory for the reentry of these diseases with the decreasing protecting power of these antibiotics. Moreover, because of the 'communicability' factor of these virus, there may be a natural query: What do we know about the spreading power (known as the epidemic effect in a broad sense) of such a disease? What statistical model can be formulated which takes into account the physical factors as well as the stochastic factors in the etiology and epidemics of such a communicable disease? It is not surprising to see that for the modern demon: AIDS (or HIV), statistician from all corners of this Planet are desperately searching for suitable epidemic models to describe the growth of this dreadful disease and to have some idea about its impact in near future. In this venture, one needs to take help from several disciplines: Immunology, Bacteriology; Cell-biology; Epidemiology and Virology are the most important ones in this shopping list. However, statisticians indeed play a basic role in any scientific study of such a communicable disease. First of all, any such study needs a careful planning so that the contended conclusions can be drawn in a valid and efficient manner; this belongs to the statistician's court of Planning (design) of experiments. Secondly, conventional statistical techniques may not be that usable in such a context. Novel tools are therefore needed. We shall stretch on some of these in the next section. Finally, without the effective collaboration with the experimental scientists, the etiology of a disease may not be that clear to a statistician desirous of statistical planning and/or analysis of an acquired data set, so that the endproduct may not be that relevant to the particular context.

In many countries there are Centers for investigations of the etiology and epidemics of communicable (or infectious diseases (and USA is no exception). More and more scientific studies conducted by such Centers are involving qualified Biostatisticians as well as Epidemiologists. It is our hope that their mutual agreement in the broad etiology and collaborative formulation of suitable stochastic models would greatly enhance the scope for such studies and these, in turn, will cast much more scientific information on the nature, growth and possible decay of such dreadful diseases.
3.15. PSYCHOMETRICS to PSYCHIATRY: Statistical Perspectives

Whereas Psychometrics deal with the Psychological theory and techniques of mental measurements, Psychiatry is designated as a branch of medicine that deals with mental, emotional or behavioral disorders. It is in this broad sense, they may be integrated into a broad discipline of theory, methodology and applications of mental measurements in sociology as well as (social and clinical) medicine. It has come to be recognized that such mental, emotional or behavioral disorders may be congenital, environmental or temporal. Although there is a saying that the tip of the Psychological theory Iceberg can atmost explain 1/9th of the entire phenomenon, it is undeniable that there has been remarkable progress in research in this broad domain, and also much in this domain has been accomplished through empirical studies. In Oriental Medicine, Psychiatry is an inseparable part of general/ family medicine (there is a saying that the very sight of the doctor can eliminate at least 50% of the illness). Such a compliment (!) for Psychiatry has been due to the fact that family physicians have traditionally relaised the importance of counseling with the patients in the proper diagnosis of a medical problem. Unfortunately, the advent of modern medical technology has been credited with the consistent shift from this traditional medical practice. Now more and more sophisticated machines are printing out numerous medical measurements before a doctor even sees a patient, and that in turn makes it more probable to bypass the counseling part of such medical advises. This is affecting the patient-physician relationship and also leads to decreasing confidence of a patient on the physician. On the other hand, life is becoming increasingly complex, and the daily stress and strain, one has to go through, make it more vulnerable for such mental, emotional and behavioral disorders. The crime records for common rapists reveal an overwhelming feature: a majority of these rapists have either some serious frustration or mental disorders. Certain types of jobs can tigger more stress/strain and thereby increase the risk of such disorders. There are also scientific evidences that many of these problems are transmitted from one gene/ration to the next one. Finally, most of the theories in Psychometrics have sound statistical interpretations. In Psychiatric treatments too, the physicians try to follow some general guidelines wherein statistical considerations and medical beliefs have migndled together in a cohesive manner. This is not at all unexpected: In most of the areas in medicine, empirical evidence led to the formulation of suitable hypotheses which were clinically tested and then put to practical testing for proper improvement of the medical treatment relating to such problems. Identifying properly the sources of variability along with the important factors affecting the experimental outcomes can still be regarded as the most important task, and statistical considerations are
equally important in this task. In fact, many of the important concepts in statistics have originated from this field. To mention just the most popular ones, we may refer to Canonical Correlations, Principal Component Models, Factor Analysis, Parallel Tests and Reliability of (mental) Tests. Mathematical modelling of such mental, emotional and behavioral disorders is being recognized as a very important area of statistical research, and probability theory along with stochastic processes will continue to serve as indispensable in this respect.

3.16. Statistics in Environment and Health Agencies.

With the increasing complexities of our environmental awareness and diversity of health problems, various agencies have been formed to oversee the tasks and promote better understanding of general public in such matters of national as well as ecological importance. The National Institutes of Health in USA may be regarded as the main Agency and various daughter agencies have emerged from it with the growing needs to attend to more specialized branches. National Heart, Lung and Blood Institute, National Cancer Institute, National Institute of AIDS, National Institute of Immunology, etc., are all notable examples of such specializations. On the environmental side, the main Agency is the Environmental Protection Agency, and sooner came the need for better coordination with health affairs, and out of this the National Institute of Environmental Health Sciences was formed. Food and Drug Administration plays a vital role in the regulation of new drugs as well as vital food products in this country. The Center for Diseases Control has its own goals. It is not surprising to see that there is also a National Center for Health Statistics. In Canada, Statistics Canada has its jurisdiction on some of these portfolios, albeit under different agencies (like Canadian Wildlife Services, Environment Canada etc).

In most of the Health Institutes, as expected, the primary setups were mostly medical/clinical, while in the Environmental ones, the experts in Environmental Sciences and Engineering were the forerunners. However, it did not take much time to realize that in each of these sectors there is a vital need for statistical planning, data collection and statistical analysis. This demand has been growing at such a pace that more and more interaction of the academic community with these agencies is taking place, and more opportunities are being created for statisticians to get involve into such applied sciences. The last but not least important citation is the creation of the National Institute of Statistical Sciences (in the Research Triangle Area in North Carolina) which is to foster statistical interactions with a broad range of scientific and technological fields. In this contemplated list, health and environment both have been singled out as the most important ones.
4. STATISTICS: TASKS AND PROSPECTS

The list of topics covered in Section 3 is admittedly a partial one. I could have continued in this fashion in citing other fruitful areas where statistical perspectives have real significance. However, I would rather check that temptation, and try to summarize my observations in the form of the following broad comments.

I. As in agricultural sciences, in most of the clinical, environmental and health science studies, generally an experimentation is contemplated; it involves a comparative study of two or more treatment groups, one of which is usually termed the control or placebo. In experiments involving physical or chemical measurements (and also in agricultural ones), the control group can be identified in an unambiguous manner, and this renders a great relief for effective comparison of the treatment group(s) with the control group. In many of the problems referred to in Section 3, although such a control group can be formed on some convention, there may not be enough control to retain its characteristics. For example, in an inhalation toxicology study, the objective may be to examine the effect of automobile exhaust on the atmospheric air quality. It is mingled with many other factors and depends a lot on the particular urban/rural area selected for such a study. The volume of traffic, type of fuel consumptions, local climatic conditions all are contributors to this complex problem. What is a control group in this setup? One may setup a particular time period of the day (say 2 to 3 PM) when the flow of traffic may be substantially less than in the rush hours, but the quality of air sampled at that time may still contain the effect of other time periods. One may set Sunday as a control and take Monday or Friday as the treatment, but the problem may still remain in a latent form. Or, in other words, the basic issue is how to deemphasize the role of control vs. treatment setups and formulate suitable regression setups with factors and their intricate dependence patterns built into the concomitant variate sets?

II. In the setup of the set of problems posed in Section 3, often, the sampling procedure is quite different from the conventional ones. For example, in the same inhalation toxicology study, how would you measure the air quality? If you decide to have a sample of the air at the ground level, it depends a lot whether you are locating the sample unit close to a heavy traffic area, residential area or somewhere else. If you decide to have samples of air at some elevation (say, 40ft. above the ground level), some of the chemical pollutants may not be that represented in such samples. Or in other words, drawing of samples has to be made in full concordance with the basic objectives of the study, and this may often make it quite complex in nature. Identification of deterministic and stochastic factors in this context constitutes a vital task. This, in turn, may generally lead to a setup where
the conventional assumption of the sample observations being independent and identically distributed random variables (or even exchangeable random variables) may not stand valid. That is, simple random sampling (SRS) with or without replacement may not be the right prescription for such problems. Consequently, the conventional statistical methodology, developed mostly for SRS schemes, may not work out properly in such complex sampling designs. In certain ecological and environmental studies, length-biased or weighted sampling schemes have been introduced and the necessary statistical methodology has been developed. This is certainly a big step in this direction, and there is a need for development of novel statistical methodology to handle problems arising in this broad spectrum in applied statistics.

III. In the conventional setup, an experiment is planned in such a way that some well formulated statistical model can be adapted to the experimental outcome. Such a model typically relates to a probability law (distribution) involving a finite set of unknown parameters; the form of this probability law may be assumed to be given or it may belong to a general class. Thus, the basic problem is to draw statistical conclusions on these parameters (or some times the law itself), subject to desired level of precision, from the experimental outcome data set in an efficient manner. In the current context, a suitable formulation of a sound statistical model is the first and foremost task. A model needs to be very relevant to the actual study setup, must take into account the intricacies ascribable to various assignable factors (underlying the experimental setup), and at the sametime, must pay sufficient attention to the latent mechanism of chance variations associated with the study scheme. In clinical, health and environmental studies, Mother Nature may play, in general, a very significant role, and human being may not have much control over her jurisdiction (although it can be made worse by careless actions). Because of these reasons, in such studies, the actual list of plausible assignable factors may not be that tenuous, nor the actual mechanism of chance variations be ascribable in very simple terms. An inadequate model may lead to improper formulation and inefficient (if not inconsistent) conclusions to be drawn from the experimental outcomes. As an illustration, consider the problem of studying the effect of Smoking on Lung Cancer. The problem may be much more complex than, whether a person smokes cigarettes or cigars (etc.) or not; "how many per day" may provide some further information but may not be enough. Is there any other member of the family who smokes? Has the family a record of smoking in the earlier generation? What type of diets this person has? Whether this person works in a closed office or is an outside worker? Age and the duration of the smoking habit, if any, as well as Sex may all be significant factors. If we need to take most of these factors in our formulation of the basic statistical model, I wonder whether we could justify the simple exponential or even
the Weibull distribution as appropriate for the failure time (i.e., the time for the onset of Cancer)? Since some of these concomitant variates are discrete (even, binary), some care should be taken into consideration in the formulation of suitable conditional survival (or hazard) functions. The celebrated proportional hazard model (PHM) due to D.R. Cox [Regression models and life tables, J.Roy. Statist. Soc.Ser.B 34 (1972), 187-220] is a valuable addition to the gallery of statistical modeling tools, and the past twenty years have witnessed a vast amount of use of this model in various areas of applications. I could only add a couple of comments on this useful technique. First, when there are a number of design and concomitant variates, if the variation in the realized values of these variables is not small and if the null hypothesis of no effect is not tenable, the PHM may not always be very appropriate. A departure from the PHM may drastically affect the conclusions derived from an assumed PHM. Or, in other words, a PHM based analysis may not be robust against plausible departure from the assumed model. Secondly, there are some situations where PHM may not be tenable. For example, consider a study with two groups: Medication and Surgery, for the treatment of clogged artery to the brain, as it generally happens to elderly people. The medication group may have a smooth hazard function while the surgery group may have a higher risk at the onset of the surgical treatment, but after that stage, there may be a sudden drop in the hazard function. Thus, the assumption of proportional hazards may not be that appropriate, and perhaps, a change-point model may have some distinct advantages over the PHM. Incorporation of concomitant variables in such studies requires additional care. The conventional linear models may not be that appropriate, and some people have considered logistic regression model and, in general, generalized linear models to bypass some of these uncomfortable points. We may refer to the recent monograph Generalized Linear Models (1989), by P. McCullagh and J. Nelder [Chapman & Hall, London] for some interesting aspects of such models. Nevertheless, such models may also not suffice in all situations arising in the context of clinical, health and environmental studies. The main difficulty may arise from the sampling scheme itself (as has been pointed out in (II)), so that the i.i.d. (independent and identically distributed) setup may become untenable. Secondly, there is a feeling that a model should also take into account the inherent biological/environmental/natural factors in a more visible way, as may be done by the use of some stochastic processes, mostly, with a Markovian structure. The recently active field of Counting Processes is an important step in this direction, and more work in this area is imminent.

IV. It is clear from (I), (II) and (III) that a challenging task is to develop appropriate statistical models which should be enough flexible to incorporate the biological or other inherent factors in a natural way and yet be amenable for simple
statistical inferential analysis. In order to include both these features, in a majority of cases, an adequate statistical model may therefore involve a large number of parameters (or a complex functional form which may not be very appealing to the applied scientists desirous of working with such a model). With the increase in the number of parameters involved in a model, the precision of statistical conclusions may go down rather drastically, unless the number of observations in the study is made to increase adequately. The rate of increase in this sample size with the number of parameters to be handled is believed to be faster than the number itself (in an i.i.d. setup), and in the current context, where i.i.d. sampling may not be the usual case, this picture may even be more complex. The actual experimental setup may not always permit this increase in \( n \). For example, in a neurophysiological study a small part of the brain of a monkey was electronically monitored to study the effect of some stimulus on the optic nerve. This small cross-section yielded 40x60x100 grids for each one of which a spike train was observed. Thus, ideally, one would have a 24,0000 dimensional counting process, and that was to be observed for several milli-seconds. That study involved a set of 12 monkeys, and we were asked to draw some statistical conclusions from the experimental data. A better mathematician in our group suggested that we use Fourier methods of analysis of these waves and limit our statistical analysis to these Fourier coefficients. A computer specialist suggested that why not we consider computer graphics and draw conclusions from these colorful graphs. I had to check temptation in going over such fancy methods, and I could only tell that with a sample size 12, we won't be able to draw any statistical conclusion for such a high-dimensional data set, unless we make some very stringent assumptions by which this high-dimensional model can be effectively reduced to a simpler model involving only 8 or 10 parameters. Although such a dimension-reduction can be incorporated in some cases, it may not be universally possible. Moreover, such reduced models are to be considered with utmost care and attention to the actual experimental setup. There are numerous examples where such reduced models would be much less informative, if not totally inappropriate. In statistical terminology, projection pursuit, partial likelihood and dimension-reduction all may be quite appropriate in the current context; however, in view of the possible complex sampling schemes (arising in such studies), it may be necessary to verify whether the needed regularity conditions are all met. This in turn is contingent on the formulation of novel statistical models as well as sampling designs for which statistical methodology developed so far can be readily amended to yield desired solutions. In this challenging task, again, the interpretations from the biological or other experimental considerations are very important in statistical resolutions of models and analysis schemes.
In a conventional setup, for drawing statistical conclusions, the Neyman-Pearsonian approach, Wald decision-theoretic approach and the Bayes methods dominate the scenario. For most of the problems discussed in Section 3, the risk functions are generally much more complex in form. The sampling schemes are usually more complex too. As such, the rationality of these approaches may openly be put to questions. Also, in many cases, a cost-benefit or some utility oriented formulation may take into account the experimental restraints and ethics in a more appropriate manner. Thus, there is a genuine need to develop more utility-function oriented formulations and then to bring in statistical methodology in a more appropriate manner to handle such a situation. As an example, we may refer back to the clinical trial setup discussed in Section 3. The decision to terminate a clinical trial may not be solely (and is rarely) based on the statistical evidence acquired from the study; clinical evidence and medical ethics play more dominant roles. A severe toxicity or side-effect of the drug used in this trial may call for an abrupt stopping of the experimentation. Imminent development of some other more potential drug may also change the course of study. Moreover, the adoption of the conventional zero-one loss or quadratic error loss functions may not appear to be that reasonable here (as the decision to advocate (or not) the particular drug may lead to a complex picture of various risks (e.g., side-effects, cost factor etc.) which need to be taken in a more effective way in the formulation of a loss function in a statistical fashion. This is clearly a signal for an utility-oriented loss function. Early termination of clinical trials has been a favorite topic for statistical discussions for almost two decades, and yet there is more room for further developments. Time-sequential procedures, progressive censoring schemes, counting processes, Kaplan-Meier estimator of survival functions, time-dependent covariates, Mantel-Haenszel procedure, group sequential methodology have all been developed in the past three decades with the primary objective of updating statistical methodology for proper adoption in a wide spectrum of applied science and technological research. Still now, largely, these are all posed in a conventional setup of structured models with i.i.d. errors. The situation may drastically change for complex sampling designs and follow-up schemes, and hence, there is an excellent prospect of extending these methodologies to a much wider situation arising in the general context of the problems treated in Section 3.

In physical sciences, statistical mechanics as well as in some biological studies, a statistical model may often be formulated in terms of appropriate stochastic differential equations; we have commented on this in connection with some neurophysiological models too in Section 3. In some of these problems, it may be possible to attach suitable physical interpretations to the algebraic constants
appearing in these equations. The physicists are notorious in this business [you may check with the Fokker-Planck diffusion equation among countably many others.] In the theory of stochastic processes too, often, these algebraic constants, appearing in Chapman-Kolmogorov differential equations, have nice physical interpretations and thereby are regarded as parameters associated with the relevant stochastic processes [you may check with the Birth and Death Process, Poisson Processes and many others.] Inspired by these developments, it may not be unnatural to think of such stochastic differential equations in the formulation of statistical models appropriate to the problems posed in Section 3. However, providing physical interpretations to the algebraic constants in such complex models may be a trifle harder problem. The neurophysiological example (with the Monkeys) may serve as a good example towards this point.

VII. There are other technical difficulties associated with the formulation of statistical methodology for fruitful analysis of experimental setups relating to the problems referred to in Section 3. For example, in the Air Pollution Problem, in the continental United States, there were nearly 50 sites where such a study was conducted; these sites differ in their geographic locations, climatic conditions, demographic and economic conditions, and in many other respect. Moreover, there are some geographical and/or climatic proximities which indicated that the observations pertaining to this array of sites were not stochastically independent. As such, there seems to be a basic issue as to how to combine the evidence from all these sites in the form of a general statement. From a statistical perspective, this amounts to combining of experiments for drawing statistical inference on the whole set. If these experiments were independent, the classical Fisher method of combining independent tests could have been adopted successfully and it would have some optimality properties too. Unfortunately, the lack of independence may stand as a barrier to a valid use of such a nice device. From an applied statistics point of view, clustering of these sites and then adopting a Fisherian detour of the step-down procedure [viz., Sen, P.K. (1983). Contributions to Statistics: Essays in Honour of Norman L. Johnson, North Holland, Amsterdam, pp. 367-377] may lead to a satisfactory resolution. In a general setup, this combination of evidence from related studies has emerged as a vital topic for statistical research with a vast potential of applications in almost all areas of applied sciences. No wonder, the currently popular terminology of Meta Analysis has drawn the attention of a large number of statistician (many of whom were originally mathematicians!), and given their input, it is a matter of diverting their attention to such applied problems, so that more satisfactory solutions can be developed for this vital problem. A second, but not less important problem is the need to develop suitable data
monitoring and data standardization techniques, so that the results from different studies may be combined more effectively. Because of the advent of modern computing facilities, there has been some good progress in this line too, and Data Management has emerged as a viable discipline with the full promise to work hand in hand with applied statistics. But, please keep in mind that an unconditional surrender to the computer packages (i.e., the Jewel Box of Statistical Programmes) may bring a disaster in our ability to make headway progress in this direction. We may be able to resolve these problems more satisfactorily if we choose to work closely with the experimental scientists (in the contemplated areas outlined in Section 3), understand the basic principles, formulate the procedure in a manner compatible with the experimental setup and try to develop adequate and novel methodology.

VIII. Because of the complexities of sampling schemes and statistical modellings for the experimental studies referred to in Section 3, the planning aspect needs a special care too. Many of the traditional aspects of statistical planning and design of experiments need to be reexamined and amended to suit the purpose. For example, randomization in an agricultural experiment is a relatively simple task, but in a clinical trial (involving volunteered patients, medical professionals and statisticians, all of whom are human beings) strict enforcement of randomization may run into conflict with the usual clinical practice. As such, it may be wiser to adopt some restricted/modified randomization principles whereby some flexibility may still be allowed without effectively violating the general setup. Do you know what Double-Blind (or Tripple-Blind) Studies mean for Clinical Trials?

Often, the experimental schemes appropriate for the studies outlined in Section 3 are follow-up (longitudinal) in nature. A batch of subjects may be followed over a period of time, and then the outcome response data may be put to statistical analysis. In Epidemiological studies, there is sometimes a Retrospective scheme which runs backward with respect to time. While it may be quite tempting to adopt the common Sequential Design of Experiments in such longitudinal studies, there are basic differences which must be kept in mind. These differences, in turn, call for a somewhat different sequential design for clinical trials and/or other longitudinal studies. The optimality criteria are generally different, the stopping rules are also generally different, and also the methodology may not run parallel. Nevertheless, the same rationality as underlying the classical sequential analysis has yielded suitable statistical procedures to handle a large class of such follow-up studies, and with more interactions between the statisticians and the experimental scientists, the situation is bound to improve vastly. Chapters 7, 12, 21, 26 and 27 of Handbook of Sequential Analysis (Edited by B.K.Ghosh and P.K.Sen, Marcel Dekker, New York, 1991) provide some general guidelines of the basic task in this direction.
Missing Values in agricultural designs have received adequate attention during the past 50 years. The same phenomenon surfaces in clinical trials, medical studies and environmental investigations in diverse and apparently more complex forms. In a follow-up scheme, a drop-out or withdrawal leads to censoring which provide only a partial information. It is possible to extract some of this through diverse statistical techniques under diverse regularity assumptions. Type I censoring (truncation), Type II censoring, random censoring, progressive censoring are a few notable cases which may crop up in such studies. It may be noted that the statistical scheme to extract information depends very much on the type of censoring as well as other compliance criteria, and there has been a spur of research activities in this domain during the past twenty years. One of the basic assumptions made in this context is the stochastic independence of the censoring and response variables - a condition that is likely to be violated in most of the cases. There is therefore a need to develop novel statistical methodology to eliminate this drawback. Confounding is another important aspect in (factorial) experiments, and sound statistical methods have been developed for taking care of this in diverse experimental setups. The situation is again much more complex with clinical/medical designs and environmental studies. In a more general setup, the question of Identifiability arises in a natural way, and confounding is just a special case of this. In clinical trials one may have more than one endpoint, and as such, there may be a genuine need to order the response variates according to the importance of the different endpoints with respect to the objectives of the study. It is also often the case that some primary variate may be difficult to record or may even be very costly, and hence, a surrogate variate is used in the study. Therefore one may question how far the statistical information can be extracted from these surrogate responses? Apart from validity, the question of loss of efficiency seems to be pertinent in this context. We may refer to a special issue (No.2) of Statistics in Medicine 8 (1989) where some of these issues have been discussed in detail.

IX. Statistical analysis has penetrated deep into the core of scientific and technological investigations. It was not too long ago, articles in medical journals used to have an acknowledgement at the end of the paper to the effect that statistical analysis (mostly, amounting to computation of mean and standard error) was made by such and such person; from statistical point of view this was inadequate and from clinical point of view too, often, the analysis used to be of very little help. As the status of statisticians has (?) improved, they are more often listed as co-authors (albeit at the end of an usually long list); they are also charged with a greater responsibility to go deeper into the planning, data monitoring and refined statistical analysis of a variety of such experimentations. This commendable, delicate task can be effectively accomplished by having a sound knowledge of statistical
methodology, thorough acquaintance with computational statistics and an indepth study of statistical inference. [How many of such jack of all trades statisticians are there?] Even in most simple models where the error components are assumed to be i.i.d., in a complex design, one may end up with suitable statistical measures whose exact sampling distributions are difficult to derive explicitly, and one may have to rely on large sample properties. In applied statistics, it may be taken for granted that large sample distributions are attracted by Normal, Chi-squared or some other simple laws, so that the task reduces to that of estimating the sampling variance or similar measures for such statistics. Resampling methods are most valuable tools in this context. They can not only provide estimates of such measures of the sampling distributions, but also, often, a nonparametric estimator of such a distribution itself. Among these resampling methods, Jackknifing and delta-method qualify for seniority, although bootstrapping has come up with its adolescence into an open world of competations with the others. The Jackknife, the Bootstrap and Other Resampling Methods, SIAM Publication, Philadelphia, 1982, by B. Efron, is an excellent introduction to this novel area, and during the past ten years there has been a phenomenal growth of research literature in this area. Although in some cases, jackknifing may not work that well but bootstrapping does so, there are other cases where an opposite picture holds. In any case, there are suitable modifications of each procedure which may work out better in such nonregular cases. The end-product is that under fairly general regularity conditions both these work out in concondance. This rosy picture may disaapear fast if we depart from the conventional i.i.d. schemes. In the field of applications contemplated in Section 3, we have noticed that i.i.d. (or SRS) schemes are not that usual, and hence, we need to fathom into the depth of such resampling methods so that they may be rendered usable in more complex sampling schemes. In the context of image analysis and/or neurophysiological studies, such resampling methods are very useful too. This scope for resampling methods should be on the top of our task list.

\( \chi \). In order to foster more effective interactions between scientists in various disciplines and statisticians (at large!), it may be wiser to arouse the statistical feeling in other sciences, and at the same time, to emphasize on the scope of applications of statistics to such a variety of disciplines by including a good consortium of such disciplines in a supporting curriculum in statistics. In the field of public health, this practice ia already adopted. Students in Epidemiology, Environmental Sciences, Health Education, Health Administration and Policy, Health behavior and Health Promotion, Nutrition, Public Health Nursing and other fields of specialization all require to take some introductory courses in Biostatistics, and Biostatistics students are, in turn, required to take number of supporting program courses in these areas. In Epidemiology and Environmental Sciences (the so called
measurement sciences component of public health), there is a far greater emphasis on biostatistics, and similarly, many of the biostatisticians are getting more and more involved with epidemiological and environmental sciences studies. The new journal: Statistics in Medicine (John Wiley, New York), since its inception about 10 years ago, has been doing a good job in fostering this interaction of statistics with medical sciences. Some of the journals published from the National Institutes of Health and its various wings have also contributed much to this needed interaction. On the other side of this fence, we have Biometrika, Biometrics and some other journals which cater more of statistical methodology applicable to a wider spectrum of applied (mostly, biological) sciences. These avenues for Applied Statistical Sciences should be kept open.

A little over sixty-five years ago, Sir R. A. Fisher in his excellent treatise: Statistical Methods for Research Workers (Oliver & Boyd, Edinburgh, 1925) aroused the scientists in a variety of disciplines of the potential impact of statistical methods in their research. With his career long devotion to genetics, eugenics, agricultural sciences and biological sciences in general, Fisher made the most outstanding contributions to the theory and applications of statistics. Even after six decades, we are trying to cross the marks of Fisher's achievements by putting his fundamental works under more coating of mathematical sophistications. But, in spirit, the gems in his work have retained their full sparkles in their adaptability to various applications. Soon after, Professor P.C. Mahalanobis initiated a highly effective research methodology in crop forecasting and related areas. Half a century later, although these methodologies still retain their utility and effectiveness, we encounter a greater task of updating them with a view to meeting the demands of the new fields of applications discussed in Section 3. Our tasks are well defined and our objectives too. Statistics started as a methodological way of analyzing data from various experimental studies, and there has been a very effective blending of theory and applications in this venture. The scenario has changed, yet in a predictable way, wherein some more challenging scientific tasks have entered into the arena and some old members have retired. To cope with these changes, we need to concentrate more on statistical theory, extend its base to include for more complex designs and more complex sampling schemes, develop more effective data monitoring tools, to foster more interactions with applications and to work out more efficient (yet simple) statistical analysis tools to accomplish our tasks. This prospect of applied statistical science seems to be indeed very bright. However, it also depends on the reciprocity of the researchers in other branches of science, technology and society in their willingness to share the scientific ideas more openly so that more appropriate statistical technology can be developed and incorporated. Only then, we will be able to meet the challenges. We look forward to this happy
ending. It is indeed a mature time to lay emphasis on Applied Statistical Sciences, to consolidate the genesis and base of this effective discipline, and to nurture its natural and effective growth by nourishments from all walks of statistical sciences. Probability theory and stochastics processes are as much important in this nurturing as the classical statistical inference and computational methodologies. Mathematical (and statistical) modelling, statistical designs and planning of experiments, practical sampling techniques, more effective descriptive statistical measures, resampling techniques to nurture the study of sampling distributions and above all the eager participation of exchange of ideas from all walks of science, society and technology are the necessary ingredients in furthering this basic objective.