STATISTICAL METHODS IN PRIMARY PUBLIC HEALTH CARE

Klaus Krickeberg*, Le Chatelet, F-63270

ABSTRACT
The set of statistical methods used in primary health is related with the same domains as general public health, that is epidemiology, design and maintenance of sanitary strategies, the administration and the sanitary economy. Nevertheless, the primary attention needs certain special statistical methods which depend of the particular structures, such as the relations between public and private sectors, the equipment of the basic sanitary center, the degree of centralization of the sanitary system, and fundamentally the structuration of the communications among its parts. Some typical problems are discussed.

Key words: Epidemiology monitoring and surveillance, sensitivity, specificity.

MSC: 62P10

RESUMEN
En conjunto los métodos estadísticos que se emplean en la sanidad pública primaria conciernen los mismos dominios que en la sanidad pública general: a saber la epidemiología, el diseño y el seguimiento de estrategias sanitarias y la administración y economía sanitaria. Sin embargo , la sanidad primaria necesita ciertos métodos estadísticos especiales que dependen de su estructura particular, por ejemplo de las relaciones entre el sector estatal y el sector privado, del equipo de los centros sanitarios básicos, del grado de la centralización del sistema sanitario y sobre todo de la forma de las comunicaciones entre sus varias partes. Se discuten algunos problemas típicos.

Palabras clave: monitoreo epidemiológico y vigilancia, sensitividad, especificidad.

Almost all of us are mathematicians and so we are naturally going to start with a few definitions. By “Public Health” we mean the entirety of the activities which are directly concerned with health and which deal not with a specific single individual, for example a patient, but with a group of persons, i.e. with what the statistician calls a population. Such activities can be of various kind. They may be preventive or curative treatments which are planned and executed in the same way for all members of a population according to a unique plan. More generally, they may consists in the design, realization, follow-up, and evaluation of a health strategy, e.g. a program for the systematic vaccination of children, or a program to prevent the death of children suffering from diarrhoea induced dehydration, or a scheme for improving or rehabilitating the environment. There exist many other components of public health like health economics and health management of an entire country, of a province or of a hospital.

All the activities which I have just mentioned are based on scientific studies and research. A large part of this research deals directly with health in populations. It belongs therefore itself to public health. Much of it is of statistical nature. I would like to single out one specially important class of such studies, viz. epidemiological ones. Let us recall the definition of the field of Epidemiology too. This field consists of the investigation of the distribution, in the statistical sense, of diseases and other health defects within populations, and of the influence[1], or action of various factors on this distribution, Krickeberg [1992].

The prominent component of this definition is the concept of a ‘factor’. In order to understand the basic ideas in what follow, it will be useful to spend a bit of time by looking at various categories of factors. In the first place we are naturally thinking of aetiological and causal factors, micro-organisms, chemical substances, radiation etc. There are also risk factors tied to behaviors like the habit of smoking cigarettes.

Next we have genetic Epidemiology which is the study of genetic risk factors. The age and the sex of a person as well can be conceived of as factors which influence the distribution of diseases, and so can be of the place where a person lives and time. For example, the distribution of the factor ‘time’ amounts to the

E-mail:krik@math-info.unif-paris.fr

[1]This was the main invited lecture of the IV Conference on Operations Research held in Havana, March 2000 and to which this volume has been dedicated.

temporal evolution of the distribution of a disease.

However, from a purely formal and mathematical point of view there exist other factors of a quite different nature, viz, diagnosis and treatment. A medical treatment applied within a population, be it preventive or curative, obviously influences the distribution of the disease treated, that is the purpose. Studying this influence amounts to studying the efficacy of the treatment which is a statistical concept. It is done by the so-called clinical trials.

But does the diagnosis of a disease in the sense of a decision rule to be applied within a specific population also influence on its distribution? It does of course, because the frequency of the disease in the two subpopulations defined by a different results of the diagnosis, e.g. by “positive” (the physician claims that the disease is present) and “negative” (he asserts the absence of the disease), is normally not the same unless the diagnostic scheme on the distribution of the disease is described in the quantitative, statistical way via the concepts of its sensitivity, specificity, and predictive values. Let me also recall their definitions. The sensitivity is the probability of a positive result among the diseased persons, the specificity is the probability of a negative result among the persons free of the disease, and the two predictive values are, respectively, the probability of the disease being present when the diagnosis is positive or being absent when the diagnosis is negative. Looking at the diagnosis and treatment from such a formal and theoretical angle is in fact quite practical as we will see presently.

So far regarding public health, but what is primary health care? According to traditional ideas it consists of the contacts of the people with the health system at the basic level, below the level of specialized treatment and hospitalization. Hence it includes consultations with the family doctors and in basic health centers and policlinics. There exists a definition by the World Health Organization [WHO] in the so-called Alma Ata declaration [1978] which makes it automatically part of the public health but which is much too long and longwinded to be quoted here.

The concrete form of primary public health care varies of course enormously from country to country. I hardly know its structure in Cuba. In order not to get lost in generalities I am going to illustrate the issues at hand mainly within the framework I know best, viz. the Vietnamese health system. This system is characterized by the existence of a dense public network of fixed communal rural and urban health stations which carries primary health care and fulfills most of the functions which in other countries are being taken care of by family doctors of family medical assistants, be private or state employed. In addition it plays an important role in implementing preventive measures. There are also part-time voluntary health workers who depend on and are supervised by their communal health stations. Licensed private doctors are rare although quacks are not. This network is very well organized but, Vietnam still being a rather poor country, the stations are poorly equipped and cannot offer sufficiently many drugs. We are going to see presently that these facts have some bearing on the type of statistical problems which arise.

However, structural features carry a heavier weight in their respect. Within each province the network of health stations is run in a rigorously centralized fashion by the relevant Provincial Hygiene Department and the theory, though not altogether in practice, the whole system of primarily health care in all Vietnam is even centralized in the capital Hanoi by the National Institute of Hygiene and Epidemiology in technically supported by a Health Information System which plays a crucial role. In many so-called developed countries like Germany, France or the USA such a system does not exist. In Vietnam it consists of a regular flow of reports to be made , for the most part monthly, by the communal health stations to, among other offices, the District of Hygiene Department, from there on quarterly to the Provincial Hygiene Department, and finally from them to the National Institute of Hygiene and Epidemiology. These reports concern more or less everything: diseases diagnosed, treatments, and management, specially the economic side. They have mainly the form of statistical indicators to be computed locally by the staff of the communal health stations of the District and Province Hygiene Departments. It is above all this health information system which on one hand poses a lot of statistical problems but on the other hand allows us to treat and to solve interesting and important problems of statistical nature which, in the absence of such an information system, are not easy to deal with.

This is the frame work in which we are going to operate. Let us know enumerate the main areas of primary public health care in which relevant statistical problems come up, and then discuss some examples from some of them:

- Monitoring the state of health of the population and the general disease pattern.
- Classical epidemic surveillance.
- Planning implementing monitoring and evaluating health strategies at the primary level.

- Management of primary health care.
- Epidemiological studies.

Monitoring the state of the health and disease patterns is fundamental for the most of primary health care. We have to distinguish two types of problems. On one hand there is the hidden pathology which does not give rise to a contact with the health system, and on the other hand there are the cases which initially take place at the primary level. The type and extent of hidden pathologies depend of course very much on the quality of the medical services in general . In some developing countries even many severe cases of chronic diseases are not being taken care of by the medical system, let alone acute diseases like measles or even deaths of children by diarrhoea or pneumonia. In many countries, developed or developing ones, some practical symptomless aliments like high blood pressure largely do not come to the attention of the health worker or the physician.

The only means to estimate the prevalence of chronic diseases which escape any contact with the health system are the cross-sectional sample surveys, i.e. sample surveys at a given fixed moment in which every person in the sample is interviewed about her or his state of health at that moment. Often fairly sophisticated sampling plans are used in order to reduce costs. Sample surveys also appear in many other contexts in primary health care, e.g. to find a vaccination coverage; they constitute in fact one of its main statistical tools. However, given that they are very classical tools I will not pursue this further.

Regarding contactless cases of acute diseases the situation is much more complicated because in the ordinary cross-sectional sample surveys there will rarely be enough cases in the sample at the moment of the survey. To follow a cohort of people is hardly ever feasible in the primary context. Sometimes another method is tried, viz. a cross-sectional survey with retrospective diagnosis or retrospective autopsy, respectively. Here the question asked to a person in the sample is not whether this had been the case at some time during a certain period of the past, e.g. during the last two weeks. The answer is based on symptoms which the patient is supposed to remember. If the question is about death, family members are being consulted. This procedure obviously works only for certain diseases and even then it is marred by many mistakes. As far as I know of the International organizations like WHO or UNICEF which uses it routinely in developing countries has ever attempted to apply any of the well-understood methods of handling measuring errors in sample surveys in order to take these mistakes into account. To lay the basis for such an application would certainly be a very worthy statistical research project. I would naturally involve a Bayesian approach. The priori information needed would come from small special studies where the retrospective diagnosis can be checked. The type of study can easily be done as part of normal clinical activities, specially it here are medical registers or records as they exist within the health information system. We will illustrate the same idea a bit later in the framework of another problem.

Let us pass to the study of disease patterns as they come out of contacts with the primary health system. If there is no health information system covering the diseases in question, we are again left with sampling methods. I would like to look now at the situation where such an information system does not exist, as is the case of Vietnam. At first sight we might think that the problem at hand were a pure problem of data transmission and that the statistical methods were reduced to adding up figures in order to calculate the indicators we are after. For example in order to find the number of cases of measles in the given month, just count the number of consultations with that diagnosis ‘measles’ during that month, given that in Vietnam a case of measles usually does lead to a consultation in a health station. However a closer look reveals that we are again facing the problem of errors and gaps which arise during the transmission of indicators already computed.

What are the main sources of errors? The most important of course wrong diagnosis. They occur in all countries and on the secondary level as well but they present a particularly serious problem in primary health care of developing countries. They depend above all on the equipment of the primary health facilities and the absence of sufficient equipment , on the training and the experience of the health personnel. In a country like Vietnam, which is in many aspects highly developed but is poor materially, diagnosis in health station s are almost all purely clinical. The health statistics have no laboratories and the use of outside laboratories at the secondary level is severely restricted. Let us take two examples, measles and shigellosis, i.e. bacterial dysentery. Both are mainly diagnosed at the primary level. The official health statistics as issued by the Ministry of Health reflect the number of diagnosed cases which have been reported through the information system. However, we know or can find out, within certain margins, to which extent these clinical diagnosis are wrong. More precisely spoken, we can estimate the predictive values of the clinical diagnostic procedures used by the average health worker by conducting a few statistical studies where the clinical diagnosis is confronted with the result of a laboratory test. Let me just recall that such statistical analyses follow the schemes of general epidemiologic studies because as we have seen in the beginning, they are in fact about the influence of the factor “ diagnosis” . In Vietnam they have been done for measles and should not be hard for shigellosis. It turns out that the estimated predictive values do not vary too much between health stations, and it is then a simple application of Bayes' theorem which leads to much improved estimates of the true incidence of these diseases.

An even more interesting and useful application of these ideas is the comparison of several different clinical diagnostic schemes in order to select the "best" one in some statistical sense. For example we might look for a standardized clinical diagnostic scheme which yields, again in the statistical sense, the best estimates of the true incidence. WHO advocates standardized clinical diagnostic schemes in the framework of the program CAD [Control of diarrhoeid diseases] and ARI [Acute respiratory diseases], based on the general medical experience and no doubt with view of high sensitivity and specificity studies.The problem is particularly urgent for shigellosis because physicians now use very different clinical diagnostic schemes, i.e. the way in which they take into account the basic symptoms "headache, fever, blood in the stool, mucus in the stool" varies widely between countries. We will return to this question.

With these problem we have in fact already reached our third domain in which statistical problems abound, viz. primary health strategies. We will comeback to a similar problem in that context, using the example of malaria. Before doing so however, let me take up the question of other errors in the information system, different from wrong diagnosis. As hinted at before, in a perfect computerized information system using a nation-wide network of large capacity cables, in theory the only errors arising are those of missing or wrong entries at the primary level, and that may already be serious enough. The problem is aggravated by the fact that in many countries by far too much information is being asked for which is neither evaluated nor used and never will. Hence the health personnel in the primary institutions is not very motivated and often has no time , to provide all the basic data correctly. If, moreover, the information system is partly or entirely functioning on paper as it is and will be for a long time in most developing countries, we are facing the additional problem of missing reports and or errors when copying and consolidating , i.e. aggregating reports.

Let me illustrate this problem, and an attempt to solve it, with the help of a deliberately simplified example. Suppose we would like to know the number of diagnosed cases of dengue fever in a given month and a given district. This district consists of 12 communes each of which has a health station. Unfortunately only 7 of them have sent their report for this month to the District Hygiene Department. The common procedure is to add up the incidences, i.e. the number of cases, for these 7 communes, and to report this sum to the Provincial Hygiene Department as the incidence for the entire district. Incidence figures appearing in official health statistics including those of WHO are usually originated in this way. In the "World Health Statistics Annual 1980/81" [1981]. I have seen the incidence of influenza in all France for the whole year given as 1 [one]: that was the case which by chance got reported. Such figures are of course absurd, they are gross under-estimations. Nevertheless, a bureaucrat would be loath the correct then but for a mathematical statistician it is obvious that we can generally get much closer to the true incidences by using appropriate estimates. The most elementary way would be a trivial extrapolation based on population figures. There are more sophisticated methods which take into account the incidences observed in the past in each of the 12 communes including seasonal patterns. Going one step further we could also take into consideration the geographical structure of the 12 communes and regard the problem as one of interpolating a stochastic process in the plane.

For isolated statistical studies in the usual sense there exist indeed many theories and practical methods for handling missing or erroneous data including software, but here the situation is different: we do not deal with a single and limited study, but we are confronted with this kind of analysis every month. Hence we are looking for a method which can be applied routinely under the prevailing conditions. The District Hygiene Department rarely has a computer, and even if it has, it will not always be feasible to use very complicated software. So we are confronted with the following challenging statistical problem: given the particular structure and material conditions of the system of primary health care including its health information system, develop algorithms for statistical error correction which can be applied routinely at minimal cost.

As a second domain where statistical surveillance issues play a prominent role we have mentioned classical epidemic surveillance. In principle this is part of the first domain, viz. of the one of monitoring disease patterns, and it is largely also situated at the primary level, but it is marked by three additional particular features; the first one is speed of transmission of the relevant information, the second one is a mechanism in order to rapidly identify unusual events like an unusual increase of the incidence of such and such disease, and the third one are rules for action in response to such information. The second feature is entirely statistical and can be treated as a problem of level-crossings of the stochastic process. This has been done by various people and also put into practice in some countries, e.g. in some of the States of the USA although not on the federal level. I do not think that there are any challenging basic problem left and will therefore not dwell on the matter much more. Y think the only interesting unresolved question left now is how to integrate epidemic surveillance into general information systems for primary health care.

Now to the third domain, health strategies at the primary level. This is vast field indeed, and as said before Y will proceed by examples. I have participated as a UNICEF consultant in various primary health programs in Vietnam and Cambodia and I am going to draw on my own experiences. Sometimes the statistical problems one encounters are most elementary. That is the case in the above-mentioned classical programs CDD and ARI which concern acute disease and are, by the way, very similar to each other in operational respect although the aetiologicy and the manifestations of the two illness are completely different. Their statistical components are reduced to monitoring incidence and mortality of the various forms of these diseases and to following-up the treatment and the logistics of the drugs involved which are oral re-hydration packages for CDD and antibiotics for ARI.

Programs regarding chronic diseases like malaria and tuberculosis present already much more interesting statistical problems. Again, the two programs at hand are operationally quite close to each other, hence let me confine myself to malaria. In many developing countries including Vietnam and Cambodia the treatment of suspected malaria cases needs to be started immediately even before the presence or absence of plasmodia in the blood can be ascertained by a microscopic examination of the blood sample or by some more modern method. Most communal health stations do not have a microscope, and sending the blood sample to the District Malaria Station and waiting for the result may take a long time, especially in remote areas where malaria is endemic. Hence it is crucial to work with a good preliminary clinical diagnostic scheme, and that is a problem at which we have already looked in the context of measles and shigellosis. We want a scheme of very high sensitivity because a delayed treatment of certain forms of malaria often entail death, but also a reasonably high specificity in order to avoid costly and heavy treatments of healthy subjects. In the present context of malaria and in contrast to measles and shigellosis, we need not even start a special study to find these characteristics of a given diagnostic scheme because eventually the presence or absence of the disease will be known and documented at the health station for practically every patient; hence we can do with a study which is entirely based on the records or registers which underlie the information system. At the same time we can again investigate, using the same cases , the characteristics including the predictive values of other fictious diagnostic schemes which are also based exclusively on clinical symptoms, and then select the one which appears to be best in view of the actual implementation in daily practice.

Malaria is a good example where not only the diagnostic part of the strategy is situated within primary health care and poses relevant statistical problems but where almost the entire strategy takes place on the primary level apart from hospitalized cases. The usual strategy in most countries consists of a combination of preventive measures with early treatment. The main preventive measure for the last 15 years or so has been treated bed-nets which are supposed to repel or even kill the vectors, i.e. the mosquitoes “anopheles”, by the chemical substance with which they are impregnated, and at the same time they separate them from man in a purely mechanical fashion. Early treatment often means, as we have seen, treatment by drugs before the final diagnosis is established. In both components of the strategy I have deliberately disregarded phonemes of the resistance of the anopheles or the plasmodia respectively.

Now, how we evaluate the whole strategy which consists of all of these components; prevention, preliminary clinical diagnosis, and treatment? Evaluating a clinical diagnostic scheme had been a classical statistical problem. The statistical unit have been a single patient. Basically we were comparing different patients who behave independently of each other. The whole strategy, however, is not applied to a single person. It concerns a community, e.g. a village or a province or region; it is Public Health in the very strict sense. The statistical units in studying the effect of treated bed-nets are no single persons and not even families because there may have been selected by a sampling plan, the mosquitoes repelled there might attack families without a bed-net more intensively than before. Regarding the evaluation of the treatment , we might think at first that it suffices to study its effects in different patients regarded as independent but again this is not true because there exists what is called the indirect effect of the treatment strategy in the entire community: if some patients are rapidly cured , they will no longer transfer plasmodia to the anopheles by which they are bitten, and this reduces the infective potential within the whole community.

Hence the natural statistical unit in the study of such a strategy would be a community, not a case or a patient. We would have to compare communities in which different strategies are being applied. So-called community trials of this kind are often done by various international organizations but as far a I can see on somewhat shaky statistical bases, there are hardly even enough communities to be considered as independent and to be compared. We are facing a particular case of so-called intervention trials. To strengthen their statistical bases would be a worthwhile enterprise.

In order to get an idea about the direction in which one could go for doing this, let us look at a closely related intervention problem, viz., vaccination trials in primary health care. For many diseases, a vaccination campaign in a community does not only have a direct effect by immunizing individual people and thus protecting them from infection but , also an indirect one because an immunized person who does not contract the disease can normally not infect others or in the case of vector-born diseases cannot infect the vector.

A solid mathematical-statistical theory of malaria vaccination trials was developed a few years ago not far from here, at Emory University in Atlanta, Georgia: Halloran (1991). It starts with a model for the spread of the disease in the host populations, man and anopheles, which takes direct and indirect effects of the vaccination into account. Such a model belongs to the theory of dynamical systems and consists of a set of partial differential equations. It contains certain parameters, some of them describing the vaccination strategy, and the behavior of the solutions is studied as a function of the parameters. The statistical part is made up by methods for estimating the parameters in the concrete situations. Three months ago I met one of the authors of this theory and asked her whether it had actually been used in one of the past or ongoing malaria vaccination trials. The answer was “no”. I think that people conducting such trials often do not understand the idea of modeling and do not see the necessity of employing indirect effects correctly. In any case, a solid theory of community trials should also start out with a model of the underlying situation.

Before leaving the domain of statistical problems in health strategies let me just mention a last example, viz. drug surveillance. I am thinking of drugs which are being used systematically within a particular program or strategy of primary health care, following within a particularly defined therapeutic scheme. The surveillance concerns both secondary effects and resistance. Sometimes a few cases of these are already cause of alarm and sometimes we are only interested in unusual increases of a certain magnitude, but in any event we are facing statistical problems not unlike the ones in epidemic surveillance. Let me describe an example in which I encountered in a paedriatic policlinic in Phnom Penh and which involves two drugs. There were many cases of dysentery and no facilities for a sufficiently rapid stool analysis. The physicians thought that on the basis of clinical symptoms alone they could not distinguish with satisfactory specificity between bacterial dysentery i.e. shigellosis, and amoebian dysentery. Hence the following strategy was adopted in that clinic: administer always both a drug against shigellosis, i.e. an antibiotic, and an antiamoebian drug. This is of course a waste and, worse, there might exist synergetic phenomena. Each of the drugs had certainly been systematically tested in view of side effects , but the pharmaceutical companies involved had hardly imagined that they would be given together. Hence apparently nothing was known about adverse reactions due to synergy between the two medications. Monitoring such effects should be an urgent matter; it could be done with the help of a health information system. The alternative would be to change the strategy as outlined earlier based on studies of the diagnostic schemes employed.

I will not dwell on statistical problems in primary health economy and management. There are of course interesting mathematical problems in this area, e.g. from operations research, but as far as I can see the statistical part in them is always fairly elementary. It is exclusively composed of classical descriptive statistics. Hence I will skip this topic and turn to the last domain, epidemiologic studies.

We have in fact already touched upon this subject on several occasions in particular practical contexts. Let me recall that in the beginning we had defined the concept of Epidemiology and of the action of factors in a very broad way. This implies that monitoring the disease pattern on the primary level including epidemic surveillance, or investigating diagnostic and therapeutic schemes applied to individual patients or evaluating health strategies directed at populations including vaccination trials, is in fact an epidemiologic study. A characteristic feature of studies of those types is that by their very nature they are imbedded in primary health care. Some of them can even be conducted within the routine activities of the health services, specially so if there exists an information system.

By way of contrast epidemiologic studies in the usual more restricted sense are investigations set up ad hoc in order to answer a specific isolated question. A typical example: is coffee drinking a risk factor for cancer of pancreas? a typical study of this kind would start by selecting cohorts of persons to be followed over years, or by constituting groups of cases and of controls to be compared. Although many results of such studies find applications in primary health care the studies themselves are normally not situated within it.

Nevertheless, sometimes it can be fruitful to use the framework of primary health care in order to conduct epidemiologic studies which are designed in view of responding to a particular well-posed problem. Again I will be discussing a few examples, passing from most elementary investigations to fairly sophisticated ones.

About two decades ago the concept of “Epidemiology at the basis”, i.e. on the primary level, appeared in various documents of WHO. Unfortunately it has been largely forgotten by now; even some high officials of WHO do not know it. I think, however, that it can be very useful in primary health care by providing insight and motivation to the health personnel. It goes beyond the trivial making of charts. Maps are often more practical: by drawing and comparing simple maps the personnel of commune health station can illustrate and sometimes even discover relations between such and such risk factor and given disease. For example they may make maps representing stagnant waters on one hand and a map showing all cases of shigellosis on the other.

This is elementary stuff. On larger scale, going beyond a single commune, the analysis of maps is only slightly more difficult and largely restricted to the study of the factors “location” and “time”, i.e. to the so-called descriptive epidemiology. The data come very often from primary health care; a good example is malaria.

However, even classical descriptive Epidemiology non-trivial problems do arise. A well known recent one: relations between observed incidence and unknown prevalences. This is of course related to the good old Lexis diagrams in demography. The problem came up again in the Epidemiology of AIDS: to find the prevalence of seropositivity from reported cases. In this context the so-called back-calculation method was developed which involves solving and integral equation, Brockmayer and Gail [1994]. Now, with AIDS we are not necessarily in the domain of primary health care but as it happens the underlying ideas had already been employed a long time before in a more crude form the epidemiology of tuberculosis. They now return to this area in their refined mathematical form developed for AIDS. The basic data needed are all there, in the tuberculosis registers kept at the primary level.

Let me make some general remarks about primary health care as a vehicle for epidemiologic studies, in particular, if an information system exists. As I had recalled earlier, classical epidemiologic studies use two types of design. In one of them we are setting up cohorts of “subject”, i.e. of persons, to be followed over time. By its very nature such a “cohort study” runs over long periods which may make the follow-up of the subjects quite difficult. In the other type of design we take a group of “cases” and compare it with a suitable chosen group of “controls”, i.e. subjects which had not been afflicted by the disease in question. This looks much easier. Such a “case-control study” conducted about twenty years ago seemed to indicate that coffee drinking is indeed a risk factor for cancer of the pancreas, and that in fact roughly half of all cases of that form of cancer can be attributed to coffee, McMahon, Yan, Tricopoulos, Warren and Nardi [1981]. However, all later studies failed to find such an effect of coffee and epidemiologists now agree that we may drink coffee regularly without having to worry about our pancreas.

What had gone wrong in the earlier study? There may be have been several causes for its failure. In fact case-control studies have been attacked for a long time for a variety of reasons. One of the main problems is the selection of the controls so as to form a group which is more or less representative of the whole underlying population. In particular, recruiting controls among hospitalized persons as it has often been done usually gives rise to a bias which was already pointed out in 1946 by Berkson and is sometimes quoted as Berkson’s paradox although it s not a paradox at all.

Hospitals, however, that means secondary health care. The information systems in primary health care cover a much larger portion of the population and lend themselves much better to the selection of groups and cohorts to be studied without introducing an important bias. In particular I think that they can be a suitable basis for more recent and more sophisticated designs of studies which are about to replace the classical case-control studies. So called hybrid studies, in particular case-base studies, have already been employed in order to estimate vaccination efficacious with he help of a primary health information system. In such a study a group of cases is compared with a sample of subjects drawn from the entire population, cases and non-cases confounded. An even more recent design leads to case-control studies going on over time where cases and controls are being sampled for pre-defined cohorts, Langholz and Goldstein (to appear). Their analysis requires fairly advanced mathematical-statistical tools because the underlying likelihood-functions are getting quite complicated. Regarding cohort studies an information system is a natural vehicle for retrospective ones where the cohorts are defined within the registers for a certain moment of the past, and then followed up to the present, again with the help of registers. These studies cover of course only the influence of factors which had been recorded in the registers of the information system, but the evaluation of diagnostic and therapeutic schemes discusses earlier is exactly of that type.

Let me conclude by coming back to more difficult problems involving a modeling approach via systems of partial differential equations. The main one which many open questions left, is the dynamics of infectious diseases in populations including the effect of interventions at which we had looked earlier from the angle of health strategies but which is equally interesting as a purely epidemiologic problem. Measles is a noteworthy example but more extensive work focuses on diseases where the pathogenic agent passes through several hosts like dengue fever of the plague; recently Lyma disease has attracted much attention. The verification of the model and the estimation of the parameters are large based on data coming from primary health care.

Even the questions asked about interventions, e.g. about vaccinations, have nowadays become more sophisticated. For example one does no longer simply ask about the efficacy of a measles vaccination as a single number but about the temporal evolution of the efficacy to be estimated from an epidemic which means again primary health care.

Since we are speaking about measles, I would like to mention that very interesting conclusions about the dynamics of the disease have been drawn in an elementary fashion , without modeling, by re-analyzing primary health care data established during a measles epidemic in a German village in the year 1861 (Oesterle, E.: Statistiche Reanalyse einer Masernepidemie 1861 in Hagelloch).

To sum up let me say that statistical problems and methods in primary public health care present a large variety of faces, and that they range from the completely elementary to the most sophisticated many mathematical statistics problems are left, but even more possibilities are for novel applications.

REFERENCES

ALMA ATA (1981): “Les soins de santè primaires”, **Santè pour tous**. N WHO, Geneva.

BROOKMEYER, R. and M.H. GAIL (1994): AIDS Epidemiology: A Quantitative Approach. Oxford University Press, New York.

Halloran, E. (1991): “Evaluating interventions against Malaria and other vector-borne diseases”, **Bull. Inter. Stat. Inst. Proceedings of the 48th Session of ISI**, LIV(1), 2-9, Cairo.

KRICKEBERG, K. (1992): “Moderne Epidemiologie und ihre Anwendungen”, **Leopoldina** (R3), 35, 1989, 149-160.

LANGHOLZ, B. and L. GOLDSTEIN (to appear): “Case-control studies from a cohort sampling perspective”, **Statistics in Biology and health**. Springer Verlag, N. York.

McMANON, B.; S. YEN; D. TRICHOPOULOS; K. WARREN and G. NARDI (1991): “Coffee and cancer of the pancreas”, **The New England J. of Medicine,** 304, 630-633.

Oesterle, H. (----): “Statistiche Reanalyse einer Masernepidemie 1861 in Hagelloch”. Medizinische Fakultät Tübingen.

WORLD HEALTH STATISTICS ANNUAL 1980/81 (12981): “Infectious diseases: cases”, WHO, Geneva.