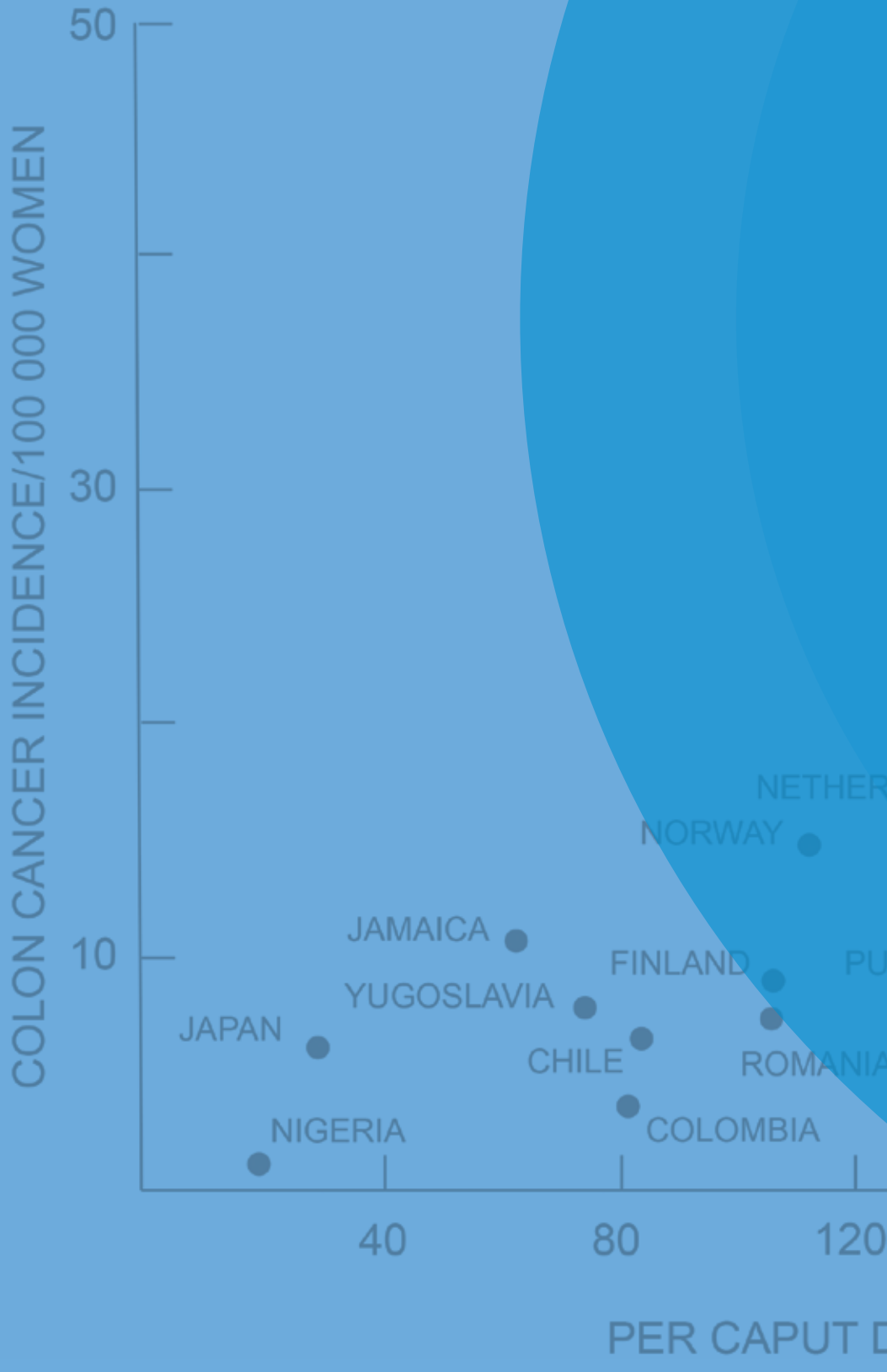
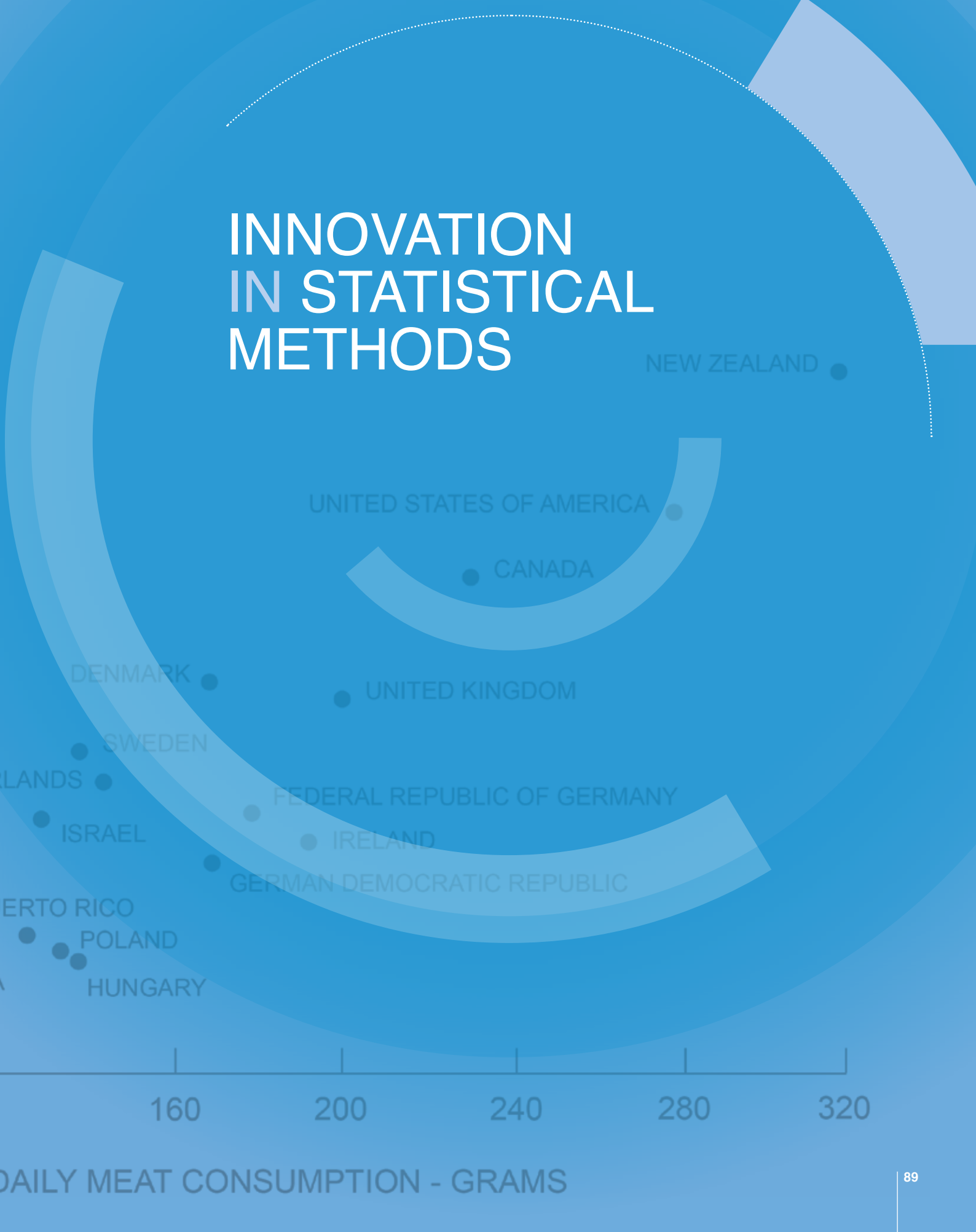

50
30
10
COLON CANCER INCIDENCE/100 000 WOMEN
NETHER
NORWAY
JAMAICA
FINLAND
PU
YUGOSLAVIA
JAPAN
CHILE
ROMANIA
NIGERIA
COLOMBIA
40
80
120
PER CAPUT D

INNOVATION IN STATISTICAL METHODS



INNOVATION IN STATISTICAL METHODS

At the time when IARC was established, in 1965, there was an obvious need for competence in the design and statistical analysis of laboratory experiments and epidemiological and clinical studies of cancer, and of diseases generally. This need was hard to meet because of a shortage of qualified people, exacerbated by the demands of developing methods well suited to studies of noncommunicable diseases and of mastering the potential of electronic computers, recently introduced and still unfamiliar.

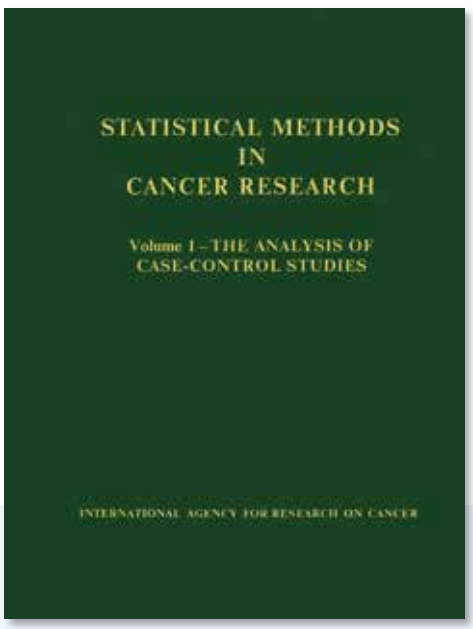
To meet this need, work on several main topics in statistical methodology was soon initiated within IARC's areas of research. The conditions at IARC, like at other institutions, reflected the status of information technology in the late 1960s and early 1970s. Jacques Estève, who was the head of information technology at IARC at that time, recalls: "I had to introduce the first data management system. In the first years of IARC, things had been pretty disordered. The epidemiologists were unhappy as they had great difficulty in retrieving their data; once they were entered into the computer, how to access them was a kind of practical mystery. The computer installation to support the new data management system occupied a large room, and provided much less computing power than today's smallest laptop. Yet over a few years the data management performance was transformed for the better." This was only the first of a series of transformations that kept IARC's computing system on a par with the constantly evolving technology.

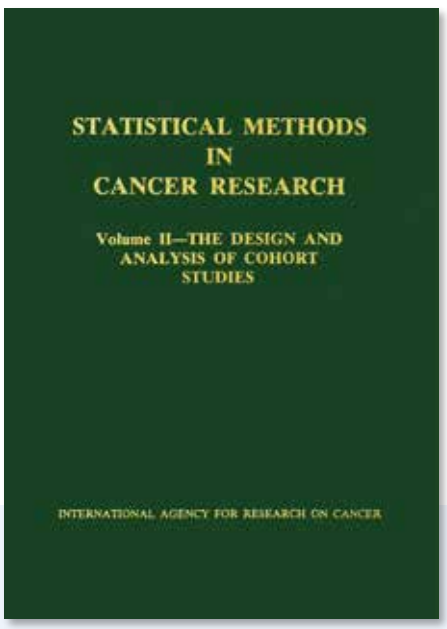
Students at work with mechanical calculating machines, the tools usually available in the late 1960s for statistical analyses of epidemiological data sets. Beyond the four arithmetic operations, these machines could calculate the sum of a sequence and the sum of the products of two sequences of numbers.

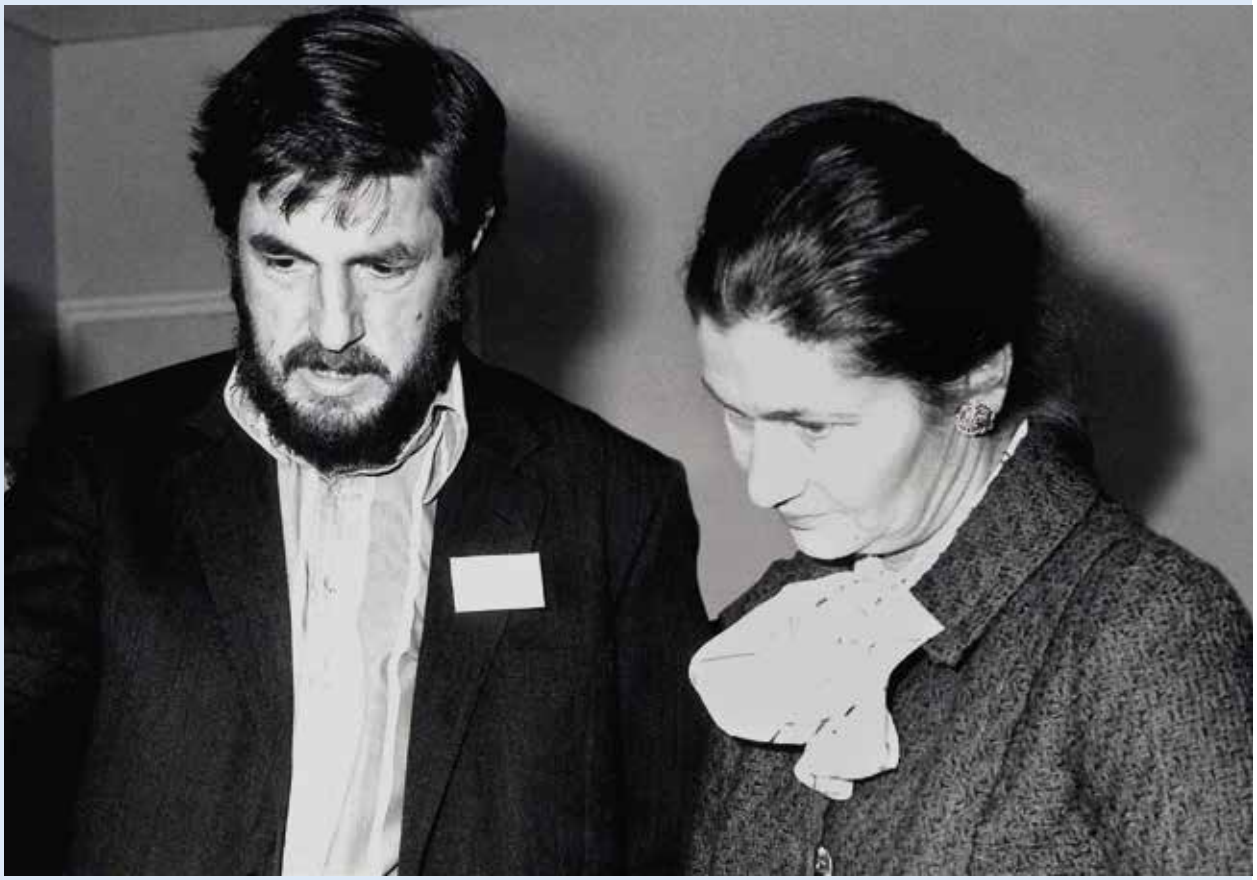
A UNIFIED FRAMEWORK FOR EPIDEMIOLOGICAL STUDIES OF CANCER ETIOLOGY

Epidemiological studies aimed at investigating causes of cancer were – and are – at the core of IARC's research. The development at IARC of statistical methodology for etiological studies produced notable results, some of which have proven to be of lasting value as they still serve as key references. This applies especially to *Statistical Methods in Cancer Research*, by Norman Breslow and Nick Day, published in two volumes: *The Analysis of Case–Control Studies* in 1980 and *The Design and Analysis of Cohort Studies* in 1987. The book is still available on the IARC website and is, quite reasonably, characterized as a classic text in the field (see "Frontline statistical research: Norman Breslow and Nick Day").

Several factors combined to make the book a success. First, it was timely. The title refers to cancer research in general, but in fact the book deals essentially with statistical methods for cancer epidemiology (although some of the methods, such as survival analysis, can also be applied to animal experiments). In cancer epidemiology, methodological innovations had been flourishing since the 1950s, aimed at solving specific problems of data analysis. However, the connections between the different new methods were not obvious, and their relative merits and limits of applicability were not well defined. In Breslow and Day's book, these methods, which had been scattered among articles in statistical and epidemiological journals, were critically reviewed and related to one another in a logically coherent framework. Second, in doing so, the authors frequently used original results from their own methodological research. Third, the presentation was at a







Simone Veil, then French Minister of Health, visited IARC on 3 November 1975 to open a symposium on environmental pollution and carcinogenic risks. Here, Nick Day shows Veil a computer display of epidemiological data analyses. Between 1970 and 1986, Day and his team made IARC a world-leading centre in biostatistics. Day subsequently moved to Cambridge, United Kingdom, where he spent the second part of his career as a professor of public health and epidemiology.

FRONTLINE STATISTICAL RESEARCH: NORMAN BRESLOW AND NICK DAY

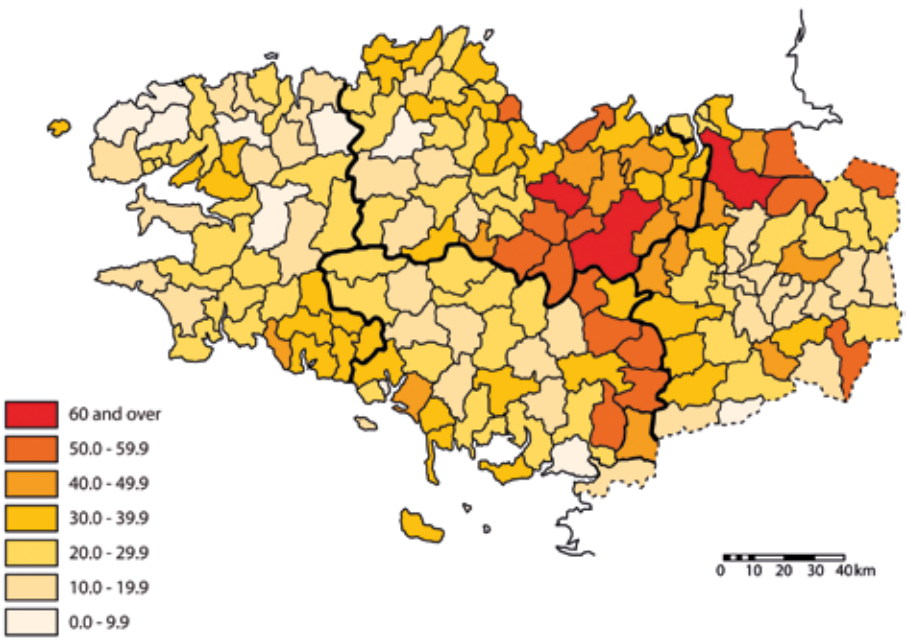
Nick Day, who is now retired and lives in Guernsey, remembers the statistical work: "The two monographs on statistical methodology played a central role for me, particularly the first one, on case–control studies. Such studies were the mainstay of cancer epidemiology at the time, and there were numerous papers, and even books, on their theory, design, and analysis. It was all a bit of a mess, with little coherence. Norman Breslow and I saw that basic statistical developments in the 1970s could provide a coherent underlying structure for case–control studies, the theoretical basis if you like, which led directly to methods that could be generally applied. We then went ahead and wrote the case–control volume. It quickly gained wide acceptance, and was translated, in whole or in part, into a number of languages. A few years ago, a review appeared in the *American Journal of Epidemiology* identifying the most widely quoted publications in the journal in the previous 25 years. Our book was at the top of the list, by a long way."

Norman Breslow, who now divides his time between Seattle and Provence (France), adds, "The case–control volume turned out to be a very successful effort, way beyond our wildest dreams in terms of how it was accepted as a textbook within the community of epidemiologists and biostatisticians. It had not been intended as a textbook – it was a research monograph to reveal the latest developments in biostatistics as related to epidemiology, and relied very heavily on research work that Nick Day and I were both involved in in Lyon at the time. I think it had an impact because of the level it was oriented towards: not a theoretical statistical text nor an epidemiology text spending a lot of time on a line of study and the development of questionnaires, collection of data, and that sort of thing. It was addressed towards mathematically qualified epidemiologists and also statisticians. For example, in about 1980 I started in Seattle a course directed towards second-year epidemiology students, PhD students, and first-year master's students in biostatistics, and we used the case–control monograph as a textbook. It was very successful, and I imagine that similar courses happened at many other universities, but epidemiologists, too, liked the book because it used real data, real examples – it did not dwell on mathematics aspects but tried to emphasize what was useful in answering questions of interest to them."

Norman Breslow enjoys returning to IARC (here attending a recent seminar), where he worked for four years during the 1970s, closely collaborating with Nick Day, in particular on the production of their book on statistical methods for cancer research. At the School of Public Health of the University of Washington, Seattle, USA, Breslow founded and developed a first-class biostatistics department, where he is currently professor emeritus.

> *In 1972 WHO had installed an IBM 360, which at the time was a very powerful computer. We wrote programs in FORTRAN and prepared a set of Hollerith cards to instruct the computer how to analyse the data, followed by 600 or 1000 data cards. I would rise at 5:00 am, walk up to the Gare des Brotteaux in Lyon, get on the overnight train going from Barcelona to Copenhagen, with a stop in Geneva, where I would have breakfast in Gare Cornavin, take a bus up to the WHO headquarters, work all day feeding my cards into the machine, and then in the evening I would reboard the train to Lyon.*
>
> – Norman Breslow, former IARC scientist

level that respected theoretical and formal rigour while being mostly accessible to readers with a limited mathematical background. Fourth, and most importantly, the statistical analyses were illustrated step-by-step by applying them to real data from epidemiological studies. This was uncommon in the methodology books then available, and even in those published later. A strong connection was maintained between the methods and the substance of epidemiological investigations, such as studies of the relationship between alcohol consumption and oesophageal cancer, hormones and endometrial cancer, or ionizing radiation and lung cancer.



Mortality rates for oesophageal cancer in the Brittany region (deaths per 100 000 population per year during 1958–1966), by canton. Rates in Brittany were markedly higher than the average rates in France. Within the region, major variations occurred among cantons. A relationship with different levels of alcohol consumption was suspected, and epidemiological studies were initiated to test this hypothesis.

ANALYSING ANIMAL CARCINOGENICITY EXPERIMENTS

Typically, animal carcinogenicity experiments involve exposing groups of animals (e.g. rats) to a substance (e.g. a chemical whose molecular structure makes it a suspected carcinogen) and following them up and recording the numbers of observable tumours in various organs. If the substance is indeed carcinogenic, more tumours will be observed among the exposed animals than among a comparable group of unexposed animals (controls).

However, the tumours may become observable in very different ways: because they affect the skin or visible mucosa, because they affect an internal organ and they cause the animal's death, or because the apparently healthy animals are sacrificed at a fixed time. A naive way of counting the tumours, not at all uncommon in the literature, is to simply tally all the tumours, irrespective of the way in which they were observed. However, this may lead to grossly wrong conclusions in the comparison between exposed and unexposed animals. It is intuitive that a tumour found at autopsy when an apparently healthy animal is sacrificed has a different relevance than a tumour that has already caused the animal's death. The differently observed tumours need to be counted and compared separately. This requires sophisticated statistical analysis capable of combining the results of the separate comparisons between exposed and unexposed animals.

These methods have an interesting relationship to what we can or cannot observe in human populations. When we focus on a single cause of death (e.g. lung cancer), what we can actually observe are those deaths from lung cancer that have not been precluded by earlier deaths from all other causes. In animal experiments, the sacrifice of animals at fixed times has a similar role to all other causes of death in human populations. As in animal experiments, after earlier deaths from other causes, some lung cancers can be found at autopsy. In animal experiments, the fixed times of sacrifice are the main determinant of the average age at death of the animals. Similarly, it is mainly the ensemble of all other causes of death – rather than the single cause of death of interest (lung cancer) – that determines the average age at death of people dying of lung cancer. This is contrary to what might seem to be the case at first glance.

A 2014 survey of books on epidemiological and statistical methods in the biomedical literature showed that Breslow and Day's book currently receives 100–200 citations per year in research contexts, as a reference for now well-established methods or for teaching purposes (see "Case–control studies"). The Breslow–Day test, first introduced in the book, is often found in research papers as a statistical test of whether risk (e.g. of lung cancer in smokers compared with non-smokers) is the same in different subgroups (e.g. men and women).

In Breslow and Day's book, cancers were considered as occurring in a cohort of people specially assembled to study the causes of cancer. One can also consider cancers, or deaths from cancers, occurring in populations within defined geographical areas, or cancer-related deaths or recurrences of cancer occurring in groups of patients. Statistical methods to deal with these two situations were presented in 1994 in an IARC book co-authored by Jacques Estève, *Statistical Methods in Cancer Research: Descriptive Epidemiology*. It details methods for analysing data as typically gathered by cancer registries, including examining how cancer occurrence evolves over time, as well as geographical variations in cancer frequency and their

correlations with geographical variations in factors such as income, air pollution, alcohol consumption, and diet. The book also covers a key topic for evaluating the effectiveness of cancer treatments: the analysis of cancer patients' survival.

Survival analysis is also at the core of animal experiments to test whether a substance is carcinogenic. Supplement 2 of the IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, published in 1980, has a 100-page annex on statistical methods, produced collaboratively by IARC and external statisticians. This annex highlights with great clarity how findings from carcinogenicity experiments in animals should be correctly analysed (see "Analysing animal carcinogenicity experiments"). A substantial expansion, with a focus on formal statistical models of analysis, followed in 1987 with the IARC book *Statistical Methods in Cancer Research: The Design and Analysis of Long-Term Animal Experiments*, of which Jürgen Wahrendorf was the senior co-author. The issues developed in these publications are a further example of the value of IARC's contributions to the analysis of cancer data and are relevant not only for cancer research but also for long-term toxicological experiments in general.

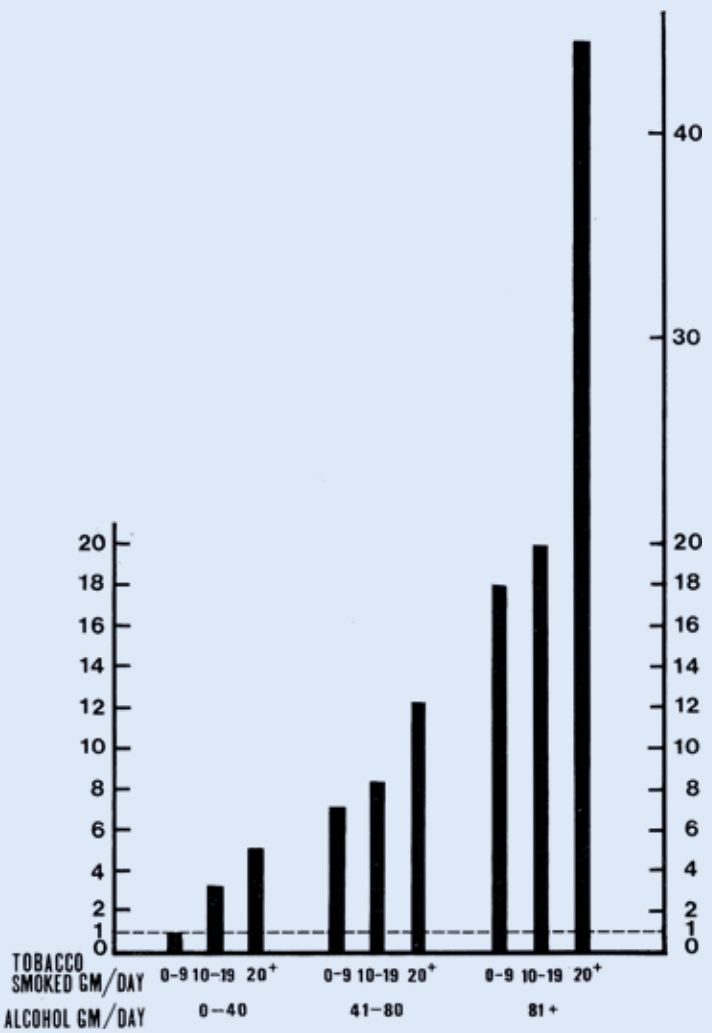
CASE–CONTROL STUDIES

The observation of a high frequency of oesophageal cancer in the Brittany and Normandy regions of France raised a question: Could this be due to the notoriously high consumption of alcoholic beverages (some of high proof, typical of these regions), or to tobacco smoking (as suggested by previous studies in other parts of the world), or to both? A first, relatively quick answer to this question would come from finding out whether people with oesophageal cancer in fact drank more alcoholic beverages or smoked more than healthy people of the same age and sex.

A study to address this question was organized in hospitals of Ille-et-Vilaine, a department of Brittany, between 1972 and 1975. It included 200 patients affected by oesophageal cancer ("cases") and 778 unaffected people ("controls") from the communes of this region. In this typical case–control study, the usual consumption of wine, beer, and spirits was assessed by interviewing the study participants using a standardized questionnaire. The quantity of the different beverages was converted to grams of alcohol per day. The results were clear: the majority (85.5%) of the patients with oesophageal cancer had consumed 40 grams or more of alcohol per day, whereas only about half (50.2%) of the controls had.

It seems reasonable to infer that a person who currently drinks more alcohol than somebody else is at a higher risk of developing oesophageal cancer in the future. But how can we obtain the actual measure of interest, the risk of future cancer for different levels of alcohol consumption, when we only have data on cases (plus controls) that have already occurred? Do we need to carry out a cohort study assessing alcohol consumption in a large group of people and following them up for 20 or 30 years to see how many cases of cancer occur in the different categories of consumption? Indeed, this is the direct measurement needed to conclude that alcohol causes oesophageal cancer. However, it turns out that there is no need to wait many decades; essentially the same risk measure can be derived for a given population from either a case–control study or a cohort study. A formal justification of this seemingly magical solution had been put forward in the early 1950s. The book *Statistical Methods in Cancer Research* placed it within a coherent logical and probabilistic framework, showing that case–control studies (treated in Volume 1 of the book) are basically equivalent to cohort studies (treated in Volume 2). In practice, this equivalence holds provided that some conditions are satisfied, particularly concerning the way in which cases and controls are selected.

This fundamental development in methodology, presented in Breslow and Day's book, unifies the different study designs. This made it possible to use the same, rather than disparate, statistical methods of data analysis to deal with issues arising from these studies. For example, a study in which subjects with different levels of alcohol consumption are compared would need to control for extraneous factors (e.g. tobacco smoking) that might produce a false effect of alcohol or influence its strength of effect. In fact, as the graph shows, people who have a high consumption of alcoholic beverages (equivalent to 81 grams or more of alcohol per day) and also smoke 20 or more cigarettes per day – represented by the column at the extreme right of the graph – have a risk of oesophageal cancer more than 40 times that of people who consume 0–40 grams of alcohol per day and smoke 9 or fewer cigarettes per day (the extreme left column of the graph).



Relative risk of oesophageal cancer in Ille-et-Vilaine, a department of Brittany, by categories of daily tobacco smoking and alcohol consumption.

A NOVEL EPIDEMIOLOGICAL STUDY DESIGN

Reliable data analysis does not depend only on statistical methods; these are highly dependent on the way in which the data have been collected, and hence the design of a study matters as much as the analytical approach. The Gambia Hepatitis Intervention Study ranks among IARC's key projects. Its substantive relevance for establishing the etiology of liver cancer and testing the preventive effectiveness of the vaccine against hepatitis B is outlined in the chapter "Viruses and vaccines". Equally important from a methodological viewpoint was its novel study design.

The Gambia Hepatitis Intervention Study originated in the mid-1980s under particular circumstances. A vaccine was available that was known to be effective against hepatitis B. The research question was whether preventing hepatitis B infection (i.e. preventing newborns from becoming carriers of the hepatitis B virus) would prevent the later occurrence of primary liver cancer. Initially, it seemed that the only ethically admissible way to answer this question would be to start administering the vaccine to all newborns in a given year and then compare (several decades later) the liver cancer occurrence in vaccinated people with that in the unvaccinated people who were born before the vaccination programme started. This is known as a "pre–post" comparison. Such an approach is fraught with potential biases because many other factors, which vary over time and have nothing to do with the vaccine, could induce a change in cancer occurrence and detection.

> *What impressed me enormously was the Gambia hepatitis B programme and the absolute commitment of the Director at the time to ensuring that IARC could continue vaccination after the numbers needed in the vaccinated and unvaccinated groups were complete. Most researchers would say, "We will now continue with our research programme and we're very sorry but you'll have to go and look elsewhere for the money to continue the vaccinations", but that was not the approach taken by IARC. I thought that was really an amazingly good thing to do.*
>
> – Bruce Armstrong, former IARC Deputy Director

The study design was considered ethically uncontroversial, as confirmed by the IARC Ethics Committee, which had recently been established (see "The IARC Ethics Committee"). But the design was scientifically weak, a serious handicap considering the considerable investment of resources that would be demanded by the project over a projected 40-year period. However, one major practical constraint to vaccine delivery soon emerged that was turned into a scientifically strong study design. In fact, it would have been logistically impossible to start administering the vaccine to all newborns in The Gambia in a given year – more than 60 000 newborns, scattered across rural areas. The only feasible procedure was to introduce the vaccine gradually over several years.

THE IARC ETHICS COMMITTEE

The IARC Ethics Committee was established in 1982. It has two specific tasks: first, to verify that research projects in which IARC participates and that involve human subjects have received clearance by the relevant ethical committees at the country level, and second, to assess whether it is ethically appropriate for the Agency, in the light of its research mission and public health role within the framework of the World Health Organization, to participate in such projects. Projects must receive clearance from the IARC Ethics Committee before they begin. Any subsequent modifications of the study protocol must also be submitted to the committee for ethical clearance. Currently, the committee is composed of researchers and laypeople. A majority of its members, including the chair and vice-chair, are from outside the Agency.

Members of the IARC Ethics Committee at a recent meeting.

The crucial methodological innovation was to choose at random – rather than by convenience or in a systematic way – the newborns to be vaccinated each year. Actually, clusters of newborns (i.e. local vaccination teams), rather than individual newborns, were chosen at random. During the first year of the programme (1986), about 25% of all newborns, coming from areas covered by four vaccination teams chosen at random from a total of 17 countrywide, were vaccinated (to be compared with the 75% who were unvaccinated). During the second year, 50% were vaccinated. During the third year, 75% were vaccinated, and finally during the fourth year, all newborns were vaccinated. This design made possible an unbiased comparison between the randomly chosen vaccinated and unvaccinated subjects within each of the first three years of the programme. The random choice of newborns to be vaccinated was ethically unobjectionable since it was non-discriminatory and impartial.

This design, first implemented in the Gambia Hepatitis Intervention Study, is both scientifically and ethically sound and has entered standard methodology as the "stepped-wedge" trial design. The principle is that an intervention is assigned sequentially to the trial participants, either as individuals or as clusters of individuals, over several time periods. Which individuals or clusters receive the intervention in each time slot is determined at random, and by the end of the random allocation all individuals or groups will receive the intervention. This type of design has been used, and continues to be used, in a variety of studies within and outside the field of cancer research, particularly in the evaluation of the effects of vaccinations, screening, and health education programmes.

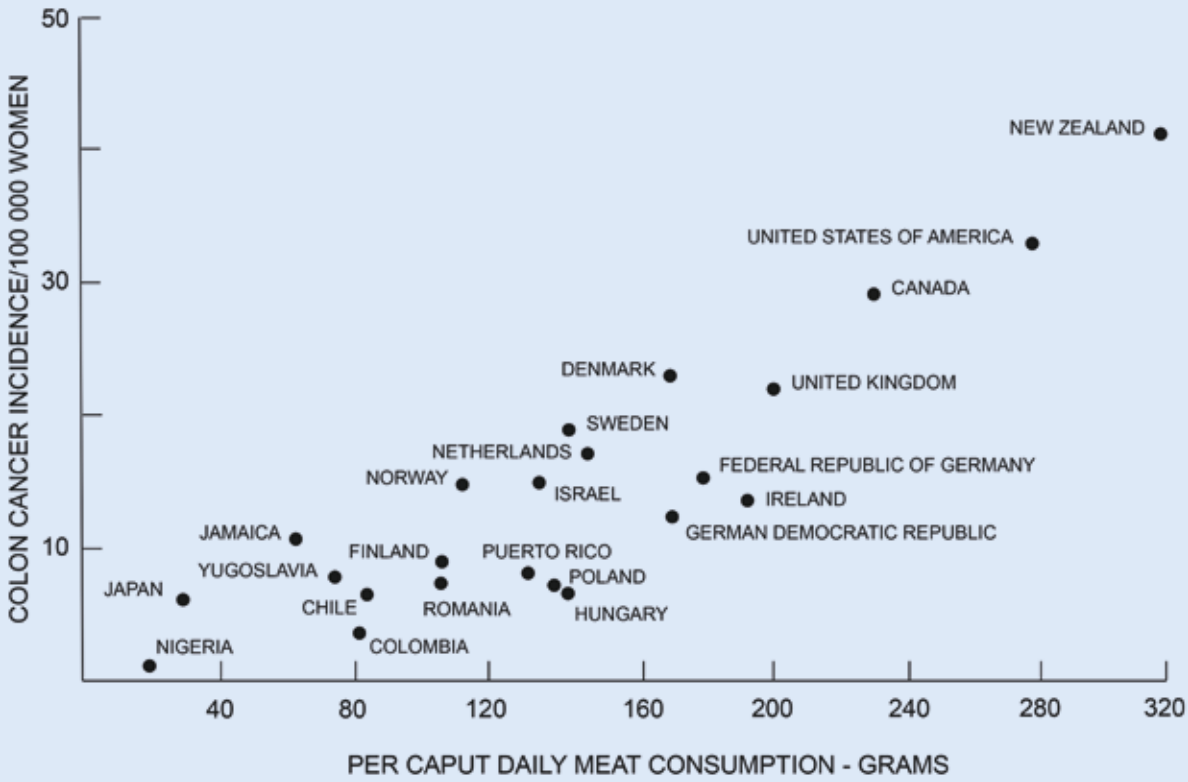
ANALYSING MULTICENTRE EPIDEMIOLOGICAL STUDIES – A KEY IARC ACTIVITY

Breslow and Day's synthesis consolidated a methodological basis that could be used as a standard starting point for a great variety of specific developments. At IARC, research in statistical methods has become more specialized over the ensuing decades and is now embedded within the different types of epidemiological studies. However, some areas of work maintain a more general perspective. One example is a recent paper from IARC, "Penalized loss functions for Bayesian model comparison". Although the title sounds highly esoteric, in fact this research addresses the very general and fundamental issue of how to choose the best model in the analysis of any data set (e.g. how best to formalize the mathematical relationship between intake of various foods and the occurrence of colon cancer).

A second topic of broad relevance is the analysis of data from multicentre epidemiological studies. Conducting investigations in multiple populations was inherent to the scientific rationale for the establishment of IARC, particularly because in the mid-1960s this type of study was not common in cancer research. For example, multiple populations may be those in different geographical areas, chosen because they may have widely different lifestyle habits. Or multiple populations of workers exposed to the same potential cancer hazard (e.g. a chemical) at various factories may be chosen to achieve a total population large enough to attain a high sensitivity for detecting an increase in risk if it exists. Another advantage of multicentre studies is the possibility of verifying whether the results obtained within the different populations are consistent with each other. For example, finding the same inverse relationship between the intake of vegetable fibre and the frequency of colon cancer in different populations would be strong evidence in favour of a causal preventive role of vegetable fibre intake. In science, replicability, or at least consistency, of results – as is feasible in multicentre studies – is the most stringent criterion for judging causality. The methods for assessing consistency, although simple in principle, are fraught with complexities in practice (see "Combining epidemiological results from multiple populations"). Optimizing these methods is a continuous area of research in biostatistics at IARC.

COMBINING EPIDEMIOLOGICAL RESULTS FROM MULTIPLE POPULATIONS

Two basic types of comparison are possible when investigating factors that possibly cause cancer. The first is a comparison between individuals, measuring for each person the level of exposure to the factor of interest (e.g. daily meat consumption). The second is a comparison between large groups of people, typically the populations of a region or a country, estimating for each population the average level of exposure (e.g. average meat consumption). This second type of comparison has often been used in studies of the role of diet in cancer because it exploits the large variations in intake of different foods and nutrients that spontaneously occur between populations with vastly different cultural and dietary habits.



A clear relationship is seen between the average daily meat consumption (in grams per day) by women in a country and the incidence rate of colon cancer in women in the same country (in 1975): the higher the consumption, the higher the rate.

As shown in the graph, when the frequency (incidence rate) of colon cancer is plotted against the average daily meat consumption for women in different countries, a striking relationship emerges of increasing incidence rates with increasing dietary meat consumption. Does this speak in favour of meat consumption as a cause of colon cancer? It does, but there are a multitude of differences between the populations of women in different countries, other than just meat consumption. Therefore, one cannot be sure that the relationship seen in the graph is not produced by one or more of these other factors, known or unknown.

The relationship seen in the graph, termed “ecological” because it involves whole populations in different environments (countries), can only be regarded as suggestive. It needs to be confirmed by studies carried out at the level of individuals, termed “analytical” studies, which offer much better possibilities of measuring consumption for each person and ruling out factors other than meat consumption. Multicentre international studies – the type of epidemiological investigation in which IARC has developed unique expertise – combine the advantages of both approaches, ecological and analytical, and make it possible to check the consistency of the results obtained at the two levels.

If consistency holds, a picture closely similar to the one in the graph would be obtained within every country (or, more generally, for every study centre). The points on the graph would represent not countries but a group of women who are comparable in all characteristics but with different meat consumptions, measured individually. (Data for men would need to be looked at in the same way.) However, assessing whether a “close similarity” exists requires sophisticated statistical methods, incorporating the treatment of possible – and, in reality, inevitable – errors in measuring exposures like meat consumption. If the results pass the tests of consistency and show the same relationship between colon cancer occurrence and meat consumption within each country and for all countries combined, then the conclusion that meat consumption is a cause of colon cancer would be strongly supported. Methods that combine in a single analysis “within-centre” results and “between-centre” results are acquiring wide relevance in epidemiology, as reflected in the concluding sentence of an IARC paper developing this approach: “The use of multilevel models, which constitute a very powerful approach to estimating individual vs aggregate levels of evidence, should be considered in multicentre studies.”