INHALATION TOXICOLOGY: AWARENESS, IDENTIFIABILITY, STATISTICAL PERSPECTIVES AND RISK ASSESSMENT

by

Pranab Kumar Sen
Department of Biostatistics
University of North Carolina

Institute of Statistics
Mimeo Series No. 2134

August 1994

INHALATION TOXICOLOGY: AWARENESS, IDENTIFIABILITY, STATISTICAL PERSPECTIVES AND RISK ASSESSMENT*

BY

PRANAB KUMAR SEN
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

SUMMARY

The air we need to inhale may not be safe any more,
material fibres and carbons permeate in high score.
Auto-exhausts and industrial fumes invade our lungs,
while the alveoli nurture the immense carcinogen bugs.
Yet these are mostly identifiable and deserve our awareness,
albeit, statistical perspectives dominate the eco-wilderness.
How would you assess the inhalation risks without statistics?
In doing so, can you simply rely on bioengineering gymnastics?

AMS Subject Classifications:

Key words and phrases: Accelerated life testing; alveoli; asbestosis; automobile exhausts; bioassy; bio-equivalence; biologic markers; black lung disease; cancer; carcinogen, case study; censoring; clinical epidemiology; clinical trials; dosimetric study; ecology; environmental pollutants; environmental smoking; field study; hazard; industrial waste (exhausts); logistic regression; lung cancer; measurement errors; meta analysis; molecular biology; mutagenesis; OSL; odd ratio; pooling of information; proportional hazard model; risk; semi-parametric model; smoking; virology.

*Keynote address at the P.C. Mahalanobis Birth Centenary International Environmental Conference, Calcutta, December 20, 1993.

1. INTRODUCTION

Our environment is endangered with pollutions and toxicants of all sorts. Airborne particles, chemical (industrial) exhausts and waste products, automobile exhausts, environmental tobacco smoking and thinning of ozone layers are all adding significantly to the grave risk for human life on this planet. Although these life threatening problems are not exactly of very recent origin, their intensities have increased very drastically in the past decase or two, and awareness of this serious risk is a must for our survival. In fact, assessment of this risk is by far one of our most challenging tasks. In this quest, an interdisciplinary approach is essential, and (bio-)statistical principles play a very fundamental role. We find it convenient to introduce a few specific problems to illustrate the nature and magnitude of the inhalation toxicology problem.

a) **Galif Street, Calcutta Affair: Ink Factory Workers' Tale.**
Let me start with the Indigo, a blue vat dye obtained from the Indigo or other plants. The principal coloring matter $C_{16}H_{10}N_2O_2$ of natural Indigo is usually synthesized as a Blue Powder with a coppery luster; it used to be commonly used for the manufacture of ink powder and tablets. Laundromat use of Indigo is also common for preserving whiteness of clothes. The workers at the Galif Street Factory used to be referred to as the Blue Ghosts. Excepting for a small piece of clothing wrapped around the loin, their full bodies used to be exposed to the blue dust particles inhibiting the whole complex, and even the smell was unusual. Humidity has a profound effect on the prevailing hygenic conditions. Call it occupational hazard or not, there were something unusual:

(i) These workers were mostly out of state (e.g. from Bihar) and they used to work there only for a few years (less then ten) and return to their villages, often, with some sickness.
(ii) They used to live almost totally inside the compound (including nights) and cook their foods just across the road by the Maratha Ditch Canal slopes.
(iii) Skin diseases were very common among them.
(iv) Smoking (mostly, bidi) was very popular among them, so was the respiratory diseases (including T.B.).
(v) Not much is known about the transition of such diseases to their next generation, but suspected so.

In my opinion, the situation was primarily an inhalation toxicological one, although it might have been more complex due to water contamination (drinking) and toxicity in foods (cooking). This raises the question of awareness of occupational hazards and their grim effects beyond the normal physical/chemical levels of interactions.

b) **Black Lungs Disease: Mineworkers' Story.**
Areas of prevalence in India: Coal fields of W. Bengal and Bihar States. Sub-soil mining and drilling problems-air contamination with carbon particles and dust, water contamination. Mainly inhalation of carbon particles (and CO too) resulting in blockage of free brathing air passages

and congestion in lung membranes. Genotoxicity is quite pertinent. Smoking practice, coal-burning ovens for cooking and wintertime heating devices accelerate the progress of lung diseases.

c) **Asbestosis (Mineral Fibres):** Used as a good and noncombustible insulation material for housing and some applicances, asbestos created a lot of controversies when a few years ago it was established that such fibrous materials can cause serious blockage in the air passage to lungs, and often, this penetration goes beyond the biochemical level to some molecular levels. Mutagenesis is thus quite relevant to such an aftermath.

d) **Automobile Exhausts:** The B-B-D Bag, the commerical nucleus of Eastern India and the government quarters of W. Bengal, offers a typical scenario of air borne particles of gas and diesel exhausts with a dangerous level of CO and other pollutants. This inhalation toxicity invades all walks of life (from Akhbarbadsah to Hairpada Kerani), although the less affluent ones are exposed to a greater extent and are thereby subject to a higher risk. The picture is not that much different in other parts of Calcutta, and the effects of 'smog' can be identified clearly in wintertime. There is a vital component of carcinogens and their impacts are gradually surfacing in serious scientific studies.

e) **Garbage Disposal Dilemma:** The piled up garbage along the roadsides with hot and humid climatic conditions is quite apt for decomposition, and affects the inhalation system as well as the drinking system (through possible sub-soil water contamination). The public awareness and hygenic landfill sites are essential for minimizing the impact of such decomposed gaseous particles.

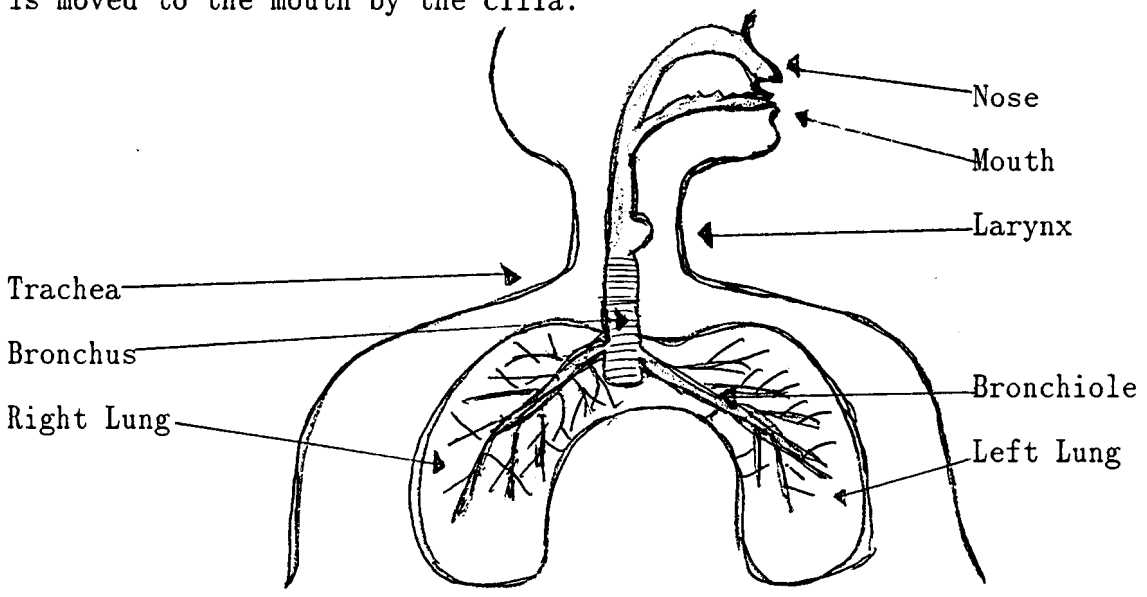
f) **Industrial Toxic Waste Products Disposition:** This includes Oil Refinaries (sulphur particles), Nuclear Power Plants (radiation effects), Iron and Steel Plants, Chemical Plants (including fertilizers and pest control products) and all other plants emitting (carbon rich) airborne particles from chimneys. These particles make their way through human respiratory organs and add to inhalation toxicity at the chemical, biochemical as well as molecular levels. Such effects can be perceived almost everywhere.

There are a thousand and one such sources of inhalation toxicity. In order to comprehend the gross impacts of inhalation toxicology, it is essential to look at our inhalation system from a biomedical perspective and then to assess the environmental impacts. Perhaps then it is easier to highlight the statistical perspectives with access to mutagenesis. Keeping these in mind, in Section 2, we present the basic biomedical explanations, followed by the environmental impacts in Section 3. Statistical considerations are laid down in Section 4. Section 5 deals with matagenesis in the content of inhalation toxicity. Assessment of risk in an important task requiring delicate statistical tools as well as sophisticated planning of the study. The last two sections deal with a motivated introduction to such (statistical) procedures. It is strongly desired to have good deal interactions between scientists (and environmental engineers) and statisticians at large, so that a mutual

understanding of this problem of grave concern may bring light at the other end of the tunnel.

2. BIOMEDICAL EXPLANATIONS

Respiration, the process of breathing, is the mechanism by which somatic cells use oxygen to produce energy and generate CO_2 and H_2O. The movement of air in and out of the lungs makes O_2 available to the body and removes CO_2 and H_2O. The respiratory system inhales air through the mouth and nose where it is warmed and moistened. Inhaled air is filtered by coarse hairs that line the nostrils which trap (larger) dust particles. Smaller particles are trapped by mucus, a sticky fluid produced by the cells lining the passage between nose and mouth. This mucosis continuously moved away by the beating of minute hairlike projections (cilia). From mouth, inhaled air travels through pharynx (throat), larynx (vocal box) and trachea (windpipe); at its base, the trachea divides into two bronchi, one entering into each lung. The trachea and bronchi are stiffened by rings of cartilage. The trachea and bronchi also produce dust-trapping mucus which is moved to the mouth by the cilia.



Within each lung, the bronchi branch out into smaller and successively into many thousands of even narrower bronchioles which lead to millions of air sacks (alveoli) of the lung tissue where gases are exchanged. Each alveoli is meshed with small blood vessels (capillaries) carrying blood from the heart containing CO_2 and H_2O. Inhaled O_2 passes into the blood, and, in return, CO_2 and H_2O are released into the alveoli to be breathed out. The O_2-rich blood in the capillaries flows into the pulmonary vein towards the heart for recirculation. After air has been breathed out, CO_2 builds up again in the blood stream. The cells in the respiratory center are quite sensitive to CO_2 concentrations. Cycles of contraction and expansion of the diaphragms (below the lungs) keep the respiratory process continuing. However, this may not always be a quite process.

The presence of dense dust (air borne) particles in the nose can trigger off sneezing. Irritants or too much mucus in trachea and bronchi

cause coughing. Speech is also a noisy breathing. The larynx is the controller of this aspect. Bronchiectasis is a chronic disorder of the bronchi and bronchioles; the tubes become weak and stretched and may not allow normal drainage of fluid secretions from the lungs. Bronchial pneumonia is a contageous infection (mostly in the bronchioles) caused by pneumoccous type virus, and lung cancer is the presence of malignant (cancerous) tumor in the lung. This tumor may even grow in the alveoli and is a serious disorder. It may not be detectable until at an advanced stage when it may be too late to arrest its malignant growth. Moreover, the cancer etiology is still not previsely known, although significant work has been in progress with respect to various influencing factors (carcinogens) promoting its incidence rates.

Toxicity in our inhalation system may arise due to various causes. Although virology used to be regarded as one of the most significant explanator in this respect, it is quite clear that there are some other important factors which deserve a serious appraisal. As was the case in Bhopal, India (Union Carbide) toxic gas leakage, often toxic substances (mostly, in the form of airborne particles) are inhaled, and they immediately start having reactions on the inhalation system, ranging from chemical to biochemical to molecular levels of penetration. Far more worse is the silent factor which cripples the inhalation system like a slow poison and may not be that perceptible in a small period of time. Sometimes we attribute such factors to our normal "aging process", although environmental awareness has cast significant concerns on such toxic pollutants some of which can be identified clearly. Often, the effects of such toxic pollutants go far beyond the biochemical levels onto the mutagenesis level. For this reason, it is imperative to bring in the relevance of environmental impacts on inhalation toxicology with due emphasis on genotoxicity or mutagenesis. A complete treatment of this complex phenomenon is beyond the scope of this writeup. Rather with due emphasis on the statistical perspectives, we provide here an appraisal of such environmental factors. Awareness and identifiability considerations dominate our depictation of the scenario. Again, in view of the emphasis on inhalation toxicology, we pay more attention to the most relevant environmental pollutants.

3. ENVIRONMENTAL IMPACTS

Although scientists and medical researchers may not have yet a solid grip on the cancer etiology, environmental impacts are being identified with greater reliability. The advents of modern molecular biology (and biotechnology) have opened the doors for fruitful research in this field, and, indeed, statistics plays a vital role in this context.

In a healthy humán, somatic cells are produced to enable growth and to replace cells that die through wear and tear; new cells are formed when existing cells divide. Normally, this division takes place at a controlled rate to maintain equilibrium. But, cancer cells divide at an uncontrolled rate forming a tumor which may spread anywhere in the body. This process may be accelerated by some carcinogens. Smoking of tobacco products (both active and passive) has been identified as carcinogenic for lung cancer.

Carcinogens damage body cells and can eventually cause at least one cell to become cancerous, which with time (i.e., over generations of cell divisions) form a 'mafia', and the cancer process initiates. Industrial chemicals (such as arsenic, asbestos, coal or tar products), diesel and gasoline products (and exhausts) can create serious environmental hazards. Occupational hazards are also evident in many chemical plants: chemical carcinogens in the form of air borne particles have grave effects from inhalation toxicological point of view. If smoking is stopped and levels of atmospheric pollutions are reduced, the chance of developing lung cancer may drop at a steady rate. Such carcinogens may have profound genotoxicity effects. Identification of such mutagens and assessment of their hazard (risks) constitute the most important task (for our survival), and a multidisciplinary approach to combat with this grave phenomenon is indeed highly desirable.

As has been mentioned earlier, the environmental impacts on inhalation toxicology can be broadly categorized as

(i) Chemical actions of environmental agents and pollutants;
(ii) biochemical impacts;
(iii) virological annexations,

and (iv) interactions at the molecular biological level. Chemical actions generally resemble the usual laboratory experimentations, excepting that the present setup relates mostly to a less controlled experimental setup, and hence there may be considerable degree of unexplained or chance variation which may obscure to a greater extent the true dose-response relation. Nevertheless, statistical planning and inference tools are generally available to extract such a "true" regression function subject to specified margins of error. Virological annexations are generally based on suitable stochastic modeling, (viz., epidemic model etc.) which incorporate stochastic variation along with the deterministic models in a generally statistically satisfactory manner. But, in the context of inhalation toxicology, the biochemical and molecular biological interactions are generally far more significant, and they pose some challenging statistical problems too.

As has been pointed out in the last section, human beings differ as much in respect to their biochemical measurements as they do with respect to their physical ones (e.g., height/weight etc.). Thus, we need to take into account interperson variations with respect to their metabolism, immunity and susceptibility to air borne particles (environmental pollutants). For this reason, a fully deterministic biomathematical model may not be very reasonable, and any such deterministic factor has to be judged in the light of intrinsic variations as is normal with the population under study. From this point of view stochastic modeling and stochastic analysis are far more appealing then the pure deterministic ones. Modern biostatistics (theory and methodology) aim to achieve this dual objective in an effective manner.

Coming back to inhalation toxicology we may note that the hairs that line the nostrils and mucus produced by the cells in the passage between nose and mouth provide a first step protection against air borne particles. Nevertheless, more refined particles make their way through the larynx and

trachea where also mucus are produced and the cilia move these dust-absorbed mucus back to the mouth. Inspite of this second step filtering, more refined particles find their way through the bronchioles all the way to the alveoli. For cigarette smokers, workers in asbestos factories, coal mines and in many other occupationally hazardous places, the inhaled air contains an excessive dose of such air borne particles containing CO and other toxic elements which hit the alveoli with high concentration and continual deposit of such toxic particles blocks the air sacks in the alveoli as well as initiates chemical actions with the CO_2 and H_2O rich blood in the lungs. This may also reduce the blood circulation and reduced flow of O_2 in the alveoli hampers the purification of the blood (from CO_2 and H_2O contaminations). From a biochemical point of view the inhaled toxic particles are quite prone to interact biochemically with somatic chemicals. At the molecular level, it may be quite appropriate to assume that alveoli blockage by toxic particles may interact with the process of somatic cell division, and may actually nurture a potentially cancerous cell which may trigger off the formation of a tumor leading to lung cancer in due time. Although still in a biochemical setup, such molecular ations require a different level of penetration wherin carcinogeneity, mutagenesis and statistics are all needed to depict the outcome variables in a proper perspective. The modern discipline of chemometry is devoted to the study of such biochemical actions, while in molecular biological studies, there is a need to lay emphasis on mutagenesis aspects as well. In this setup, it may be quite natural to stress on some 'regression' type of models, although such a regression function may be quite nonlinear (if not nonparametric) in nature. Moreover, measurement error models (both in quantitative and qualitative setups) are also relevant in this context. On the top of that such developmental processes are generally characterized by prolong incubation period, although their advanced activity period may be considerably short; often, such disorders are detected only at a very advanced stage without much hope for a successful treatment. Thus, latent effects, unobservable responses, segmented and highly inhomogeneous growth patterns and other form of nonstationarity may vitiate the scope for adaptibility of (generalized) linear models or simple Poisson type of stochastic processes, and formulation of nonparametric or nonlinear models needs to be based on biological assumptions conforming to the actual experimental setup. This brings the relevance of modern statistical theory in inhalation toxicology.

4. STATISTICAL PERSPECTIVES

Based on the discussion made in the preceding two sections, we may present the following basic statistical issues:

(i) Regression diagnostics vs. ecological perspectives.
(ii) Validity of conventional statistical analysis.
(iii) Stochastic clouds over deterministic horizons.
(iv) Measurement errors and misclassifications.
(v) Latent effects and identifiability problems.
(vi) Transformations on variables and generalized linear models.
(vii) Representitive sampling shemes.
(viii) Bioequivalence and accelerated life testing.
(ix) Genotoxicity and mutagenesis.

(x) Sensitivity and specificity issues.
(xi) Model validity and rebustness considerations.
(xii) Modes of experimentation and statistical considerations
(xiii) Optimal designs of inhalation toxicology studies.
(xiv) Data collection, monitoring and reliability.
(xv) Implementation of statistical conclusions.
(xvi) Risk assessment.

There may be numerous other pertinent statistical issues varying from individual studies to others, but in this overall assessment, we shall not go into their details.

In an agricultural or laboratory experimentation on the comparative performance of two or more treatments/factors/stimuli, it is customary to introduce a simple (linear or generalized linear) model, pertaining to the deterministic factors, mostly, in conjunction with a simple probability structure on the stochastic component. Often, a suitable transformation is made on the response variable and/or the predicting variables (or regressors), so as to be able to express

$$Y = \text{response variable} = f(\underset{\sim}{x}) + e,$$

where $\underset{\sim}{x}$ stands for the regressors, $f(\bullet)$ the regression function, and e for the error component. In a simple setup, $f(\bullet)$ is assumed to be of a given functional form (involving some unknown parameters) while a suitable assumption is also made concerning the probability law for e. To justify this basic setup, usually, $\underset{\sim}{x}$ relates to a discrete set of points, and randomization techniques are adopted to ensure that probabilitistic logic can be adapted for the e. This is the convetional regression setup which includes linear, generalized linear as well as various nonlinear models as special cases. In various inhalation toxicological studies, such a simplistic approach may not turn out to be very reasonable for various reasons. First and foremost, the lack of control is the basic factor. There may not exist any "ground zero" in the sense that an experimental place completely free from all such airborne particles may not exist. Secondly, the control of the extent of such pollutants may not be feasible, although it may be regulated to a certain extent. Thirdly, unlike the usual laboratory experiments, the exact amount of doses of such atmospheric pollutants may not be precisely determined (bring in the issues of misclassification of the doses as well as comparable order of measurement errors for these doses). Fourthly, even if such doses could have been controlled to a certain extent, the subjects under study may have great variability with respect to their exposures to such toxic doses as well as their metabolism and immunity to combat with them. Thus, simple or even complex biomathematical models may not be adequate unless the intrinsic variability in the dose levels, exposure levels and response levels are properly incorporated into such a model. Simple stochastic processes (viz., inhomogeneous Poisson or birth and death procceses) may not serve the purpose adequately for the same reasons. Ecological perspectives are therefore very important in such model selection and drawing valid (statistical) conclusions. Unlike simple physical/chemical reactions, the atmospheric toxic particles may not always have an instanteneous effect on

our inhalation stystem (unless the dose level is exceedingly high to induce so). Rather, the reaction (resembling 'slow poison') may be quite slow and it may take quite a while to detect any disorder which may be attributed directly to such toxic substances. Or in other words, the incubation period is generally long and the latent effects may make it difficult to have a proper bookkeeping on the dose and exposure records for individual subjects. There may be therefore some genuine "identifiability" problems, for the conventional statistical procedures to be adaptable without any reservation. Sixthly, the dose levels or exposure records are typically nonegative quantities, and the usual response variable (Y) is also so. This feature may generally lead to a somewhat skew-shaped distribution for the error component (e) and often to a nonlinear $f(\bullet)$. To eliminate this drawback, some transformation (on Y and/or $\underset{\sim}{x}$) are generally advocated: The log-transformation and the Box-Cox power transformations are the most popular ones. In the choice of such transformations (on Y or $\underset{\sim}{x}$) if only mathematical simplicity or empirical evidence is given the prime importance, the endproduct may not be very satisfactory: They should have good justifications from biological/ecological/environmental point of view as well. Suppose that we can consider the transformation:

$$Y \to Y^* = \phi(Y) \text{ and } \underset{\sim}{x} \to \underset{\sim}{x}^* = \underset{\sim}{\psi}(\underset{\sim}{x}),$$

where ϕ and $\underset{\sim}{\psi}$ are well defined. Then, we may rewrite

$$\begin{aligned} Y^* = \phi(Y) &= \phi(f(\underset{\sim}{x}) + e) \\ &= \phi(f(\underset{\sim}{\psi}^{-1}(\underset{\sim}{x}^*)) + e), \end{aligned}$$

which may not be representable as $\phi^*(\underset{\sim}{x}^*) + e^*$, say, with the property that e^* has a simple distribution not affected by the deterministic component $\phi^*(\underset{\sim}{x}^*)$. Thus, a simultaneous achievement of linearity of regression and close normality of the errors may not be always possible. There is, however, a more serious concern: How to choose effectively the elements of $\underset{\sim}{x}$ and how to measure them precisely? Moreover, many of the components of $\underset{\sim}{x}$ may be stochastic in nature, leading to a regression setup when both (the independent and dependent) variates are subject to (random) errors. Determination and precise measurement of Y also pose serious problems. We illustrate these points with the inhalation toxicology problem referred to earlier.

Four years ago, while visiting Indian Institute of Technology, Pawaii (Bombay), I was introduced (at the biomedical engineering laboratory) to a typical "Lungs Model" nicely formulated there. This biomedical model takes into account the lung capacity, breathing powers, a given level of atmospheric air quality and a flow rate for dust particles/CO and other gases and relates to a rate of developing lung congestion (more in the spirit of fluid dynamics). Similar models have been considered by researchers all over the world. They indeed provide a goal heuristic justification for incorporating biomathematical mmodels in inhalation toxicology. Nevertheless, such a model suffers from several important drawbacks:

(i) Is the lung capacity of a person accurately definable? Does it vary with time (age)?

(ii) The breathing power has been identified as a random variable whose distribution may depend very much on several important factors, such as smoking or not, regular physical exercise or not, asthma or other respiratory illness, age etc.

(iii) How one is going to define a given level of atmospheric air quality which is inhaled? It may depend very much on the type of occupation a person has, a place where a substantial part of the day is being spent and others.

(iv) The flow-rate for CO/dust particles or other potentially carcinogens is also not precisely defined. An active smoker may have a different rate than a passive one; in the same asbestos factory, the dust flow rate may depend on the actual type of occupation one has. Also, in big cities, the type of exhaust gasolines, thermal power smokes and other pollutants may affect different persons differently depending on their exposure level as well as immunity/body metabolism.

(v) The dependent variable (Y), the lung congestion may be obtained from X-ray plates and other medical diagnostic devices. However, judged qualitatively from such plates, there is scope for considerable 'error in measurement' of the Y's.

(vi) As has been mentioned earlier, some of the inhaled airborne particles are potentially carcinogen. How come a pure deterministic model relates precisely to such stochastic outcome (as lung cancer)?

(vii) The mechanism by which lung congestion induces a lung tumor leading to lung cancer is largely biochemical to genotoxic. Tomography is a natural candidate to penetrate into this, and statistical perspectives are therefore quite significant.

(viii) There are many other explanatory variables which may need to be taken into account in formulating a mathematical model. Nevertheless, inclusion of a large number of such variables may render the model as too complex, and may not reduce the variability to the desired extent.

(ix) Without a basic feeling of the carcinogenic process, a mathematical model may fail to match the biological presumptions.

(x) The process of collecting actual data needs to be sufficiently objective, so that probabilistic models can be formulated, and these may then yield proper conclusions.

There are many other considerations.

5. MUTAGENESIS AND INHALATION TOXICOLOGY

The past two decades have witnessed a significant upliftment of public awareness of the grave concern due to environmental impacts. Intermixed with these problems has been the fear of spreading human genetic illness, whether somatic or germinal. **Toxicology** and **epidemiology** are the two prominant scientific disciplines that have undertaken to dealing with these grave concerns, and statistical principles are indispensible in this respect too.

Genetic toxicology or **mutagenesis**, the most developed of the various toxicological subspecialities, focuses on the study of **Agents** that damage **DNA** and related genetic material. Such agents have the capability to alter

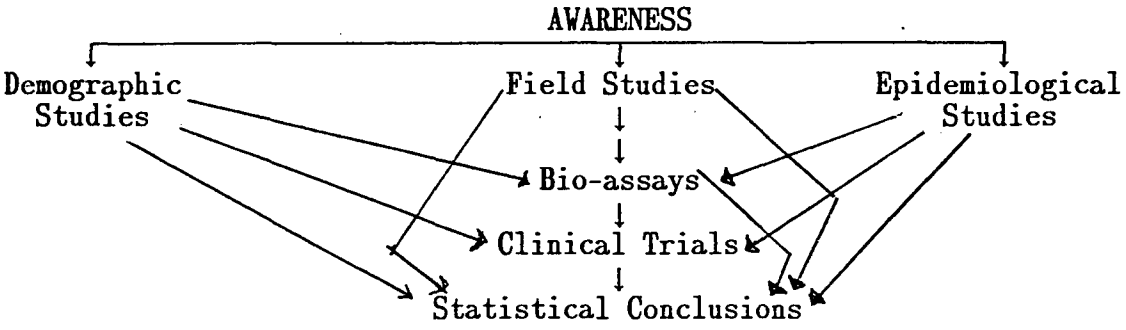
the **human gene pool** with unknown but potentially deleterious consequences for future generations. Initially, genetic toxicology focused heavily on screening environmental agents for mutagenicity and persumptive carcinogenecity. DNA damage may also be a factor in other human diseases besides cancer, and it plays a role in **aging process** too. Currently, there are increasing number of additional uses of tools of genetic toxicology to develop and evaluate **biologic markers** for epidemiologic studies, promising a whole set of new and challenging research problems as the laboratory is incorporated into epidemiologic studies. This scenario has led to a diversion of the field of classical epidemiology into

(i) **Genetic epidemiology**;

(ii) **Clinical epidemiology**;

and (iii) **environmental epidemiology.**

Each of these divisions plays a vital role in inhalation toxicology. Their basic goal is the assessment of risk for regulatory policy as well as improvement of public health. Statistics plays a fundamental role in this risk assessment. As we have noted earlier, most **toxic exposures** occur chronically at levels that are low, variable and measured with substantial error. Also, **monitoring** is time consuming, expensive and vulnerable to serious bias. There are confounding issues underlying such risk assessment, However, there are various means for gaining statistical information. Typically, a flow chart can be presented as follows:



Demographic/epidemiological studies are mostly observational ones, so that it is essential for drawing statistical conclusions that certain basic regularity conditions are satisfied. Among thse, the following are the most important ones:

(i) Quality and representativeness of collected data.

(ii) Magnitude of data-set to achieve the general goals of the study,

(iii) Elimination/reduction of subjective bias and nonsampling errors etc.,

(iv) Estimation of precision level by proper designing and monitoring of data collection.

Guided by the awareness and subsequently by demographic/field/ epidemiological studies, statistical considerations can be incorporated in

a fruitful way. Since such studies relate to human exposure to potentially carcinogens and response variable is often failure or sacrifice of life, normal experimental protocols may not be followed indiscriminately. Thus, such studies are generally conducted in several phases:

Phase I. Gathering of epidemiological evidence (without any Laboratory experimentation). There are some genuine questions regarding the interpretation of epidemiological evidence and only statistical principles underlying the design of the study scheme, collection of data set and other demographic information can lead to a proper study. This is often not the case in reality.

Phase IIA: Animal Studies (including bioassays). Often, cats, dogs, monkeys and guinea pigs are used as surrogates for human beings in such carcinogenicity studies. Their scope as well as limitations should be clearly kept in mind in drawing statistical conclusions, and trying to applying them to human populations. Although such studies can be made in a laboratory setup with greater control, they may be of little help for human studies when there may be a large number of uncontrollable factors.

Phase IIB: Dosimetric Sudies. These relate to identification of the effective dose of inhaled toxicants as well as molecular dosimetry of chemical carcinogens. Epidemiological implications and assessment of risk are both based on sound statistical principles.

Phase IIC: Mechanistic Studies. Such mechanistic approaches may relate to pathobiological effects of specific inhaled toxicants (such as fibres and tobacco-related chemicals, CO etc.) in human lung cells in vitro or to assessing (at the biochemical level) the inhalation toxicity and hazards of some specific toxicants (such as Methylisocyanate and alipathic monoisocyanates). Again, these are mostly laboratory controlled studies, and their appropriateness for human exposure study in a relatively far less controlled setup may be questionable.

Phase III: Case Studies. These are largely epidemiological nature, and are generally guided by the considerations pertaining to Phase I and II studies. In the context of inhalation toxicology and carcinogenicity, Formaldehyde, Butadiene, Benzene, Diesel and mineral fibres are important elements, while environmental tobacco smoke and exhaust emissions from automobiles and Polycyclic Aeromatic hydrocarbons have also been identified as significant agents for carcinogenicity. In order that the wealth of information gathered from such case studies can be incorporated in planning some human exposure study in an epidemiological setup, there are a number of ethical, medical and statistical issues which need to be checked carefully.

Phase IV: Clinical Epidemiology and Clinical Trials. This marriage between clinical and epidemiological approaches is often possible when from Phase I through III studies it may be possible to eliminate some of the major medical/ethical concerns through determination of appropriate dose levels/living styles whereby the subjects in the placebo/control group are not subject to any extra risk and statistical monitoring is feasible in a

manageable way. Clinical trials are now being conducted to focus on epidemiological as well as clinical undercurrents in various medical as well as environmental problems. Statistical methodology has emerged as the key factor in this dissemination of vital scientific knowledge in various areas which were once thought to be out of scope for a valid and efficient scientific investigation. In the United States, the National Institutes of Health and its various tributaries have fostered fruitful interactions in such probes during the past two decades, and more support is anticipated in the future. Environmental Protection Agency (EPA) and the National Institute of Environmental Health Sciences (NIEHS) are particularly taking bold steps in such environmental problems and the National Institute of Statistical Sciences (NISS) has also launched on reserach work with due emphasis on statistical perspectives. In mutagenesis studies, NIEHS has a special role. Similar agencies in Canada, Great Britain and some of the western European countries have joined in a global task force to combat with environmental threats, and in the third world too, there is a pressing need to institute on such activities with special emphasis on some of their local problems. Inhalation toxicology appears to be one of the vital issues in almost every country and continent as well. Since climatic conditions, socio-economic conditions and other factors vary from one place to another, there is a need to pay due attention to these factors in a proper assessment of risk.

6. ASSESSMENT OF RISKS

Statistical perspectives in inhalation toxicology, as outlined in Section 4, clearly depict the need to developing sound statistical methodology to assess the risk in an interpretable and meaningful manner. In view of the fact that there is a definite amount of stochastic element in the outcome of an event (viz, failure due to lung cancer), the assessment of risk is usually made in terms of the **probability of certain events** (which may be time dependent) or in terms of certain **intensity functions** relating to the stochastic processes arising in the flow of such an event. For example, in order to assess the impact of smoking on lung cancer, we may study the probability of developing lung cancer over a (say) ten year period for smokers and nonsmokers of a certain age group, and comapre these probabilities to draw conclusions on the **relative risk (RR)** of smoking. This approach is closely related to the classical **epidemiological risk** formulation, and we shall discuss this in detail. **The survival function (SF)** is the complement of the usual distribution function, and is characteristically related to the so called **hazard (or failure rate)** and **cumulative hazard rate.** In the context of studying the impact of environmental pollutions on lung diseases, it may be more appropriate to consider a **hazard regression (HR)** approach where the regressors are the various environmental factors (some continuous and some discrete/ordinal variables). The HR approach has been paved by the doctorine of **partial likelihood function** and the so called **proportional hazard model (PHM)**, introduced by Cox (1972, 1975).

In the context of inhalation toxicology, the **event** of concern is the development of (lung) cancer or death due to it. Thus, the **risk** can be defined as the probability of an individual's developing lung cancer over a

specified period of time, conditional on not dying from any other cause during the period; by increasing this period of time, one may define the **cumulative risk** in a similar manner. Because of the fact that there are generally numerous concomitant variates relating to an individual, such a risk is also dependent on such covariates. Thus, corresponding to an exposure period $t(>0)$ and a set of covariates $(\underset{\sim}{z})$, we may write the risk as

$$\pi(t;\ \underset{\sim}{z}),\quad t > 0,\ \underset{\sim}{z} \in \mathfrak{Z},$$

where $\pi(t;\underset{\sim}{z})$ lies between 0 and 1, and $\pi(t;\ \underset{\sim}{z})$ is nondecreasing in t. The covariate vector $\underset{\sim}{z}$ may be partitioned into two components: environmental factors and other natural covariates, i.e., $\underset{\sim}{z} = (\underset{\sim}{z}'_1,\ \underset{\sim}{z}'_2)'$. Two important statistical tasks to study are

(i) the growth of $\pi(t;\underset{\sim}{z})$ as a function of t;

(ii) comparative picture of $\pi(\bullet;\ \underset{\sim}{z})$ for different grades or levels of $\underset{\sim}{z}_1$ (and $\underset{\sim}{z}_2$).

Often, the choice of a suitable t (>0) is made on some extraneous considerations, and one considers a logistic regression model wherein it is assumed that for a chosen t,

$$\pi(t;\underset{\sim}{z}) = \{1 + \exp\ (-\{\alpha + \underset{\sim}{\beta}'\ \underset{\sim}{z}\})\}^{-1},$$

where α and $\underset{\sim}{\beta}$ may depend on t. In this setup, we may consider the logit transformation:

$$\lambda(t;\underset{\sim}{z}) = \log\ \{\pi\ (t;\ \underset{\sim}{z})/[1 - \pi(t;\ \underset{\sim}{z})]\}$$

$$= \alpha + \underset{\sim}{\beta}'\underset{\sim}{z}.$$

Through stratification, matching or other techniques, in actual practice, the covariates $(\underset{\sim}{z})$ are sorted into relatively homogenous subsets and for each subset $\lambda(t;\underset{\sim}{z})$ is estimated from the experimental outcome. Then a linear model (more precisely a generalized linear model) approach is then incorporated to estimate α and $\underset{\sim}{\beta}$ in an efficient way. These estimates may then be used to draw conclusions on the effect of environmental factors on the risk $\pi(t;\underset{\sim}{z})$. If $\underset{\sim}{z}$ has only one component assuming only two values (say, 0 and 1), one may also use the conventional odd ratio:

$$\frac{\pi(t;1)[1-\pi(t;0)]}{\pi(t;o)[1-\pi t;1)]} = \lambda^*(t),\ \text{say}.$$

There are various conventional categorical data analysis tools which may be used to draw statistical conclusions; we may refer to Sen and Singer (1993, Ch.6) for some details. In epidemiological investigations, as has been remarked earlier, sampling designs may be quite different from the conventional i.i.d. sampling scheme: case-control studies, follow-up studies, retrospective studies and some other forms of matched studies are notable cases of such designs. Kleinbaum, Kupper and Morgenstern (1982) contains an excellent introduction to such statistical analysis schemes in epidemiologic studies. The actual sampling design may even be more complex due to **censoring** (withdrawal/dropouts) and possible competing risk factors.

Kalbfleisch and Prentice (1980) and Fleish (1973) contain useful account of such research developments in a variety of simplistic situations and there is good scope for further developments to suit more complex models as may generally arise in practice.

For a d.f. F defined on $\mathbb{R}^+ = (0, \infty)$, the survival function $S(t)$ is equal to $1-F(t)$, so that $S(0)=1$, $S(\infty) = 0$ and $S(t)$ is $\downarrow$ in $t\epsilon\mathbb{R}^+$. Whenever F admits of a pdf $f(t)$, one may define the **hazard or failure rate** $h(t)$ as

$$h(t) = f(t)/S(t) = -\frac{d}{dt} \log S(t).$$

If we denote the **cumulative (or integrated) hazard rate** $H(t)$ by $H(t) = \int_0^t h(u)\, du$, $t \geq 0$ (so that $H(0) \geq 0$ and $H(\infty) = \infty$ and $h(t) = H'(t) > 0$, $\forall t \geq 0$), we have by definition

$$S(t) = \exp\{-H(t)\},\ t\epsilon\mathbb{R}.$$

As such, the risk may equivalently be assessed from the hazard rate $h(t)$ or the cumulative hazard rate $H(t)$. Cox (1972) made an elegant use of this relation in a conditional setup to formulate the **proportional hazard model (PHM)** which may be posed as follows. We define the conditional survival function as $S(t;\underset{\sim}{z}) = 1-\pi(t;\underset{\sim}{z})$ and denote the conditional hazard function as

$$h(t;\underset{\sim}{z}) = h(t;\underset{\sim}{0}) \bullet g(\underset{\sim}{z}),\ t\epsilon\mathbb{R}^+,\ \underset{\sim}{z}\epsilon\mathfrak{Z},$$

where $h(t;\underset{\sim}{0})$, $t \geq 0$ is an arbitrary nonnegative (baseline) hazard function while $g(\underset{\sim}{z})$ is a parametric function which he specifically took as

$$g(\underset{\sim}{z}) = \exp\{\underset{\sim}{\beta}'\underset{\sim}{Z}\};\quad \underset{\sim}{\beta}' = (\beta_1, \ldots \beta_p),$$

and termed this as the hazard regression (on covariates). In this setup, we encounter a **semi-parametric model** (with nonparametric $h(t;\underset{\sim}{0})$ and parametric $g(\underset{\sim}{z})$). It may be noted that the risk ratio

$$\frac{h(t;\underset{\sim}{z}_1)}{h(t;\underset{\sim}{z}_0)} = \frac{g(\underset{\sim}{z}_1)}{g(\underset{\sim}{z}_0)} = \exp\{\underset{\sim}{\beta}'(\underset{\sim}{z}_1 - \underset{\sim}{z}_0)\}$$

does not depend on $t(\epsilon\mathbb{R}^+)$, and

$$\log\{h(t;\underset{\sim}{z}_1)/h(t;\underset{\sim}{z}_0)\} = \underset{\sim}{\beta}'(\underset{\sim}{z}_1 - \underset{\sim}{z}_0)$$

provides a convenient means for studying the influence of environmental concomitant factors. The PHM model has also been used on odd-ratio measures. The recent monograph by Andersen et al. (1993) is an excellent source of such models and their statistical treatments.

There are many practical situations where such a PHM may not be apropriate at all. Moreover, much of the simplicity may be lost if the covariants $\underset{\sim}{z}$ are time-dependent and/or the coefficients β are so; a case that is very likely to arise in enviornmental studies where the covariates are not only highly stochastic but also time dependent. Moreover, the basic martingale approach providing the theoretical background may not be tenable when a simple random sampling scheme is replaced by a more complex

sampling design as may be generally necessary in environmental studies. For this reason there is a genuine need to explore more general statistical methodology as may be applied to complex sampling designs which are usually adaptable in such an inhalation study.

There are some other major issues requiring close attention. As has been indicated in Section 5, the nature of an inhalation toxicological study depends on the particular framework of Phase I, II(A, B or C), III and IV studies. The extent of "controlled experimentation" varies from one phase to another, and, in turn, their statistical modeling and inference tools may also vary considerably. **Accelerated life testing** (ALT) methodology is often tried on a Phase II study to yield some general conclusions for Phase IV studies. This should not be done without proper safeguards. Particularly, for a low prevalence rate, such ALT methodology may not be very reliable. There is another basic problem requiring sophisticated statistical analysis. This is the problem of **pooling of informatin (data)** for drawing valid and more efficient statistical conclusions. In a very simple situation arising in $k(\geq 2)$ 2 x 2 tables, the celebrated **Mantel-Haenszel (1959)** procedure provides a more efficient mode of pooling information when the basic measures are concordant in a statistical sense. The same procedure applied to a 2xK table when the K categories are ordered in a certain sense. A multivariate extension of this Mantel-Haenszel procedure to more complex models arising in clinical trials has been considered by Sen (1987). There is a good scope for adopting such prockedures in inhalation toxicological studies. the classical **observed significance level (OSL)** based methods for pooling of information may be applied in a more general setup; Sen (1983) considered a modified **step-down procedure** which has certain advantages over the classical step-down procedure [Roy, 1958], and further extensions are considered by Sen (1987). One advantage of such OSL based pooling methods is that they are amenable to more complex sampling designs as well as statistics. However, as is the case with most inhalation studies, individual estimates/statistics (to be pooled) may not be indepenent and also homogeneous to a refined degree. Thus robustness considerations may dominate the scenario. For example, if we are to pool the information from independent bio-assays conducted on different species (viz., cat, dog, monkey, etc.), we may have to keep in mind that the effect of a toxic dose in the inhalation system of these subjects may vary considerably, and hence, such inter-species variation need to be taken into account in a proper formulation of a model for pooling data. Similarly, if an inhalation study is made for different geographical sectors of a city/county/district, the assumption of independence of such estimates may not be tenable and their toxic environments may not also be comparable. For such situations, often, pooling of information is made through a parametric (viz., mixed-linear) model. Validity/robustness efficiency considerations are very important in such combination of experimental evidences. **Meta analysis** is thus a very important research topic for inhalation toxiciology studies. Much more remains to be done in this respect.

Most of the statistical procedures referred to here are either based on generalized linear models (viz., McCullagh and Nelder (1989)] or some

simple semi-parametric models [viz., Andersen et al. (1993)]. In a variety of situations, nonparametric models may appear to be more appropriate mainly because they involve less stringent regularity assumptions, and they may be more amenable to complex sampling designs as may crop up in inhalation toxicological studies. On the other hand compared to some simple parametric formulations, nonparametrics may require larger data sets. Fortunately, many environmental and epidemiological studies now a days involve adequately large data banks, and the advent of modern computers has made it easier to adopt relatively complex statistical designs arising in such environmental studies. This is a golden opportunity for (mathematical as well as applied) statisticians to looking at such sampling designs, appraising of suitable statistical models and developing efficient and robust statistical inference procedures with a view to assessing the risk. In USA, Canada and Great Britain some bold steps have already been taken to promote fruitful research in this area. In India there is a galaxy of statisticians (both younger and senior ones) who have made outstanding contributions to mathematical statistics, probability theory, design of experiments and sample surveys, among other branches. I would like to conclude this presentation with an invocation of awareness and prime interest to all these statisticians (at large) towards this area of research: theory, methodology and applications. This will be highly beneficial for our safe and prosperous living on this planet.

REFERENCES

Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993). Statistical Models based on Counting Processes. Springer-Verlag, New York.

Breslow, N.E. and Day, N.E. (1987). Statistical Methods in Cancer Research. Volume I: The Analysis of Case-Control Studies. IARC Scientific Publications. Lyon.

Breslow, N.E. and Day, N.E. (1987). Statistical Methods in Cancer Research. Volume II: The Design and Analysis of Cohort Studies. IARC Scientific Publications 82, Lyon.

Cox, D.R. (1972). Regression models and life tables (with discussion). J. Roy. Statist. Soc. B 34, 187-220.

Cox, D.R. (1975). Partial likelihood. Biometrika 62, 269-276.

Dungworth, D., Kimmerle, G., Lewkowski, J., McClellan, R. and Stober, W. (eds.) (1988). Inhalation Toxiology. Springer-Verlag, New York.

Fleiss, J.L. (1981) Statistical Methods for Rates and Proportion. John Wiley, New York.

Gart, J.J., Krewski, D., Lee, P.N., Tarone, R.E. and Wahrendorf, J. (1986) Statistical Methods in Cancer Research. Volume III: The Design and

Analysis of Long-term Animal Experiments, IARC Scientific Publications 79. Lyon.

Kalbfleisch, J.D. and Prentice, R.L. (1980). Statistical Analysis of Failure Time Data. John Wiley, New York.

Kleinbaum, D.G., Kupper L.L. and Morganstern, H. (1982). Epidemiologic Research. van Nostrandu, New York.

Mantel, N. and Haenszel, W. (1959), Statistical aspects of the analysis of data from retrospective studies of diseases. J. Nat. Cancer Inst. 22, 719-748.

McCullagh, P. and Nelder, J.A. (1989). Generalized Lineaqr Models. 2nd ed, Chapman-Hall, New York.

Roy, J. (1958). Step-down procedure in multivariate analysis Ann. Math. Statist. 29, 1177-1187.

Sen, P.K. (1983). A Fisherian detour of the stepdown procedure. In Contributions to Statistics: Essays in Honour of Norman L. Johnson (ed: P.K. Sen). North Holland, Amsterdam, pp. 367-377.

Sen, P.K. (1987). Combination of statistical tests for multivariate hypotheses against restructed alternatives. In Advances in Multivariate Statistical Analysis (eds: S. DasGupta and J.K. Ghosh). Indian Statist. Inst. Calcutta, 377-402.

Sen, P.K. (1993a). Statistical Perspectives in Clinical and Health Cciences: The Broadway to Modern Applied Statistics. J. Appl. Statist. Sc. 1, 1-50.

Sen, P.K. (1993b). Perspectives in Multivariate Nonparametrics: Conditional Functionals and ANOCOVA Models. Sankhya A55, 516-532.

Sen, P.K. and Singer, J.M. (1993). Large Sample Methods in Statistics: An Introduction with Applications. Chapman-Hall, New York.