Numeracy for Everyone

5. A Circular Path of Conjectures, Observations, Interpretation and New Conjectures about Diseases and the Causes there of

Anil P Gore and S A Paranjpe

A P Gore is currently Head, Department of Statistics, University of Pune. He initiated teaching and research in ecological statistics at Pune. He has been active for many years in promoting interaction between statisticians and biologists.

S A Paranjpe is Reader in the Department of Statistics, University of Pune. She has long experience of applying statistics in the fields of nutrition, forestry, ecology, agriculture and health.

Epidemic is occurrence of a large number of cases of a disease in a locality at a time. Similar phenomenon in animals is technically called an epizootic. But one is often not particular about this. Epidemiology is the field in which scientists play detectives and try to identify 'reasons' behind the event. This is done in many ways.

Let us consider some examples.

Cholera: In 19th century London, epidemics of cholera were not rare. In an attempt to identify any commonalties among families with cholera patients, it was found that there may be some connection with the company from which the family sourced their water supply. There were several companies in the business of city water supply. Occurrence of disease seemed to go hand in hand with water supply from some companies and not others. This was surprising since all companies pumped their water from river Thames. Further inquiry showed some companies, pumped water upstream and others downstream, where city waste water was released into the river. Clients of latter companies were more likely to get cholera. Given this diagnosis, corrective measures seem clear. Is it not? Of course scientists can go deeper into the question of what in the waste water may be the cause and this may further help prevention as well as cure. Preventive and social medicine (PSM) is a subject included in standard medical curricula in India. However, it seems to suffer from low status and neglect.

Heart disease: Epidemiologists talk about factors that influence the chance of an individual getting heart attack. The chance increases with age. Higher the cholesterol content of

blood, greater is the risk. Habits such as smoking, drinking alcohol make an individual prone to heart disease. On the other hand regular moderate exercise, balanced and just adequate diet, absence of obesity make it less likely for a person to be affected. Other things remaining constant, males may be more vulnerable than women. A fast and stressful lifestyle can make achievers more susceptible while a sedate and regular lifestyle may make one less susceptible to heart attack. All such conclusions are based on statistical analysis of sample surveys.

Sexually Transmitted Diseases: Now a days we all are very concerned about AIDS. But for a very long time public health experts have been concerned about spread of syphilis and gonorrhea which have been known to man for centuries. Epidemiologists have tried to recognize patterns in the number of cases recorded by hospitals and other facilities. Rather elaborate differential equation models are developed to understand the patterns. Let us cursorily consider one such model called the S-I-R model. 'S' stands for susceptible, the population of individuals who can get the disease. 'I' stands for infective. Susceptibles have to come in contact with the infective to get disease. 'R' stands for removed. These are cases which are cured or immune or dead or physically isolated. Numbers in each category change depending upon the interaction among individuals.

Attempts to fit such models to data on gonorrhea in USA lead analysts to realize that there were two kinds of infective individuals. The core cases were those with promiscuous habits. They tend to acquire the disease quickly and transmit it to several susceptibles before taking treatment and getting cured. The opinion therefore was that protecting the core type through regular check-up and other preventive action was essential for control of disease. The non-core patients had become infected by a stray event and were not a major threat. As a consequence, an elaborate interview was introduced in case of each patient taking antibiotic treatment. The interview helped social workers identify the core type cases which were then kept under careful follow up.

Part 1. Why Quantification?, *Resonance*, Vol.4, No.5, 19-30, 1999.
Part 2. Dice of Life, *Resonance*, Vol.4, No.9,14–23, 1999.
Part 3. Just for Ecologists, *Resonance*, Vol.5, No.5, 15-25, 2000.
Part 4. Numeracy in Research Planning, *Resonance*, Vol.5, No.11, 14-25 , 2000.

If 50% people smoke and cancer occurs randomly, then we expect about 50% patients to be smokers.

Smoking and Cancer: This has been one of the most controversial areas in epidemiology. As you probably know, tobacco is a new world plant (i.e. from American continent) and was introduced into Europe and then Asia after discovery and conquest of America. Its use spread like wildfire. In 1940's medical experts began to suspect that use of tobacco was responsible for occurrence of oral cancer and cancer of respiratory organs. Lung cancer was earlier a rare disease. In 50's it became more common. This was attributed to smoking.

How does one check this? One can take cancer cases and look into the habits of those individuals. Suppose we find that a large percentage of cancer patients are heavy smokers. Can we say that smoking 'caused' cancer? Here is a counter argument.

Smoking is a very common habit. Since percentage of smokers in society is large, it will be large among cancer patients also.

This argument will not stand if percentage of smokers among patients is much larger than in the population. If 50% people smoke and cancer occurs randomly, then we expect about 50% patients to be smokers. In a study on lung cancer and smoking habits among British male doctors, it was observed that out of 179277 doctors, 109368 (61%) were smokers. 201 doctors were detected for having lung cancer. Out of these 195 (97%) were smokers. So we have a strong reason to suspect that smoking has something to do with cancer.

Another method is to examine longevity of life among smokers and non-smokers. Thus in another study on death rates from coronary disease among British male doctors, it was observed that among nonsmokers about 55% doctors survived the age 75 years while among smokers the figure is only 44%. One can get such figures for various ages and plot what is called a survivorship curve. This curve for smokers turns out to be below that for nonsmokers (see *Figure* 1). One can of course raise doubts about how individuals in the study were selected.

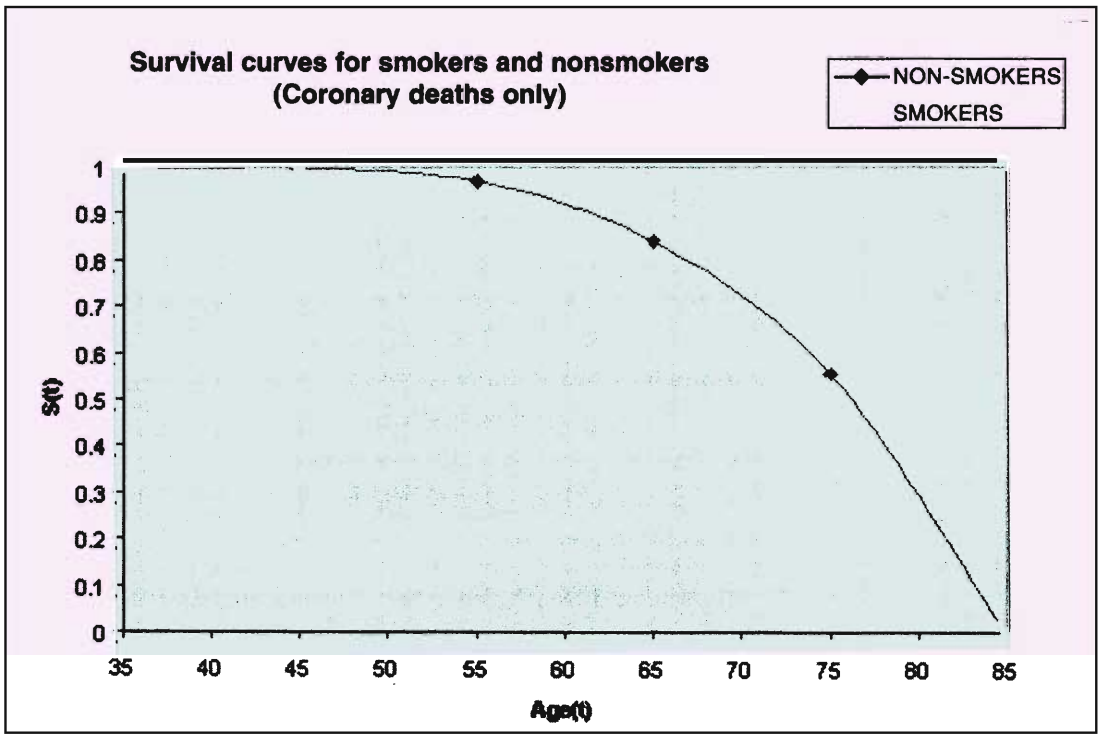


Figure 1.

In the US and UK, nearly 60,000 individuals were tracked for years to check if percentage of cases developing lung cancer is similar or different among smokers and nonsmokers. It turned out that heavy smoking increases the chance of lung cancer over twenty times. But even this and subsequent studies in fifties with massive sample sizes could not finally clinch the issue. The reasons were basically logical. One argument in defense of tobacco the 'vile weed', was that among smokers, not only were deaths due to lung cancer more common, but also deaths due to many other diseases. Tobacco, simply could not have been responsible for all these diseases. So perhaps there were some hidden biases in sample selection. Perhaps people with inclination to smoking also had a hereditary tendency to catch the diseases. If such were the case, giving up smoking does not change the genetic constitution and hence may not change the tendency to get diseases.

The only way to answer such doubts is to carry out a properly designed randomised experiment. In that experiment, randomly

The current wisdom is that tobacco consumption in any form increases the chance of getting cancer.

selected volunteers will have to agree to take up smoking at an adequately high level and continue it for years. You can see why such experiments are not possible. However, experiments were indeed done on animals and it was found that tobacco does cause lung cancer.

In midsixties, the Surgeon General of USA accepted a report of an Advisory Committee on tobacco and declared that smoking is injurious to health. The current wisdom is that tobacco consumption in any form increases the chance of getting cancer. In India today, gutkha, a mixture of condiments and tobacco has become very popular. Doctors everywhere are advising that it can cause problems in oral health rather quickly and the habit needs to be resisted.

Hygiene and Death: Have you heard of the name Florence Nightingale? She was called 'the lady with the lamp'. She brought solace to wounded soldiers of the British army which suffered very heavy casualties in the Crimean war against Russia. In addition to nursing soldiers, Sister Nightingale did something most extraordinary. She dwelt over all the information about deaths of soldiers and discovered a fact which everyone found very striking. More British soldiers died away from the battle field, in the barracks, than on the battleground. She repeated such a study in India and the same was true here too. More redcoats died sitting at home (in barracks) than while fighting. They generally succumbed to diseases caused by filth and lack of hygiene. The report recommended that proper and clean accommodation with healthy surroundings is a must for the soldiers. This lead to the development of cantonments in India. Florence Nightingale was made a Fellow of the Royal Statistical Society for her creative use of statistical data to influence public policy.

Florence Nightingale recommended that proper and clean accommodation with healthy surroundings is a must for the soldiers. This lead to the development of cantonments in India.

To illustrate how epidemiological summary of information regarding a disease helps, let us see the following paragraph about rotavirus.

Rotavirus: is the most common cause of severe diarrhoea in children all over the world. Children get this infection early in

life (below 2 years) and develop antibodies. That is why adults are generally immune to it. In developing countries 0.75 million children die of rotavirus each year. This is 6% of all deaths among children under 5 years. In India about a quarter of all hospitalized cases of diarrhoea are due to rotaviruses. Males seem more susceptible to it than females. Winter is the season favorable to rotavirus attack.

In India about a quarter of all hospitalized cases of diarrhoea are due to rotaviruses.

Antibiotics useful in other types of diarrhoea are of no use in rotavirus infection. But it is difficult to distinguish between the two infection types. Misuse or overuse of antibiotics can be harmful. The only way to identify the diarrhoea type is a biochemical test (ELISA). It can be expensive and time consuming. However, given age, sex of the patient and month of occurrence of diarrhoea, it is possible to assess the chance of the case being rotavirus. If the chance is low, antibiotics can be used. Otherwise an ELISA can be asked.

Breast Cancer and Mammography: Perhaps you may know women are vulnerable to breast cancer. As in any disease early detection of breast cancer increases effectiveness of the treatment. A test for detecting this condition is called mammography. But who should take this test? If the test is recommended indiscriminately, most will be negative and one may be accused of colluding with test laboratory. It is perhaps better to consider various risk factors such as age, age at birth of first child, number of children born, dietary habits, etc. So every general practitioner should have a simple numerical formula to calculate the risk that a given patient may get breast cancer. He should also have a threshold value. Mammography can be recommended in cases which cross the threshold. To our knowledge such values are not available for various Indian groups. Bodies such as Indian Medical Association can take a lead in development of this formula.

Every general practitioner should have a simple numerical formula to calculate the risk that a given patient may get breast cancer.

Many such numerical aids in diagnostics are possible but at least in India, they are not used commonly. Part of the reason is lack of numeracy among doctors.

Clinical Trials

Human population is exploding on earth. This is because of lowered mortality rates due to wonder drugs, vaccines and other medical innovations. Diseases like cholera, typhoid, tuberculosis, malaria, pneumonia, etc. are fully and quickly curable, thanks to medicines. In fact with vaccines we have eradicated small pox, yellow fever and other epidemic diseases. The story of how these wonder drugs are discovered is an enthralling saga of modern science.

Statistics plays a key role when it comes to testing of new drugs. In the fifties, Jonas Salk developed a vaccine against poliomyelitis. Its efficacy had to be checked before extensive use. Now polio is a disease which afflicts children and leaves a very tragic impact by way of paralysis of a limb or even death. It's epidemic occurs mainly in summer. The American President during second world war, Franklin Roosevelt, was himself a victim of polio. This was one reason behind the great impetus to research in USA on prevention as well as cure of polio.

Efficacy of polio vaccine was put to test in 1954. One million children participated in the test. Why so many? Polio is not a common disease. Even at its best (or worst?) its attack rate is only say 1 in thousand, and vaccine is expected to cause reduction in that value. Rare events are in general difficult to assess and hence the large 'sample size'. Deciding how many individual subjects are needed in a test to make it adequately sensitive is one of the most important (and difficult) questions.

Deciding how many individual subjects are needed in a test to make it adequately sensitive is one of the most important (and difficult) questions.

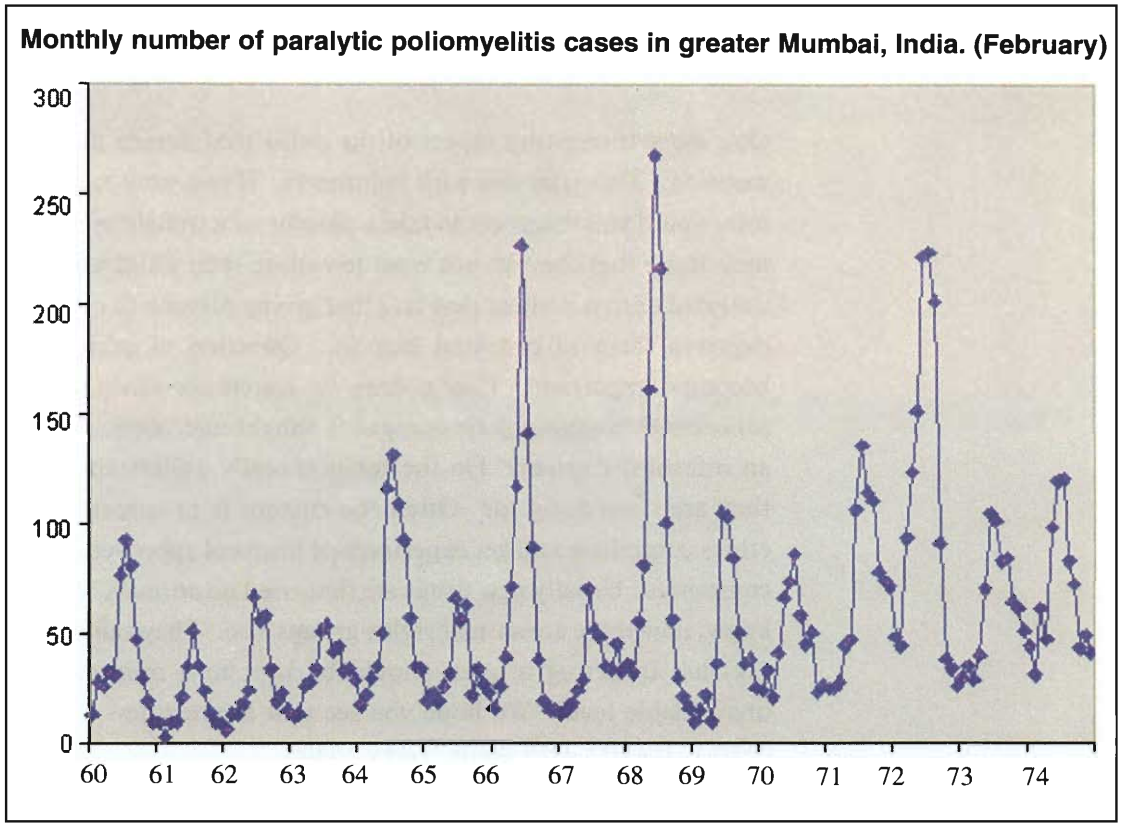
The test used placebo controls. This means some children were given vaccine and others only salt solution. The latter provide a valid bench mark for comparison. Incidence rate among inoculated children must be significantly lower than the rate among children getting placebos. Simultaneous application of vaccine and placebo ensures that comparison is in the same season and under similar conditions. If the comparison were made with incidence record in a previous year, a skeptic could have argued that any vaccine can be proved to be effective by comparison

with a really bad year. (The extent to which incidence of polio varies from month to month and year to year can be seen in the graph for Mumbai (*Figure* 2). Do you know the trick of shortening a line segment without touching it? Draw a longer line segment beside it.

Volunteer children were assigned randomly to vaccine or salt solution. This ensures that no unknown bias is allowed to vitiate the experiment. It was also a double blind trial.

Now what is this? If two drugs are being tried and everyone knows who is getting which drug, then it is an open trial. This has some risks. A patient knowing that he/she is given experimental (unproven) drug may feel stressed and this may affect the response. To protect against such a risk, patients are kept blind i.e. they are not given information about which drug they get.

Figure 2.



A double blind trial is one in which neither the patient nor the doctor knows who gets which treatment.

This is a single blind trial. In this system, only the doctor who administers the drug knows who gets what. The doctor normally would have some inclination (favourable or unfavourable) to the drug under trial. This may affect the response. A double blind trial is one in which neither the patient nor the doctor knows who gets which treatment. In a clinical trial in which we participated as statisticians, the experiment was over, the data were in. Yet we were only told that group 1 was given drug R and group 2 was given drug S. We protested. That was for some reasons. Firstly, any bias on our part cannot affect the data. Secondly, if it is found that we are biased, analysis could be done by someone else. Lastly, some finer aspects of analysis need knowledge of which is the new drug and which is well established. If the drug being tested has a response not as good as an established drug, we will discard the new drug. But even if the new drug is slightly better, we will not favour it. Demands on it are stringent. It must turn out to be 'significantly' better than the established drug. This requires a so called 'one tailed test'. But let us not get too technical.

One more interesting aspect of the polio trial design deserves mention. This trial was with volunteers. If you were to volunteer, would you volunteer to take a placebo or a trial drug? Some may argue that they do not want to expose their children to an untested entity. Others may feel that giving placebo to children deprives them of potential benefit. Question of ethics also becomes important. Can a drug be tested on unsuspecting innocents? Suppose their consent is sought and obtained, is it an informed consent? Do the subjects really understand what they are consenting to? Often the custom is to constitute an ethics committee and get experimental protocol approved by the committee. Usually new drugs are first tried on animals. As you know, now there are animal rights groups also. They can and do ask that inflicting of pain should be kept to a minimal and unavoidable level. We hope you see that the business of drug trials is fraught with many tricky angles.

Drug Screening

This refers to scrutiny of a large number of chemicals hoping to encounter one or a few useful drugs. This is nearly as difficult as searching a needle in a haystack. Often there are hundreds or even thousands of compounds which, according to scientists, have some prospect of being useful against a disease. Success rate is frustratingly low. One estimate puts it around one per ten thousand. In a search for anticancer drugs, laboratory mice in which some cancerous growth has been induced may be used, a handful of them getting each compound on test. A relatively large control group is also maintained. At the end of the experiment, cancer tumors may be removed from each mouse and weighed. Lower weight of tumors as compared to those in the control group would be taken as suggestive of usefulness of the compound. A drug on trial is either rejected i.e., further research on it is discontinued or it is included in a shortlist for further investigation.

If a drug seems promising it is subjected to a more detailed quantitative study. Its dose is varied and the effect of this variation on the disease is examined.

One can of course make errors. A useful drug may get discarded early or a useless drug may be allowed to tag along causing wastage of resources. If a drug seems promising it is subjected to a more detailed quantitative study. Its dose is varied and the effect of this variation on the disease is examined. The number of animals used at this second level experiment is usually large. If only 3 or 4 mice may have been treated with a drug in the earlier round, now the number may be raised by one order of magnitude to 30 or 40 say. So a useless compound that sneaks into this level, uses up resources to test 10 compounds at the first level. Hence it is a tricky thing to decide the criterion for selection in round 1. Statisticians constantly strive to make the decision rule more and more efficient.

One way to improve screening efficiency is to introduce what is called a sequential procedure. In the procedure described above, in round 1, a drug is either discarded or selected for round two. In a two stage procedure we have three options. Discard the drug, retain it for round two or thirdly take some more observations on it before making up your mind. This increases the cost

of testing in round 1 but reduces the chance of erroneous decision.

We have to bear in mind that no matter how intelligent the screening procedure may be, in the final analysis, its success depends on how effective the 'good' drugs are. If they are only marginally better than placebos, not much success can be expected.

Exercises

1. Prepare a questionnaire to check relationships, if any, between food habits (vegetarian/non vegetarian), number of times a person brushes teeth, use of cold drinks/ soft drinks/ ice-cream and dental health. Collect information on 50 children in a suitable age group say below 10 years and prepare cross tables of dental health and other variables.

Address for Correspondence
A P Gore & S A Paranjpe
Department of Statistics
University of Pune
Pune 411 007, India.

Suggested Reading

[1] N E Breslow, Cohort Analysis in Epidemiology, *A Celebration of Statistics*, The ISI Centenary Volume, A C Atkinson and S E Fienberg (eds), pp 109-143, 1985.

Please Note:

Physics Nobel Prize 2000
Resonance, January 2001

Page 82, 1st column, last paragraph, the line

"Another desirable effect of the lower base doping ..."

should read as

"Another desirable effect of the higher base doping..."

The error was pointed out to the author by a reader of *Resonance*, P V Venkatakrishnan, Chennai.