The contribution of epidemiology

P. van den Brandt, L. Voorrips, I. Hertz-Picciotto, D. Shuker, H. Boeijing, G. Speijers, C. Guittard, J. Kleiner, M. Knowles, A. Wolk, A. Goldbohm

Summary

Epidemiologic studies directly contribute data on risk (or benefit) in humans as the investigated species, and in the full food intake range normally encountered by humans. This paper starts with introducing the epidemiologic approach, followed by a discussion of perceived differences between toxicological and epidemiologic risk assessment. Areas of contribution of epidemiology to the risk assessment process are identified, and ideas for tailoring epidemiologic studies to the risk assessment procedures are suggested, dealing with data collection, analyses and reporting of both existing and new epidemiologic studies. The dietary habits and subsequent disease occurrence of over three million people are currently under observation worldwide in cohort studies, offering great potential for use in risk assessment. The use of biomarkers and data on genetic susceptibility are discussed. The paper describes a scheme to classify epidemiologic studies for use in risk assessment, and deals with combining evidence from multiple studies. Using a matrix approach, the potential contribution to each of the steps in the risk assessment process is evaluated for categories of food substances. The contribution to risk assessment of specific food substances depends on the quality of the exposure information. Strengths and weaknesses are summarized. It is concluded that epidemiology can contribute significantly to hazard identification, hazard characterisation and exposure assessment. © 2002 ILSI. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Epidemiology; Risk assessment; Exposure assessment; Hazard characterization; Food chemicals; Food composition tables; Biomarkers

Contents

1. Introduction ...........................................................................................................................................................389
  1.1. The situation specific for dietary risk assessment ..........................................................................................389
  1.2. Outline of the chapter ..................................................................................................................................390

2. Introduction to epidemiology ..........................................................................................................................390
  2.1. Epidemiologic study designs .........................................................................................................................390

Abbreviations: ATBC = Alpha-Tocopherol, Beta-Carotene Cancer Prevention Trial; BHA, butylated hydroxyanisole; BHT, butylated hydroxytoluene; CARET, Beta-Carotene and Retinol Efficacy Trial; CC, case-control studies; CDC, US Centers for Disease Control; DALY, disability adjusted life expectancy; EAN, European article number; ELR, excess lifetime risk; EPIC, European Prospective Investigation into Cancer and Nutrition; FDA, US Food and Drug Administration; GEP, good epidemiology practice; GSTs, glutathione S-transferases; HBV, hepatitis B virus; HBsAg, hepatitis B surface antigen status; HLE, healthy life expectancy; ISPE, The International Society for Pharmaco-Epidemiology; JECFA, Joint FAO/WHO Expert Committee on Food Additives; JPHC, Japan Public Health Center-Based Prospective Study; MRFIT, Multiple Risk Factor Intervention Trial Group; NAT 2, N-acetyltransferase 2; NCI, US National Cancer Institute; PCB, polychlorinated biphenyl; PLM, post-launch monitoring; RCT, randomised controlled trial; RR, relative risk; UCL, upper confidence limit.

* Corresponding author. Tel.: +32-2-771-00-14; fax: +32-2-762-00-44.
E-mail address: jkleiner@ilsieurope.be (J. Kleiner).
1. Introduction

For the process of risk assessment of ingredients deliberately added to food, well-controlled experiments in animal models are legally mandatory. Although useful data of human origin are sometimes available (particularly for food safety evaluations), they are generally not included in quantitative risk assessments. A more extensive use of human data has been proposed by many investigators since real-life exposures reflect the actual situation. In contrast, animal studies have limited capability due to the requirement of both interspecies and high-to-low dose extrapolations (Shore et al., 1992; Hertz-Picciotto, 1995; Shore, 1995; Samet et al., 1998). For interspecies extrapolation, if no specific information is available, similar rates of absorption, metabolic pathways, rates of activation or detoxification, and rates of elimination are assumed in both animals and humans (Hertz-Picciotto, 1995; see also Dybing et al., 2002). Other limitations for comparability are interspecies differences in breathing rates, organ sizes, basal metabolism, rates of cell turnover and life spans. Interspecies extrapolation also requires extrapolation of equivalent exposures, however uncertainty exists due to differences in the required units to be used (mg per kg body weight, mg per surface area, cumulative lifetime mg per kg body weight) (Hertz-Picciotto, 1995). In high-to-low dose extrapolation, the risk assessed at high dose is used to estimate the risk at lower (often more realistic) levels of exposure (often by application of a mathematical function), although actual measurements at this level generally have not been made.

Human data on both adverse and beneficial health effects of substances in food have less uncertainty and are primarily derived from epidemiologic studies having an experimental design (such as clinical trials or intervention studies) or an observational design (e.g. case-control studies and cohort studies). Experimental designs of studies in human have the advantage that they can better control for confounding and have theoretically, therefore, more evidential strength. However, due to ethical, financial and practical reasons, it is often not possible and often not even necessary to conduct an experimental study on adverse effects in humans. In particular, if a specific health outcome has a long induction period (as in the case of cancer) or if the chemical to be investigated has a postulated serious adverse effect on health, intervention studies may be unfeasible. Observational epidemiologic studies, therefore, can contribute substantially to evidence on the risk of human exposure to dietary chemicals. In addition, experimental human studies require active participation which usually results in highly selective groups in which a measure can be tested. Therefore, data tailored to the general population might only be found in observational epidemiologic studies.

Despite these shortcomings of controlled animal studies, and even in the presence of human data, incorporation of human studies (both experimental and observational) in the risk assessment process has rarely been put into practice in the area of food-associated risks. This seems to be due to a number of reasons, including:

- No explicit role for epidemiologic data in risk assessment required for regulatory purposes, not even for substances to which humans have been exposed long-term (e.g. setting ADIs and maximal permitted levels in food products).
- Misunderstanding across different disciplines (i.e. toxicology vs epidemiology) of concepts and methods used in these disciplines.
- Lack of suitable data, either because potentially useful data are not analysed or published or because such data are not reported in the manner that is required for risk assessment.
- Epidemiology cannot contribute to screening of new substances, which have no history of use by humans.

With the exception of the last argument, it has been suggested that these reasons are not basically prohibitive and that more can be done to clear the barriers to include epidemiologic data explicitly in the risk assessment process.

1.1. The situation specific for dietary risk assessment

Risk assessment of orally ingested compounds needs a specific approach per se, since eating is unavoidable. Unlike xenobiotic agents which, if found to be toxic, can at least in principle be dispensed with; food is
indispensable and many individual nutrients are essential for basic physiologic functions (Saracci, 1993). Furthermore, in contrast to xenobiotics, for which the lowest possible exposure is best for health, nutrient intake has an optimal range, and levels above and below which harmful health effects may occur (Saracci, 1993). The definition of the optimal range is a continuous scientific process using animal and human data and performed by institutional bodies that formulate dietary recommendations. It follows a risk/benefit approach and is based more on physiological and epidemiologic than toxicological principles. For the evaluation of safety of entirely new substances or known substances used in much higher concentrations such as in novel foods, fortified foods and food supplements, a toxicological screening is required as an important first step in the risk assessment process, since it is considered unethical to expose humans deliberately to substances with unknown health effects. As a second step (pre-launch), intervention studies/clinical trials in humans can generate short- and medium-term data on safety as well as efficacy (if desired) of the substance. After the market launch of the novel substance, observational epidemiology is very well suited to contribute to the third step; that is, addressing the issue of long-term safety. Post-launch monitoring is a type of observational study, specifically designed to monitor the long-term safety of a novel substance.

1.2. Outline of the chapter

This chapter addresses the potential of human studies to contribute to risk assessment. It focuses on observational epidemiology in particular, as the perceived gap between epidemiology, mostly based on observational studies, and experimentally-based toxicology needs to be bridged. With the exception of randomised controlled trials, which are considered as epidemiologic studies, human volunteer studies are outside the scope of this review. This certainly does not imply that human volunteer studies are not valuable for hazard assessment of chemicals in food. The roles of such studies (more extensively discussed in Appendix B) vary from obtaining information on absorption, metabolism and kinetics of a substance and on tolerance, acceptability and palatability of a product, to development and validation of a biomarker.

The chapter introduces the epidemiologic approach, identifies areas of contribution of epidemiology to the risk assessment process, suggests ideas for tailoring the presentation of epidemiologic study results to the risk assessment procedures, and summarises the current state-of-the-art regarding combining and evaluating epidemiologic studies.

In section 2, a concise description is given of the different experimental and observational epidemiologic study designs and issues of confounding. Section 3 is a discussion on the main differences between toxicological and epidemiologic risk assessment. Some examples of the use of epidemiologic data in the risk assessment process are presented in section 4. Section 5 considers ways to improve data collection, analysis and reporting of both existing and new epidemiologic studies with the aim of expanding their potential use in the risk assessment, while section 6 focuses on the use of evidence from epidemiologic studies in risk assessment. Using a “matrix approach”, the potential contribution to each of the steps of the risk assessment process is evaluated for each of the categories of food substances (section 7), and conclusions and recommendations are formulated in section 8.

Many of the examples used throughout this review are derived from cancer epidemiology for several reasons. Cancer epidemiology has provided good examples on how epidemiologic data can be applied to risk assessment (Moolgavkar et al., 1999a), and is a paradigm of a chronic disease with latency. Also, much methodology has been developed in cancer epidemiology. The use of such examples certainly does not imply that epidemiologic risk assessment is not applicable to areas other than cancer.

2. Introduction to epidemiology

Epidemiology evaluates the association between environmental exposures or other characteristics of study subjects (e.g. demographic variables, genetic susceptibility) with disease risk or other health outcomes (Szklo and Nieto, 2000) in attempting to explain observed differences. Associations between the characteristics of a population and its disease risk are subsequently investigated to determine whether these factors directly alter the disease risk or act as a marker for another underlying factor directly influencing the disease risk. Epidemiology is the only scientific discipline that directly addresses phenomena of disease occurrence in the human population with the aim of explaining and clarifying them as well as advising public health agencies regarding preventive measures. Analytical epidemiologic studies generate human data suitable for use in the hazard identification and hazard characterisation steps of the risk assessment procedure.

2.1. Epidemiologic study designs

Epidemiologic studies utilise experimental as well as observational study designs. For observational studies, further segmentation is made based on the unit of observation: in ecologic studies, populations or groups of individuals form the basis, whereas individuals are the unit of interest in cohort studies and case-control studies. Whereas in animal experiments efforts are made to generate genetically and environmentally homo-
geneous situations except for the factor under study, such situations are not usually found in human populations. Human populations are heterogeneous in behavior and genetic susceptibility. Observational studies consider this real-life situation and assess it with respect to disease risk. It is obvious that sorting out the major components of risk for specific diseases in human populations requires sophisticated approaches. Observational studies, in which persons have not been randomly assigned to exposed vs unexposed groups, may be affected by bias distorting the factor — disease association. Bias can also occur apart from genetic differences in susceptibility because study participants also differ in many other aspects relating to exposure. In experimental studies, random allocation of subjects to different treatments will minimise the effect of this variation. In observational studies, random allocation to exposure is not possible. Instead, epidemiologists utilise many methods, as in the design and in the statistical analysis of an observational study in order to minimise most sources of bias, which will be discussed in this section. High quality studies are usually those in which such methods are appropriately applied. From the above considerations, the basic epidemiological designs can be briefly summarised to include: (a) experimental studies, such as randomised controlled trials (RCT) (clinical trials, intervention studies) and (b) observational studies, such as cohort, case-control and correlation studies.

A summary on different forms of bias including confounding can be found in Appendix C.

2.1.1. Experimental studies as a paradigm: randomised controlled trials

In some circumstances, experimental studies in humans can be performed analogous to animal studies. Experiments can be carried out in a hospital setting, with patients (clinical trials) or with healthy individuals in the population to study, for example, health effects of dietary changes (intervention studies). The experiment generally is the most clear-cut way to study a dietary cause–effect relationship. This study type, generally designed as a randomised controlled trial (RCT), allows conclusions regarding whether a measure affects the disease rate directly or should be considered as a marker of another underlying cause affecting disease risk.

Experiments are said to be “controlled” when an untreated group or a group treated according to a standard method is involved, with the aim of controlling changes in disease incidence occurring independently of changes in exposure. As in animal studies, human subjects in randomised experiments are allocated to the treatment or control group at random, minimising the chance of grossly unequal distribution of potential confounding variables among the study groups. Ideally, the experiment is carried out in a double-blind fashion, where both the subject and the observer of the study outcome are blinded to the exposure status, thereby avoiding subjective influences on the results of the study (i.e. respondent and observer bias).

For ethical reasons, human experiments are generally carried out to study positive health effects. Therefore, human experimental studies can only be carried out once sufficient observational data have shown that the benefit is reasonably probable and/or toxicological screening has given reasonable proof of safety (Willett, 1998). Most human randomised controlled trials are conducted to establish efficacy and safety of a single agent in a specific situation. Owing to the limited size of most trials, less common adverse effects of exposures cannot be detected in experimental studies. When long community-intervention trials are not feasible, investigations will need to rely on large-scale observational studies. Also, because time of follow-up of the study population in experimental studies is generally limited, adverse effects occurring many years later will not be identified. For dietary intervention studies aimed at health outcomes such as cancer, a major limitation is the long follow-up time needed, thus leading to problems with compliance and drop-out, which may cause selection bias. Also, the danger exists that an unexposed control group adapts to a diet similar to the one used for the exposed study group if the treatment diet is thought to be beneficial, as occurred in the Multiple Risk Factor Intervention Trial of coronary disease prevention (MRFIT) (Multiple Risk Factor Intervention Trial Research Group, 1982). Intervention studies are best used to study beneficial outcomes of minor components of the diet such as trace elements or vitamins since these can be formulated into pills or capsules. Because of the enormous costs of large-scale intervention studies, exposure is usually limited to a few or even one exposure level, thus limiting information regarding the exposure–response relationship. Additionally, exposure levels generally will be chosen that will give the highest probability of “success”. Subsequently, studies to determine the minimal effective dose are rarely conducted. Examples of well-known large-scale dietary intervention studies are the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Trial (ATBC) and Beta-Carotene and Retinol Efficacy Trial (CARET) trials. In the ATBC study, subjects were provided pills containing vitamin E and/or beta-carotene or a placebo. In the CARET study, subjects were given a combination of beta-carotene and vitamin A (retinol) supplements. Subjects were followed up for 4–6 years for cancer incidence and mortality. More details of the ATBC and CARET study are illustrated in section 4.

2.1.2. Observational studies

As shown earlier, controlled human experiments such as intervention trials have many limitations, mainly imposed by ethics and costs. However, individuals often
willingly or unwillingly expose themselves to many potentially harmful or beneficial factors. Observational studies, which do not involve by definition intervention and randomisation, therefore have a major role in providing evidence on the association between dietary exposure and health effects. Two different types of observational studies can be distinguished: cohort or follow-up studies and case-control studies.

2.1.2.1. Cohort studies. Cohort studies start with a population initially free of the health outcome under study (Fig. 1). Exposure is assessed at the start of the study (and is sometimes assessed several times during follow-up) and may include (besides questionnaire/interview data), the collection of biological materials. During follow-up, some subjects develop the disease (or other health outcome) and others do not. The risk of developing the disease (defined as cumulative incidence) is related to the exposure, and is generally expressed in terms of relative risk or relative rate, comparing risks (rates) among exposed and unexposed individuals. Sometimes it is also expressed in terms of added or excess risk or rate or risk differences. A main advantage of the cohort design is that at entry, subjects are free of the disease in question. Therefore, the disease will not influence exposure (information) of the subjects. To be more certain of this, in cohort studies, cases diagnosed in the first few years of follow-up can be excluded to avoid possible information bias due to preclinical disease. Both in cohort studies and in case-control studies, confounders may not be equally distributed between exposure groups. However, using appropriate statistical analysis methods these factors generally can be controlled for if they are adequately measured. In other words, it is a misconception that observational epidemiologic studies must necessarily suffer from confounding bias. A general limitation of cohort studies is that large populations are required, especially if rare health outcomes are studied. Substantial follow-up time is required for diseases of long latency periods. Both large sizes and long follow-up time make cohort studies relatively expensive.

Existing databases from which information about exposure can be derived (such as those kept by health insurance companies, pharmacies, and purchase panels used in marketing studies) and registries containing health outcomes (such as cancer, mortality or birth defect registries) can sometimes be used to link exposure to health outcome in a very economical and efficient manner. Additionally, in the recruitment phase of a study, existing databases, for example from municipalities, can contribute to reduce costs.

2.1.2.2. Case-control studies. In contrast to cohort studies, case-control studies start with a group of patients, or persons known to have the disease or condition (Fig. 2). These cases are compared with a control group, selected to represent the population from which the cases arose with respect to the exposures of interest. The exposure information is collected concerning a defined time prior to diagnosis of the disease. A case-control study usually compares the odds of past exposure to a suspected risk factor between cases and controls to obtain an odds ratio, which is an estimate of the relative risk. An advantage of case-control studies compared to the cohort design is that no long follow-up period is required to obtain results. A second advantage is that for rare diseases, a much smaller sample size is needed, while the odds ratio is a very close approximation of the relative risk. A main disadvantage of case-control studies is that the disease can influence exposure information provided by the subjects ("recall bias") or recent exposure really is different from what it used to be, because of the disease ("presence-of-disease bias"). For example, the dietary pattern of a case-to-be may be influenced consciously or unconsciously by (subclinical) symptoms of the disease. Another form of bias that can be present in case-control studies is "selection bias", which can occur as a consequence of improper selection...
of study participants. Selection bias is present when the relationship between exposure and disease is different in those subjects actually involved in the study from those who theoretically were eligible to participate (thus, including those who in the end did not participate). Low response rates in cases and/or an inappropriate choice of the control group may give some indication on the possible presence of selection bias (for more detailed information; see Appendix C).

**Prospective vs retrospective studies**

Confusingly, the words *prospective* and *retrospective* often are used to indicate cohort and case-control designs, respectively. In general, prospective studies can be defined as those in which exposure and covariate assessments are made before the cases of illness occur. In a retrospective study these assessments are made after the cases have already occurred. A cohort study can be retrospective (better named: historical) when the cohort is identified and assembled on the basis of information recorded in the past and “followed up” to the present time. Many occupational cohort studies are historical or retrospective, in contrast to cohort studies on diet which are often prospective. A case-control study can be prospective by using exposure data that are collected before the diagnosis of disease.

2.1.2.3. **Nested designs.** Study designs that combine the features and advantages of case-control studies and cohort studies are nested case-control studies and case-cohort studies. These study types combine the reduced chance of bias in exposure assessment occurring in cohort studies with the efficiency from case-control designs.

Collection of exposure data (e.g. questionnaires, biological samples) is performed for the full cohort at the start of the study, thus preventing recall bias or presence-of-disease bias. The processing and (chemical) analysis of the collected material is only done for cases and a sample of the cohort (controls). For nested designs, the case-group consists of all (or a representative sample of) individuals who develop the outcome of interest in the defined cohort over a defined follow-up period. The designs differ in the way controls are selected. As both cases and controls, however, are selected from the same source cohort, selection bias will also be less likely than in case-control studies where the source population is not well defined. Both the nested case-control designs and the case-cohort design are very efficient in providing risk estimates using cases occurring in a full cohort, whereas processing of questionnaires or analyses of, for example, serum samples is needed only for cases and control groups (Figs. 3 and 4). This importantly reduces costs in large cohort studies (Völovsics and van den Brandt, 1997; Szklo and Nieto, 2000). Many recent large cohort studies have an associated “biobank” in which samples of biological material such as blood, urine, toenails, DNA from swabs are being stored. When a specific question arises, for example exposure to polychlorinated biphenyls (PCBs) and the risk of breast cancer, blood samples of cases and controls from the biobank can be analysed for PCB content (Wolff et al., 2000).

2.1.2.4. **Ecologic studies.** In ecologic studies (also called correlation studies or aggregate data studies), the unit of observation is usually a whole population. This can be, for example, a population in a particular geographical area or at a particular time. The general aim is to study geographical trends or time trends. In this type of study, the mean value for both exposure (e.g. population per capita consumption) and outcome such as disease rate in the population) is obtained for each unit of observation (for example a country). A traditional example is a study of per capita meat consumption vs colon cancer incidence in various countries, illustrated in Fig. 5.
The association between exposure and outcome can be expressed as a correlation coefficient, although linear regression or Poisson regression is both possible. The design has the advantage that contrasts in dietary intake between populations can be very large. Also, average diets within a population are more stable than are the diets of individual persons within a country. Finally, the health effects rate on which international studies are based are usually derived from relatively large populations and are, therefore, subject to only small random errors (Willett, 1998). Serious limitations, however, are that other factors associated with the population (such as genetic predisposition, other lifestyle practices, etc.) can confound results. Population per capita consumption may only be weakly related to the diets of those individuals at risk of disease. It is assumed that the available food is eaten by humans, not taking pets into account. Also, differences between populations in quality of measurement of both exposure and disease (technology of diagnosis) may bias the results obtained. Although ecologic studies have unquestionably been useful, they can, however, be completely misleading and therefore not adequate to provide sufficient evidence for causal dose–response relationships between dietary factors and health outcomes (Willett, 1998).

### 2.1.3. Rank ordering of epidemiologic study type in respect to contribution of evidence

A classification of the different epidemiologic study designs with respect to their potential for bias to occur and, consequently, the strength of evidence they provide, and the costs involved is shown in Fig. 6. It indicates that intervention trials provide the strongest evidence for a causal relationship on risk and (due to the ability of the design to control for confounding and bias) have the lowest chance for potential bias to occur. However, they are the most expensive and usually the least feasible studies. The less expensive cohort studies assess exposure and select study participants before the health outcome of interest occurs and thus provide relatively strong evidence. Although the cheaper case–control studies generally assess exposures retrospectively in subjects with and without the health outcome, the resultant evidence is more debatable. This is particularly so in the case of dietary exposures, due to possibility of selection bias, recall bias and/or presence-of-disease bias to occur. The lowest costs are associated with correlation studies but, as mentioned previously, they provide weak evidence and are much more susceptible to bias (van den Brandt, 2000). Some investigators have stated that observational studies cannot, by definition, establish causality of a relationship based on a statistical association. However, if several high-quality studies, such as those in which biases are shown to be minimal are available, and these consistently show a dose–response association, then observational studies may well contribute to conclusions about causality. The power of observational epidemiologic studies was already established 50 years ago when it was discovered through such studies that smoking caused lung cancer.

### 2.2. Issues of confounding

Confounding is present in a study if a specific exposure of interest is associated with another exposure that is related to the disease or health outcome of interest. Observational studies are particularly susceptible to confounding, since the exposure of interest is not allocated and controlled as in an experiment, but just observed in a real life situation. For example, in the Netherlands Cohort Study on Diet and Cancer, it has been shown that a higher consumption of tea is associated with a lower lung cancer incidence (Goldbohm and van den Brandt, 1996). However, smoking is a strong risk factor for lung cancer and appears to be inversely associated with tea drinking: subjects who drink more tea tend to smoke less. Therefore, smoking confounds the association between tea drinking and lung cancer.

In animal experiments confounding can also occur, but it should and can be minimised by a proper study protocol in which circumstances related to housing,
feeding, etc., should be exactly the same for exposed and control groups. Small experiments may nevertheless suffer from substantial confounding simply because random allocation does not always distribute the confounding factors evenly between the experimental groups.

In observational studies, confounding has to be controlled in the data analysis stage of a study. Nevertheless, in the design stage of an observational study, anticipation of potential confounding factors is extremely important, because such factors have to be measured as accurately as possible together with the exposure of prime interest. Such anticipation mainly requires knowledge of other (possible) risk factors for the disease in question. Stratification and multivariate analysis (modelling) are the analytical tools used to control for confounding effects and also to tease out different risk factors for one endpoint.

### 2.2.1. Stratification

In the example of tea consumption and lung cancer in the Netherlands Cohort Study (Goldbohm and van den Brandt, 1996), the crude, unadjusted association is represented in Fig. 7. When the association between tea and lung cancer is studied separately for current smokers, ex-smokers and non-smokers, the association is almost absent within each stratum of smoking status (Fig. 8). This leads to the conclusion that once smoking is taken into account, there is no association between tea drinking and lung cancer incidence. In other words, the crude association found between tea drinking and lung cancer can be “explained” by the confounding effect of smoking. It means that tea drinkers have a lower risk of lung cancer only because there are fewer smokers among tea drinkers than among non-tea drinkers. The effect of a confounder can be controlled for by stratification as in the above example. A weighted average of the stratum-specific estimates can be calculated to obtain a smoking-adjusted relative risk.

Stratification also allows a straightforward and simultaneous examination of the possible presence of both confounding and effect modification. Effect modification is present when the association between exposure and health outcome is not the same between the strata investigated, such as a particular association that may be present only in men, but not in women.

### 2.2.2. Multivariate regression analysis

Stratification is a comprehensible way to adjust for confounding. However, in most observational studies, the investigator has to deal with several confounders simultaneously. Stratification on each level of each confounder leads to very small or even empty cells defined by subgroups of subjects. A solution lies in the use of multivariate regression techniques.

Multivariate analysis refers to a series of analytical techniques, each based on a more or less complex mathematical model, which are used to carry out statistical adjustment; that is, the estimation of a certain measure of association between an exposure and an outcome while “controlling” for one or more possible confounding variables. This can be used in linear regression, logistic regression and survival analysis techniques.

In the example of tea and lung cancer, multivariate regression is the method of choice to adjust not only for smoking status, but also for quantity and duration of smoking.

### 2.3. Ethical and legal issues

Informed consent and restraints on the use of information and biological material concerning individual persons are necessary measures to protect people’s privacy and autonomy. However, legislation and regulations concerning research with healthy persons and patients differ widely between countries, even within Europe despite the harmonisation of some of the legislation. In some instances, regulations are such that an epidemiological study is seriously hampered in the possibilities of investigating health effects of a certain exposure, for example by prohibiting linkage to mortality registries. Regulatory bodies should pay more attention to these issues and develop generally applicable policies to protect subjects against the improper use...
of their data without endangering the conduct of serious epidemiologic research.

2.4. Conclusions

Although experimental studies such as clinical trials and intervention studies among individuals provide the strongest evidence for causal associations, it is often impossible to conduct such studies. High-quality observational epidemiologic studies can also provide sufficiently strong evidence for causal associations and have less practical limitations. The influence of confounding on the results of observational studies can be largely reduced by appropriate design and data analysis.

3. Comparison between toxicological and epidemiologic risk assessment

Differences exist in perspective and approach between risk assessment based on toxicology and risk assessment based on epidemiology. Some of the differences are in definitions, others are in the nature of the investigations (observational vs experimental), and still others result from the research questions being addressed. It is interesting to note that most epidemiologic studies are not conducted with the aim of risk assessment in mind, whereas toxicologic studies often are performed with the sole or primary purpose to be used in risk assessment or at least to inform the regulatory process.

The integration of epidemiology into risk assessment generally, and hazard characterisation specifically, requires clear communication among disciplines. For this reason, Table 1 presents a comparison of the two approaches, including, among other characteristics, differences in the study population, the measurement of exposure, the quantification of effects, the analytic strategies and the sources of uncertainty.

3.1. The nature of the investigations

Outside of randomised controlled trials, which tend to be fairly uncommon in studies of dietary factors and health, most epidemiologic studies are observational. As

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Main differences between toxicological and epidemiologic hazard characterisation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Toxicological data</td>
</tr>
<tr>
<td>Study design</td>
<td>Mainly experimental, animals generally followed simultaneously</td>
</tr>
<tr>
<td>Study population</td>
<td>Selected, genetically homogeneous animals of same age/stage of development</td>
</tr>
<tr>
<td>Exposure</td>
<td>Controlled exposure (dose and duration)</td>
</tr>
<tr>
<td>Definition of risk</td>
<td>Probability of an adverse effect, during the study period, at a given level of exposure</td>
</tr>
<tr>
<td>Quantification of effect for chronic disease</td>
<td>Incidence (proportion of animals with adverse effect over duration of study)</td>
</tr>
<tr>
<td>Comparison of exposed and unexposed</td>
<td>Difference in incidence</td>
</tr>
<tr>
<td>Analysis of dose–response relationship</td>
<td>Various approaches, see Edler et al. (2002)</td>
</tr>
<tr>
<td>Evaluation of evidence and use of studies</td>
<td>Weight of evidence approach applied to different studies; selection of high-quality study with strongest difference or steepest slope</td>
</tr>
<tr>
<td>Sources of uncertainty</td>
<td>Interspecies differences • Interpretation of dose response with respect to threshold or low-dose extrapolation • Human variability • Statistical imprecision</td>
</tr>
</tbody>
</table>

RCTs = randomised controlled trials.

Such as nutritional status, genetics, etc.
Risk of an event is defined as the probability of the event happening in a specific population over a specific time span. When the time span is a standard period, such as a year, these risks can be expressed as incidence rates, and are calculated as the number of cases divided by the number of person-years at risk in the population. In toxicologic studies, the time span is sometimes ignored, if there is not a problem with early mortality, since all animals are initiated into the study at the same time and are followed for the same time. In toxicologic or epidemiologic studies, the time span will depend on the outcome of interest. For acute and relatively short-term effects (e.g., teratogenicity), human and animal studies differ little in the approach to defining and measuring risk. When chronic diseases are of interest, the time span in the experimental assays is usually for most of the lifetime, particularly for carcinogenicity studies. In epidemiology, however, it may not be possible to observe the exposed population for the entire relevant period when interest is in diseases with long induction times. The problem is that it is usually unfeasible to follow people during their whole life span. What is measured, therefore, is not ‘total’ incidence, but rather an age-adjusted incidence rate.

Even in such circumstances, the lifetime risk (which, in toxicologic studies, is referred to as the incidence) can still be calculated from the incidence (or mortality) rate as a function of age. The incidence rate is the most frequently used parameter in epidemiology to indicate the occurrence frequency of a disease or health outcome. It is defined as the number of cases per 100,000 person-years. The incidence rate can be estimated from cohort studies for each level of exposure, if the proper data (number of cases and person-years at risk for each category of exposure and age) are reported. As most epidemiologic studies focus on calculating rate, risk or odds ratios, rather than risk or rate differences, few investigators actually report all basic incidence data. The data required for calculating an incidence rate, however, should be available from the dataset of a cohort study.

A difference between the risk concept as used in toxicology and epidemiology, also relates to exposure. In toxicology, risk is always being used in connection to a specific level (zero or otherwise) of exposure to a specific substance. In epidemiology, risk and rate can also be used to describe the frequency of occurrence of an event in the population without regard for exposure.

3.3. Measures of effect

Risk assessment, whether based on epidemiology or on toxicology, aims to characterise the number of excess cases of disease Y which will occur in a population of size Z due to exposure to agent X at dose level D (Hertz-Picciotto, 1995). Many epidemiology studies, however, are concerned only with determining the risk of disease Y in the presence of agent X relative to the risk of disease Y in the absence of X. Often, the level of exposure to X is left vague and/or includes a wide range. Usually, it is only after an association with an unspecified level of exposure to X has been established in numerous studies that investigators begin to grapple with the question of ‘dose-response’. Even then, it is often only to examine whether the degree of risk increases qualitatively with an increase in the exposure, without specifying the actual levels of exposure. In the epidemiologic literature a preference exists for the presentation of relative risks: the ratio of the risk (or rate) of an outcome in the exposed group(s) to the risk (or rate) of the outcome in the unexposed group in a specified time period. In toxicology, the measure of effect is usually expressed as the absolute difference in risk between the exposed and unexposed group. For cohort studies in epidemiology, there is not any obstacle to examining risk or rate differences. However, in case-control studies for which disease incidence in the source population is not known or not estimable, only relative risks (through odds ratios), not absolute risks or rates, can be calculated. Consequently, risk differences cannot be estimated. Nested studies can provide absolute risks or incidence rates. Relative risk models are the most generally used models in epidemiology, primarily because they appear to describe the relationship between exposure and the age-specific risk of a disease better than absolute risk models. Some have argued that relative risk models also provide a
convenient framework for communicating information about risk (Krewski et al., 1999). Others have argued that risk differences are more fundamental and communicate the public health impact of an intervention most clearly. In fact, the preference for relative risk rather than risk difference may stem from the lack of availability of software packages that can appropriately model risk on an additive scale. This obstacle has recently been overcome. Differences between toxicologic and epidemiologic approaches for handling dose–response analyses stem largely from the difference in emphasis regarding use of additive vs. relative scales for comparison.

3.4. Evaluation of evidence

Conclusions of epidemiologic studies should be formulated in a way that can be better interpreted by risk assessors. A major requirement is guidance to the way “negative” or “non-positive” studies are to be interpreted. In the 1986 EPA cancer risk assessment guidelines it was stated that “if adequate exposure data exist in a well designed and well-conducted negative epidemiologic study, it may be possible to obtain an upper-bound estimate of risk from that study” (EPA, 1986). Historically, there has been a tendency to give positive studies more weight than negative studies, with the view that one cannot prove the null hypothesis, but that positive studies disprove it (Shore et al., 1992). In traditional risk assessment based on toxicologic data, only studies with a statistically significant positive result are considered, and the study showing the highest risk (among studies of good quality) is selected to be used in the final risk assessment (worst-case scenario, positive evidence approach). Epidemiologists tend to work according to the “weight of evidence” approach, in which information from all good-quality studies is combined with the aim of obtaining a risk estimate with as much precision and validity as possible (Shore, 1995). In the case that risk assessment is performed according to the “weight of evidence”, however, null results or results in the negative direction have equal importance as positive results whether they are statistical significance or not. In meta-analyses, the basic rule in deciding which studies are eligible for inclusion is to include all studies that are properly conducted, relevant and unbiased. No studies should be excluded on the grounds of size, as the totality of evidence is needed and it is irrelevant whether the individual study could or could not detect a particular size of risk by itself. However, the weight attributed to studies depends on their precision (size). Neither should any study be excluded on grounds that differentiate between those showing positive and negative results, since this could lead to the event that in one dataset a positive effect is accepted (at the 1 in 20 level of significance) whereas from 19 other similar datasets results were excluded (Doll, 1984; Shore et al., 1992).

3.5. Assessment (safety) factors

In toxicological risk assessment, safety factors, also indicated as adjustment or assessment factors, are used to take into account (1) the uncertainty with respect to extrapolation from animals to humans and (2) the variability among humans. When no other toxicokinetic and toxicodynamic data are available, default adjustment factors are being applied. Edler et al. (2002) provides a detailed description of and proposal for the use of the adjustment factors. The default factors are 10 to account for interspecies extrapolation and 10 for human variability. It was also proposed to subdivide both 10-fold factors to account for dynamic and kinetic uncertainty and variability. Additional data on the fate and behavior of the chemical in the animal and human body may be used to replace default adjustment factors by specific, evidence-based factors (see also Dybing et al., 2002; Edler et al., 2002).

Another source of uncertainty in risk assessment based exclusively on animal studies is uncertainty about the shape of the dose–response curve: studies are conducted with high doses, but have to be extrapolated to the low range of exposure to reflect the human exposure situation. This uncertainty is generally assumed to be included in the interspecies extrapolation factor.

In epidemiology, data usually refer directly to humans; therefore, interspecies uncertainty factors are not required, and for diet-related assessment there is usually no need to extrapolate outside the range of exposure in the observed data. The situation is different, however, for the variability within a human population. Here, some of the considerations in toxicological risk assessment also apply to risk assessment based on epidemiologic data. When no data are available other than an estimate of the risk from exposure in the general population, sensitive subgroups such as neonates, pregnant women or the elderly call for an application of comparable adjustment factors as in toxicological risk assessment (see Dybing et al., 2002; Edler et al., 2002). Default adjustment factors of 3.16 (square root of 10) are mentioned to cover variability within the population with respect to kinetic and dynamic characteristics. Particular attention has to be paid to subgroups with specific metabolism modifying polymorphisms, which may be much more susceptible than assumed by the default factor (see also section 5.2.1.3 for a discussion on susceptibility). Other factors that may influence susceptibility include medication usage.

Another source of uncertainty and variability specific for epidemiologic data is exposure measurement error.
When validation or calibration study data are available, it is possible to correct the results of a study for measurement error. From such studies, it has also become clear that measurement error, in the area of dietary assessment, can reduce a true relative risk of 2 into an observed relative risk of 1.3 to 1.6, depending on the accuracy of the exposure measurement (Willett, 1998). This corresponds with a 2- to 3-fold adjustment factor.

Finally, precision (statistical stability) of the risk estimates of a specific study or of combined studies plays a role in uncertainty. Using the upper confidence limit (UCL) of the risk estimate generally would address this uncertainty. The same issue, however, also applies to animal studies. The EPA takes the UCL derived from an animal study into account in the risk characterisation process; in Europe, however, this is not (yet) a common practice.

3.6. Estimation of lifetime risk based on epidemiologic data with less-than-lifetime follow-up

Epidemiologic data differ from data obtained in 2-year rodent bioassays in several ways (Hertz-Picciotto and Holtzman, 1989; Hertz-Picciotto and Hu, 1994). One of the most important distinctions is that the 2-year rodent study yields estimates of lifetime risk at each dose group. In contrast, most epidemiologic studies do not involve sufficiently long follow-up to measure lifetime risks directly. Nevertheless, any measure of association obtained from epidemiologic studies with less-than-lifetime follow-up can be used to estimate lifetime risk. The methods to accomplish this have been reported (Thomas et al., 1992; Hertz-Picciotto and Hu, 1994; Hertz-Picciotto et al., 1997; Korte et al., 2000) and are summarised briefly here.

1. The first step is to obtain a slope relating the exposure to the rate of disease in the population studied. The exposure can be expressed as a daily, weekly, or perhaps yearly intake of a substance, or alternatively, as a cumulative exposure, for each unit of person-time of observation (e.g. person-year). For some substances, there may be virtually no-one who is unexposed, in which case the comparisons are among different levels of exposure. Whether or not there is an unexposed group in the study population, a variety of mathematical forms can be used for expressing the increment in disease rate as a function of the increment in exposure. First, the relation might be linear, quadratic or some other form (polynomial, or spline relations). Secondly, the increment in disease can be additive or multiplicative over the background rate of the disease. Because age is such a strong determinant of disease risk, an additive model is less appropriate for deriving lifetime risks from partial-lifetime studies, since the added risk will be greatly influenced by the age distribution of the person-time in the study population. Thus, relative risk models are recommended in the model fitting stage. In rare instances, where, for example, risks are available across the entire age distribution, or exposures have spanned a complete generation, lifetime excess risk may actually be observable. In any case, for many epidemiologic studies of dietary factors, exposure is simply categorised and relative risks for the disease are calculated within each category, relative to a baseline or low exposure group. When the original study presents results in this form, the risk assessor will then decide what mathematical function to fit to the published relative risks. Ideally, the median exposure would be available for each category; if not, a midpoint can be used.

Once a model has been fit to the study data, the estimated model parameters can be applied to the general population. (The term “general population” will be used to indicate the population for which lifetime risk is to be estimated. It may be the total population of a country or group of countries, or it may be certain subgroups.)

2. The following additional pieces of information are needed:

(a) The distribution of exposure in the general population. If the exposure is a novel one, then a hypothetical distribution will need to be assumed. For a continuing exposure, some source of data that provides the distribution will be needed. If there is a considerable lag-time between exposure and disease, then a past exposure distribution will be more appropriate to estimate current risks. Preferably, the exposure distribution in the general population will overlap with the distribution in the studied population. If not, it is important to recognise that the exposure X disease relationship is being extrapolated beyond the range of the observed data. Also, if the distribution varies by age group and/or by gender, then it would be ideal to have age-specific and/or gender-specific exposure data.

(b) The background rate of the disease or outcome of interest in each age group in the general population. For the remainder of this discussion, we assume that death is the outcome of interest, but slight modifications can be applied to conduct these calculations for disease incidence.

(c) The background all-cause mortality rate in each age group in the general population. As mentioned above, these should be gender-specific, if either the exposure distribution or the relative risks from the published study differ by gender.

(d) Relative risks relating the exposure of interest to causes of death other than the disease of interest. This can be for a particular cause of death, or for ‘all causes of death’ or ‘all other causes of death’.
If no data are available regarding other causes of death than the disease of interest, then an assumption will need to be made; for instance, there is no increased risk associated with exposure.

3. A life table is constructed for each level of exposure of interest. The life table begins at birth and continues such that the final interval ends at either a near maximum lifetime (say 100 years) or somewhere near a mean or median lifetime (for example, 75 or 80 years in many western European countries). The next steps are slightly different depending on whether the exposure is a novel one or a continuing one, though the principle is the same.

(a) If the exposure is a novel one, then the current life table is used to calculate the lifetime risk for an unexposed population. The lifetime risk for an exposed population is derived by applying the mathematical model that was fit to the epidemiologic study to the ‘background’ mortality rates, for example those in the unexposed population, at each age (in other words, if the rate ratio in the study was 2, then the rates at each age are multiplied by 2). These new ‘predicted’ mortality rates are used to construct the life table for the exposed population. Note that the total mortality rate must be partitioned into the rate for the disease of interest, which is affected by the exposure, and the other causes of death. If the other causes of death are not influenced by the exposure, then the age-specific rates for these are added back onto the ‘predicted’ (exposure-influenced) rates for the cause of interest to obtain the ‘predicted’ rate for ‘all causes.’ If other causes of death are also influenced by exposure, then the rates for these causes must also be altered according to the appropriate models or relative risks. The predicted total mortality rates (all causes of death) are then used to construct an ‘exposed’ life table.

If an exposure rate (average intake per unit of time) was used in the epidemiologic study from which the exposure X disease relationship was obtained, then this exposure rate can be assumed at all ages, and the new rates at each age will be a function of the background rates and the relative risk for that exposure rate. If, on the other hand, cumulative exposure was used in the epidemiologic study, then the age-specific cumulative exposures will need to be calculated and the new mortality rates will be an age-specific function of the background rate, the cumulative exposure, and the slope from the mathematical model.

Once the life tables have been constructed, two approaches to comparison of exposed and unexposed life tables can be used. In the first, the cumulative risk through the last age interval is calculated for the given exposure level, and compared to the cumulative risk up to the same achieved lifetime for an unexposed or low exposure population. The difference is designated the ‘excess lifetime risk’ (ELR). In the second approach, the unconditional risks at each age (probability of surviving to that age times the probability of dying of the disease of interest at that age) are compared, and the differences are summed over the lifetime (Thomas et al., 1992). Within the life table, one can calculate lifetime risk for specific causes, considering other causes as competing risks. Simultaneously one can calculate lifetime risk for ‘other causes’ and for ‘all causes.’

(b) If the exposure is a continuing one, rather than a novel one, then existing mortality (incidence) rates must be partitioned across the various levels of exposure. At each age, the population mortality rate is a weighted average of the rates at each level of exposure, where the weights represent the proportion of the population at each exposure level multiplied by the relative risk associated with that level of exposure. Simple algebraic formulae are then used to derive the partitioned rates, to be used in constructing a life table at each exposure level (Hertz-Picciotto and Hu, 1994; Korte et al., 2000). If the exposure affects more than one cause-of-death then this partitioning can be carried out for each one, but must be done simultaneously due to competing risks. Examples applying this partitioning methodology to smokers and non-smokers in the population have been described (Korte et al., 2000).

To summarise, many rodent studies follow the animals for an assumed lifetime and the resulting proportions that die of, or with, the disease, represent the lifetime risk. In human studies, direct observation of lifetime risk is rare; nevertheless, epidemiologic analyses of disease risk over shorter time periods can be used to derive lifetime risk, under certain assumptions. The first assumption is that used in most life tables: cross-sectional age-specific mortality rates are assumed to apply to a cohort as it ages. The second assumption is that the form of the exposure–disease relation is applicable across various ages (although if there are sufficient epidemiologic data to provide age-specific exposure–response relations, then empirical data can be used and no assumption is needed). Third, the exposure–disease relation observed in the epidemiologic study is assumed to apply to a general population. This last assumption is fundamental to all risk assessments.

3.7. Conclusions

Although differences between toxicological and epidemiologic procedures and definitions can be identified, these are not insurmountable obstacles, since the estimated parameters in both cases can be converted into the desired quantity, namely risk per unit of exposure.

Because interspecies differences can be of such large magnitude, the uncertainty in epidemiologic studies, due principally to confounding and measurement error, are
considerably lower than those from toxicologic animal experiments.

Use of epidemiologic data reduces the dependence on safety or adjustment factors. Some factors, however, must still be applied (that is, if specific evidence-based information is not available) to account for human variability and for the error in measurement of exposure.

4. Examples of the contribution of epidemiology to the risk assessment process

Although the use of human data has not been legally mandated in the risk assessment procedure, epidemiology has made important contributions on several occasions in the past. In the area of air pollution, epidemiologic data are the primary source of information for risk assessment. Most other applications come from the areas of occupational and environmental epidemiology, where accidental and/or occupational, relatively high exposures are more common than dietary exposure to potentially harmful substances. In this section, some examples will be given demonstrating the potential contribution of epidemiologic data to risk assessment in the food area.

4.1. Aflatoxins and liver cancer

Aflatoxins are produced by the species Aspergillus and are considered to be liver carcinogens. Exposure arises mainly from contaminated food staples, and levels are particularly high in parts of sub-Saharan Africa, south-east Asia and Central America. Aflatoxins clearly play a role in the induction of liver cancer in these countries. Another important risk factor for primary liver cancer is infection with the hepatitis B virus (HBV).

Public concern about liver cancer risk from exposure to aflatoxins (e.g. aflatoxin B1) has stimulated a large volume of research on the carcinogenic effect of exposure to aflatoxin, both in animal models and human studies (Hoseyni, 1992).

A risk assessment of aflatoxin and human cancer was performed based on extrapolation of laboratory animal data, for which the most sensitive species (rat) was chosen using a linearised multistage model (Eaton and Gallagher, 1994). This led to an estimated potency of $6.25 \times 10^{-5}$ (ng/kg/day)$^{-1}$. Mean US aflatoxin intake is estimated to be 110 ng/kg/day, leading to an excess lifetime cancer risk in the US due to aflatoxin of $6.9 \times 10^{-2}$, which yields an annual incidence of liver cancer of 98 per 100,000. However, the actual liver cancer incidence from all causes is 20 times lower, leading to the conclusion that the rat is an inappropriate model to project human cancer risk for aflatoxin, primarily because of large metabolic differences in cytochrome P450 enzymes and glutathione S-transferases (GSTs).

Results from animal models indicate a large degree of variability in the carcinogenic potency of aflatoxin across species; hence, human epidemiologic data are needed to provide the most appropriate basis for determination of safe exposure levels for humans (Hoseyni, 1992). A large number of epidemiologic studies have been carried out, such as correlation studies, case-control studies and cohort studies. Many of the epidemiologic studies have been considered inadequate for aflatoxin risk assessment; some lacked information regarding population liver cancer rates while others suffered from the absence of quantified exposure levels. Other studies were considered to be appropriate for risk assessment, but these were all conducted in areas where the prevalence of hepatitis B infection is known to be considerably high (Hoseyni, 1992). However, they provided a high quality of data; liver cancer incidence was determined by prospective surveillance and hepatitis B surface antigen status (HBsAg) in the same subjects was determined by modern methods. (Recent) exposure to aflatoxin can be assessed by detection of urinary metabolites or DNA adducts of aflatoxin. Based on one particular study it is shown that because of the presence of data on hepatitis B infection, it is possible to separate the effects of aflatoxin exposure and hepatitis B. Then, the potency of aflatoxin, adjusted for hepatitis B prevalence, can be estimated for other populations by application of the estimated dose–response coefficient to baseline rates.

A prospective cohort study on aflatoxin exposure, hepatitis B infection and liver cancer risk was performed in Shanghai (Ross et al., 1992) involving 18,244 men, aged 45–64 years, who were followed for up to 3 years (35,299 person years). During this follow-up period, 22 incidents of primary liver cancer occurred. For analyses, a nested case-control design was used, which means that for every case that occurred some controls were chosen form the cohort for comparison. For the 22 cases, 140 controls were chosen. At baseline, dietary intake was estimated using a food frequency questionnaire, data on smoking habits and alcohol intake were collected, and a single 24-hour urine sample and 10-ml blood sample were taken and aflatoxin–DNA adducts were analysed in urine samples of cases and controls. The study was aimed at investigating the liver cancer risk dependency on aflatoxin exposure, HBV infection (HBsAg status) and the interaction between exposure and infection.

Table 2 shows the presence of several urinary aflatoxin biomarkers in both cases and controls, showing a 2.4-times higher risk of liver cancer for subjects with biomarkers present (and thus exposed to aflatoxin) compared to subjects without presence of biomarkers.

Multiple logistic regression analysis (including aflatoxin–DNA adducts, HBsAg positivity, smoking,
education and alcohol consumption) shows that exposure to aflatoxin (compared to no exposure) was responsible for a 4-times higher risk for liver cancer (Table 3), while having had a hepatitis B infection increases risk eight times. Most interesting, however, is Table 4, where interaction between the exposure to aflatoxin and the hepatitis B virus is shown. The presence of both factors (exposure to aflatoxin and the HBV virus) leads to a 60-times higher risk compared to having had none of these exposures. However, it should be noted that the effect of aflatoxin in those who are HBsAg negative is unstable and consistent with no increased risk.

Including epidemiologic studies in risk assessment leads to the estimation of the "epidemiologic potency"; human cancer risk from lifetime exposure (Eaton and Gallagher, 1994). The potency was estimated to be $8.2 \times 10^{-5}$ in HBV-endemic areas (China) and $2.8 \times 10^{-6}$ in HBV-absent areas (30 times lower than in HBV+). These estimates were confirmed by those mentioned above with relative risks of 60 in subjects with hepatitis and 2 in subjects without hepatitis (Ross et al., 1992). A detailed consideration of the potency of aflatoxin based on experimental and human data is summarised in a JECFA (Joint FAO/WHO Expert meeting on Food Additives) (1998) report which highlights some of the uncertainties in making such estimates even when such an extensive database is available, as is the case for the aflatoxins. This is a clear example of the contribution epidemiology can make to risk assessment. In the present situation, animal models failed to provide a relevant risk estimate. The effect of combinations of etiological factors, such as aflatoxin exposure and hepatitis B, studied in an observational setting, can be an important tool in future risk assessments.

4.2. The ATBC and CARET intervention trials

Many observational epidemiologic studies (mainly case-control) have found an inverse association between intake of vegetables and/or fruit and the risk of lung cancer, besides other cancers. In order to find the potentially responsible nutrients, the intake of vitamins was estimated in these studies by combining the food intake data with available nutrient databases in the 1970s and 1980s. These databases are historically focused on macronutrients (fats, protein and carbohydrates) and some micronutrients such as vitamin C, retinol and beta-carotene. Because of its presence in vegetables and fruits, beta-carotene often appeared as the interesting constituent behind the association between vegetables/fruit and lung cancer. This was supported by studies in which serum levels of beta-carotene were associated with cancer risk. There also was a biological plausibility, since beta-carotene appeared to function as antioxidant. It should be noted, however,

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of urinary aflatoxin–DNA adducts and risk of liver cancer in univariate analysis (Ross et al., 1992)</td>
</tr>
<tr>
<td><strong>Aflatoxin</strong></td>
</tr>
<tr>
<td>No aflatoxin biomarkers detected</td>
</tr>
<tr>
<td>AFB1-N7-Gua</td>
</tr>
<tr>
<td>AFP1</td>
</tr>
<tr>
<td>AFM1</td>
</tr>
<tr>
<td>AFB1</td>
</tr>
<tr>
<td>Any of these compounds</td>
</tr>
</tbody>
</table>

RR, relative risk; CI, confidence interval.

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urinary aflatoxin–DNA adducts and risk of liver cancer in multiple logistic regression analysis (Ross et al., 1992)</td>
</tr>
<tr>
<td><strong>Factor</strong></td>
</tr>
<tr>
<td>Presence of urinary aflatoxin–DNA adducts</td>
</tr>
<tr>
<td>HBsAg positivity</td>
</tr>
<tr>
<td>Level of education (high vs low)</td>
</tr>
<tr>
<td>Cigarette smoking (ever vs never)</td>
</tr>
<tr>
<td>Alcohol consumption (30+ g/day)</td>
</tr>
</tbody>
</table>

RR, relative risk; CI, confidence interval.

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction between aflatoxin exposure, hepatitis B infection and liver cancer risk (Ross et al., 1992)</td>
</tr>
<tr>
<td><strong>Aflatoxin negative</strong></td>
</tr>
<tr>
<td><strong>RR</strong></td>
</tr>
<tr>
<td>HBsAg Negative</td>
</tr>
<tr>
<td>Positive</td>
</tr>
</tbody>
</table>
that beta-carotene could also be simply a marker of intake of vegetables/fruits with no cancer-preventive property. Because only a few other micronutrients were available in the nutrient databases, only beta-carotene could turn up in the analyses. Nevertheless, following a highly publicised report in Nature (Peto et al., 1981) plans were made to design large-scale randomised controlled intervention trials to test the preventive effect of beta-carotene.

The Alpha-Tocopherol, Beta-Carotene Cancer Prevention Trial or ATBC Trial was conducted by the National Public Health Institute of Finland in collaboration with the US National Cancer Institute (NCI) (The Alpha-Tocopherol, 1994). The purpose of the study was to see whether certain vitamin supplements would prevent lung and other cancers in a group of 29,133 male smokers in Finland. The 50–69-year-old participants took a pill containing either 50 mg α-tocopherol (vitamin E), 20 mg beta-carotene (a precursor of vitamin A), both, or a placebo daily during the follow-up of 5–8 years. The Beta-Carotene and Retinol Efficacy Trial or CARET was a large chemoprevention trial that is being conducted in six areas in the United States (Omenn et al., 1996). The purpose of the study was to see whether the combination of beta-carotene and vitamin A (retinol) supplements would prevent lung and other cancers in men and women of ages 50–69 years who were smokers or former smokers, and men of ages 45–69 years who had been exposed to asbestos. Totally unexpectedly, in the ATBC Trial, 18% more lung cancers were diagnosed and 8% more overall deaths occurred in study participants taking beta-carotene compared with those taking a placebo. This was seen after 6 years of follow-up, despite an initial small benefit observed for beta-carotene in the first year of follow-up. In the CARET trial, after an average follow-up of 4 years of receiving supplements, 28% more lung cancers were diagnosed and 17% more deaths occurred in participants taking beta-carotene and vitamin A than in those taking placebos. Neither of these studies showed a benefit from taking supplements. The study group will be followed for 5 more years to determine the long-term effects of the intervention. Unfortunately, no dose–response relationship for beta-carotene effects was available from the intervention trials in humans, since single doses were used in each study. Mean blood levels of beta-carotene in these trials, however, were higher than those in a third trial, the Physicians’ Health Study (Hennekens et al., 1996), in which no excess of lung cancer and mortality was observed (Woutersen et al., 1999). In the meantime, new cohort study findings were reported in which the intake of more carotenoids besides beta-carotene were tested; this study showed that there was no inverse association with beta-carotene anymore once additional micronutrients were taken into account (Voorrips et al., 2000).

Lessons learned from these studies can be summarized as follows:

- Longer follow-up periods can provide different results than short follow-up periods.
- It is important to study disease endpoints in addition to intermediate endpoints.
- High-quality, more extensive nutrient databases can limit the chance of focussing on the wrong constituent.
- Dosages outside the normal range of human exposure may result in detrimental or other unexpected effects.
- Choice of one high dose may enhance the chance of success in efficacy trials, but seriously limits the usefulness of the results for the purpose of risk assessment if adverse effects occur.

5. Tailoring epidemiologic studies for risk assessment purposes

5.1. Existing studies

In ongoing epidemiologic studies, depending on their phase of conduct, some adaptations or additions can be made in the process of (exposure and outcome) data collection, data analysis and presentation of results in order to make the results more suitable for incorporation in the risk assessment process.

5.1.1. Presentation of exposure data and quality of exposure assessment in epidemiologic studies

5.1.1.1. The importance of reporting the appropriate data. For risk assessment purposes, the key issue in any otherwise valid epidemiologic study is the presence of quantitative data on exposure. Although epidemiologic studies are often available, they commonly do not provide absolute exposures, but rather present qualitative categories of exposures. In other words, many epidemiologists are concerned primarily with a correct rank ordering of subjects according to exposure, rather than a correct estimation of absolute exposure level. Sometimes this is because of the inadequacy of the exposure data, leading epidemiologists to feel uncomfortable with assigning quantitative measures. Nevertheless, for the dose–response stage of risk assessment, a well-conducted epidemiologic study with no quantitative information on exposure is of no use. Knowing the range of exposures within the study population, even if only a general ballpark figure, or, say, likely lower and upper limits, greatly enhances the utility of the study results. Optimally, the risk of the disease should be provided in each of several categories of exposure, where each category is defined by its lower and upper limit and reported
with its mean and median. In case-control studies, the outcomes will not be absolute risks, but rather estimates of relative risks. As any use of these data for dose–response modelling requires a single value be used to summarise each category of exposure, the mean or median is needed for risk assessment.

In nutritional epidemiology, it is common to use tertiles, quintiles, or some other quantile-based categories for food-based nutrients. In this case, it is imperative that the authors provide the upper and lower values of the absolute measurements, as well as the median or mean within each of the quantiles with special attention given to the extremes (uppermost and lowest quantiles). Since the distributions of the nutrients will differ according to the population studied, the same quantiles will not apply to other populations. In fact, use of the same cut-off points would enhance the ability to compare results across studies. Nevertheless, if quantiles are to be used, the absolute levels must also be reported if the data are to be of use for quantitative risk assessment.

5.1.1.2. Use of external exposure data to achieve more accurate absolute estimates of exposure. When studies do not provide quantitative data, there may be external information (from the same or from other, but similar populations) that can be used by risk assessors to supplement the published data and render them usable in risk assessment. With respect to dietary assessment methods, it should be emphasised that a lot of work has been done by validating dietary questionnaires for use in large studies to develop more precise methods of dietary assessment. Data from validation (or calibration) studies are particularly useful for the purpose of risk assessment. Subjects are grouped into categories such as quintiles on the basis of the surrogate method (e.g. a food frequency questionnaire); then, the “true value” for the same subjects based on the more accurate “gold standard” is assigned to the categories defined by the surrogate method (Willett, 1998). Such a presentation of the exposure data is also feasible for continuous variables by applying simple regression. A biomarker of exposure in blood or urine, provided that it is known to reflect exposure well, can be used in the same way (see section 5.2.1.1).

5.1.1.3. Use of validation data to correct for measurement error. Validation or calibration substudies, particularly within large studies, are also very useful to apply a formal measurement error correction to the observed relative risks. This procedure makes a default adjustment factor for exposure measurement error largely redundant. Statistical methods to improve the measures of association by incorporating validation or calibration substudy results have been reviewed by Willett (1998). Biomarkers could be of importance in the identification and quantification of exposure-measurement error (McKeown et al., 2001) and to calibrate exposure measurements (Riboli and Kaaks, 2000).

5.1.1.4. Enhanced use of additional databases with content of chemicals in foods and beverages. In some cases, existing nutrition-oriented studies can be enhanced by linking study data with information from external (ad hoc) databases, for example on contaminants. Of course, the additional database should have sufficient quality to be used for such a purpose (see also Kroes et al., 2002). An example is the study by van Loon et al. on nitrate intake and stomach cancer in the Netherlands Cohort Study on diet and cancer (van Loon et al., 1998). For this study, a database was assembled with data on nitrate content of all vegetables and other nitrate and nitrite containing foods. The nitrate composition data were derived from numerous samples of vegetables, collected and analysed for regulatory and quality control purposes during a 4-year period. For each vegetable, a long-term average nitrate value was established and linked to the food consumption data available from the study. In addition, nitrate intake from drinking water was based on information on mean nitrate content in drinking water for each pumping station combined with information on the distribution of drinking water from these stations and the residence of the subject. This example illustrates that it is often feasible and relatively inexpensive to use or construct databases of foods with quite accurate information on content of particular contaminants or residues, based on routinely collected data. However, before using such routinely collected information, it is important to verify how the collection of samples was established: was it likely to be a more or less random selection or were the samples taken because of suspicion of exceeding set limits? In the latter case, the estimated exposure to the substance in question would be (much) higher than the true exposure of the subjects under study, which results in underestimation of the risk associated with exposure.

In traditional tables data can be found on content of both macronutrients and (a selection of) micronutrients of separate food items. By adding new available analytical data on nutrients and, in particular, other food components that were previously not included in food composition tables [such as different carotenoids, phytoestrogens, phytosterols, individual fatty acids, butylated hydroxyanisole (BHA)/butylated hydroxytoluene (BHT)], dietary assessment can be improved and extended even in existing studies. Some of such data are present in the food industry, which may make them available for research under specific conditions.

5.1.2. Shape of the exposure–response curve

Modelling dose–response relationships is central to risk assessment (see also Edler et al., 2002). Once a dose–response model characterising the relationship
between dose and response has been developed, estimates of risk at any dose within the range of the original data can be obtained. If necessary, the model may also be used to predict risks outside the range of the original data, although extrapolations well beyond the original data range are subject to considerable uncertainty, because of the uncertain shape of the relationship (Krewski et al., 1999). While risk assessment in the field of environmental exposures has most often involved extrapolation outside the range of observed data, in the field of food-related exposures, the exposure range of concern often overlaps considerably with the range of exposures studied in epidemiology. Both empirically based and biologically based dose–response models have been developed for the purpose of risk assessment. The differences between empirical models used to describe dose–response relationships in epidemiologic and toxicological studies are based primarily on the difference in the magnitude of the response. While there is no compelling reason, on theoretical grounds, why models traditionally used in epidemiologic studies could not be applied to laboratory data and vice versa (Krewski et al., 1999), in practice, most models used for toxicologic studies (such as the multistage) are well suited for studies in which the risks span a wide range, possibly as high as 50 or 80% or higher. In epidemiologic studies, the risks evaluated are often smaller than 10 or even 2%, for which different models, such as the logistic or log-linear, are more appropriate. Recently, flexible approaches to dose–response modelling are being used more frequently in epidemiologic studies, in particular if the statistical power of a study is sufficiently large. An example is the use of spline regression, a non-parametric method to generate dose–response curves that make fewer assumptions about the shape of the curve (Smith-Warner et al., 1998). With a graphic presentation, these non-parametric regression curves can also be compared with the linear model to evaluate linearity of the associations by using a likelihood ratio test. Spline regression methods cannot be used to extrapolate outside the range of the observed exposures. Other interesting types of models are fractional polynomial models.

An additional point is that investigators should provide the incidence rates in each exposure category, or at minimum, RRs, in their published papers to enable risk assessors to use alternative models from the ones favoured by the original investigators. In some cases, a more complex model will provide a better fit, but a simpler model may be more appropriate for assessment of risks on a population-wide basis. Only if the published report presents risks or relative risks at each level of exposure, and the median exposure within each category, can the audience of risk assessors utilise the data to the fullest in dose–response assessment. Of course, providing the original data would provide risk assessors the maximum flexibility in determining the relevant dose–response relationships.

5.1.3. Mixed exposures

Observational studies in humans must face the problem of mixed exposures. There are two aspects of this phenomenon: on the one hand, any particular exposure of interest may be highly correlated with other exposures, as for instance, when one micronutrient tends to be associated with another. Alternatively, it is sometimes the mix of exposures that is of interest.

The first aspect, particularly plaguing nutritional epidemiology, is due to the fact that people eat foods, not separate chemicals. Related substances tend to cluster in the same foods, inducing a high degree of collinearity (correlation) of the substances. High collinearity makes separation of the effects of the different substances difficult. In other words, one cannot be completely sure as to which substances a specific risk or benefit can be attributed. The problem of collinearity is, however, likely to be much more important for substances occurring naturally in foods such as carotenoids rather than for xenobiotic substances.

The weakness caused by collinearity, highlights at the same time a major strength of epidemiology over toxicology: in particular, observational studies have the potential to quantify risks associated with combined exposures of (foodborne) agents, among themselves, or in combination with other substances, such as drugs or xenobiotics in the environment. As exogenous exposures other than those tested in a particular experiment can influence the health effects of a certain exposure, results from animal experiments may lack relevance to the human, real-world experience (Hertz-Picciotto, 1995). Of particular concern is the possibility of synergistic effects, in which the risk associated with joint exposure to two or more agents exceeds the sum of the risks for each agent alone (Krewski et al., 1999), for an epidemiologic example see (Hertz-Picciotto et al., 1992). Similarly, when another exposure has antagonistic effects, it will reduce the magnitude of risk or eliminate it entirely. Epidemiology allows research into such combined exposure scenarios: the larger the study, the better these interactions can be addressed.

5.1.4. Quality assurance

To further stimulate the use of epidemiologic studies in risk assessment of food constituents by guaranteeing quality assurance, several initiatives are being taken to develop specific guidelines. Guidelines for Good Epidemiology Practices (GEPs) originally have been formulated by the Chemical Manufacturer’s Association Epidemiology Task Group (The Chemical Manufacturers Association’s Epidemiology Task Group, 1991). These GEPs propose minimum practices and procedures that should be considered to help ensure the
quality and integrity of data used in epidemiologic research and to provide adequate documentation of the research methods, with specific reference to occupational studies. The International Society for Pharmaco-Epidemiology (ISPE, 1996) has specified these guidelines for drug, device and vaccine research in the United States. These guidelines propose practices and procedures in areas such as elaboration of the protocol, description of roles and responsibilities of organisations and personnel involved, study conduct (protection of subjects involved, data collection and verification, analysis and report), communication and archiving.

5.1.5. Use of ongoing studies for the purpose of risk assessment

Non-epidemiologists sometimes express the opinion that epidemiologic studies can hardly be conducted for the purpose of risk assessment, because of the costs and the time involved. This may also be the reason why risk assessors or regulators almost never advocate setting up a new epidemiologic study for this purpose.

Although in general it is true that a good epidemiologic study takes more time than a study in rodents and may also be more expensive if it has to be set up for one specific purpose only, good alternatives are available. The dietary habits and subsequent disease occurrence of more than 3 million people are currently under observation all over the world in cohort studies (Table 5); many of these studies also have the disposal of a biobank, with biological samples that can be used in nested designs. The majority of these studies have been designed to study chronic effects of diet and other possible risk factors on chronic diseases with relatively long induction times. The infrastructure of these studies can and must be exploited more specifically for risk assessment in the food safety domain. The additional costs of such exploitation are modest in comparison to a chronic rodent study.

5.2. New studies

Future and also ongoing long-term prospective studies could profit from new insights and recently developed designs and techniques. The developments that can contribute to use these studies for risk assessment purposes are described below.

---

Table 5
Large prospective studies of diet and disease using comprehensive food-frequency questionnaires (adapted from Willett, 1998)

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>No. Recruited</th>
<th>Biological specimens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adventist Health Study</td>
<td>Males and females</td>
<td>40,000</td>
<td>1976</td>
</tr>
<tr>
<td>Nurses Health Study (USA)</td>
<td>Females</td>
<td>90,000</td>
<td>1980</td>
</tr>
<tr>
<td>New York State Cohort</td>
<td>Females</td>
<td>25,000</td>
<td>1980</td>
</tr>
<tr>
<td>Canadian National Breast Screening Study</td>
<td>Females</td>
<td>57,000</td>
<td>1982</td>
</tr>
<tr>
<td>Northern Sweden Health and Diseases Study</td>
<td>Males and females</td>
<td>55,000</td>
<td>1985</td>
</tr>
<tr>
<td>New York University Women’s Health Study</td>
<td>Females</td>
<td>14,000</td>
<td>1985</td>
</tr>
<tr>
<td>The Alpha-Tocopherol Beta-Carotene cancer prevention study (Finland)</td>
<td>Males</td>
<td>29,000</td>
<td>1985</td>
</tr>
<tr>
<td>Health Professionals Follow-up study (USA)</td>
<td>Males</td>
<td>52,000</td>
<td>1986</td>
</tr>
<tr>
<td>Netherlands Cohort Study</td>
<td>Males and females</td>
<td>121,000</td>
<td>1986</td>
</tr>
<tr>
<td>Iowa Women’s Health Study</td>
<td>Females</td>
<td>42,000</td>
<td>1986</td>
</tr>
<tr>
<td>ORDET (Italy)</td>
<td>Females</td>
<td>11,000</td>
<td>1987</td>
</tr>
<tr>
<td>Atherosclerosis Risk in Communities (USA)</td>
<td>Males and females</td>
<td>16,000</td>
<td>1987</td>
</tr>
<tr>
<td>Swedish Mammography Cohort</td>
<td>Females</td>
<td>61,000</td>
<td>1987</td>
</tr>
<tr>
<td>Melbourne Collaborative study</td>
<td>Australia born Italian and Greek migrants (M + F)</td>
<td>42,000</td>
<td>1990</td>
</tr>
<tr>
<td>JPHC (Japan)</td>
<td>Males and females</td>
<td>140,000</td>
<td>1990</td>
</tr>
<tr>
<td>Nurses Health Study H (USA)</td>
<td>Young female nurses</td>
<td>95,000</td>
<td>1991</td>
</tr>
<tr>
<td>Women’s Health Study (USA)</td>
<td>Females</td>
<td>40,000</td>
<td>1992</td>
</tr>
<tr>
<td>Canadian Study of Diet, Life- style and Health</td>
<td>Males and females</td>
<td>100,000</td>
<td>1992</td>
</tr>
<tr>
<td>American Cancer Society</td>
<td>Males and females</td>
<td>184,000</td>
<td>1992</td>
</tr>
<tr>
<td>Women’s Health Initiative (USA)</td>
<td>Females</td>
<td>165,000</td>
<td>1993</td>
</tr>
<tr>
<td>Multi-Ethnic Cohort (USA)</td>
<td>Males and females</td>
<td>215,000</td>
<td>1993</td>
</tr>
<tr>
<td>EPIC* (9 European countries)</td>
<td>Males and females</td>
<td>482,000</td>
<td>1993</td>
</tr>
<tr>
<td>Women’s Antioxidant Cardiovascular Study (USA)</td>
<td>Females</td>
<td>8000</td>
<td>1994</td>
</tr>
<tr>
<td>California Teachers’ Study</td>
<td>Females</td>
<td>132,000</td>
<td>1995</td>
</tr>
<tr>
<td>National Cancer Institute (USA)</td>
<td>Males and females</td>
<td>540,000</td>
<td>1995</td>
</tr>
<tr>
<td>Cohort of Swedish Men</td>
<td>Males</td>
<td>49,000</td>
<td>1997</td>
</tr>
</tbody>
</table>

* ORDET, Prospective Study of Hormones and Diet in the Aetiology of Breast Cancer, JPHC, Japan Public Health Center-based Prospective Study, EPIC European Prospective Investigation into Cancer and Nutrition.

b Dietary data collected within a randomised trial.
5.2.1. Use of biomarkers

Among the extensions and improvements to epidemiologic methods that have become available, the use of biological material to derive biological markers (biomarkers) has attracted much attention. This strategy of collecting biological material in epidemiologic studies on a large-scale basis started in the 1980s (Table 5) and became an overall strategy of designing prospective epidemiologic studies in the 1990s. The reasons for building up large collections of biological material in the context of prospective epidemiologic studies were many, including: obtaining insight into the role of metabolic processes for disease risk (Khaw et al., 2001); using the increasing knowledge accumulating in basic science about metabolic and cellular processes and how to measure them. Thus, the use of biomarkers in epidemiologic studies will obviously lead to new knowledge about the risk and benefit of nutritional factors. Ideally, a biomarker should reflect a specific entity in the chain between exposure and medically apparent beneficial or adverse effects. Therefore, beside an accurate and reliable measurement of the biomarker, it must be validated with respect to the entity the biomarker is intended to reflect. The validation process usually requires epidemiologic studies that include endpoint data of high quality (Bartsch, 2000). Validation of a biomarker of exposure usually does not require large epidemiologic studies, but rather small metabolic studies among volunteers who are fed the substances of interest. In the case of a very harmful substance, such studies may be unfeasible and the validity of the biomarker with respect to exposure may have to be derived from animal studies and other, more circumstantial evidence. If a biomarker is well characterised and validated it can subsequently be used in epidemiologic studies. However, epidemiologic studies have limitations regarding the availability and collection of specific biological material. Hence, the collection of biological material in epidemiologic studies is usually restricted to peripheral blood, spot urine, toenail specimens and sputum. Additionally, the assessment of biomarkers in large epidemiologic studies could be expensive, depending on the methods being used. Miniaturisation of assays not only means that they use very small samples of blood or tissue but also reagent costs are dramatically reduced. For many biomarker assays a major obstacle to more widespread use is sample throughput. An important reduction in costs would be achieved by analysing biomarkers in a subsample of participants only, such as in nested case-control and case-cohort studies.

The number of well-characterised biomarkers is still limited, despite much enthusiasm for biomarkers in general. Thus for the future, the development and validation of biomarkers in the epidemiologic context is therefore an absolute need.

For the purpose of this review, we would suggest grouping the biomarkers as to whether they reflect exposure, health effects or genetic susceptibility. For chronic diseases such as cancer the placement of a biomarker along the continuum of disease progression starting from exposure, such a demarcation may be difficult and for any type of biomarker this categorisation may depend on the outcome being studied (Vineis et al., 1993).

5.2.1.1. Use of biomarkers that are representative of exposure to the agent of interest (see also Kroes et al., 2002). The accurate assessment of exposure has always been a crucial element of well-designed epidemiologic studies. In the case of aflatoxin and hepatocellular carcinoma described earlier (section 4.1), the chemical structure of aflatoxin is so characteristic that there is little if any ambiguity about the relationship between biomarkers (protein or DNA adducts) and the exposure. This may not always be the case since some dietary exposures may cause molecular changes that are “generic” such as oxidative damage. A further lesson from the aflatoxin study is that the successful use of the biomarkers of exposure (urinary metabolites and DNA

Example on acrylamide.

Acrylamide is an important industrial intermediate for polymer and resin synthesis. Exposure occurs in occupational settings and in the environment through the leaching out of monomer from various products. Acrylamide has been found to be a carcinogen in experimental animals and estimates of risk associated with low level exposures have been calculated.

Exposure to acrylamide in occupational settings has been quantitated by use of characteristic hemoglobin adducts and it has been frequently noted that nominally unexposed control subjects have readily detectable levels of the adduct. It has been shown recently that acrylamide is formed in the cooking of foodstuffs and that, at least in experimental animals, this source of exposure contributes substantially to “background” levels of hemoglobin adduct (Tareke et al., 2000). The significance of “background” levels of protein and DNA adducts in relation to risk assessment have been summarised (Farmer and Shuker, 1999). These above results for acrylamide suggest that risk may be under-estimated or inappropriately estimated if only the obvious sources are considered. Useful biomarkers of exposure will take account of exposures from all sources.
adducts) did not depend on highly sophisticated analyses or repeated measures. The improved estimation of the relative risks was based on analysis of a single-point urine sample collected at recruitment into the study. This may indicate that the biomarker resulted simply in better classification of individual exposure status. In fact a comparable analysis of the same cases within the cohort using more traditional questionnaire-based methods demonstrated this fact very clearly since classification of cases and controls into high, medium and low exposure to aflatoxin did not reveal significant relative risks of disease (Qian et al., 1994).

Biomarkers also have the advantage that they combine several routes of exposure, as is shown by the example in the text box.

The significance of “background” levels of protein and DNA adducts in relation to risk assessment have been summarized (Farmer and Shuker, 1999). These above results for acrylamide suggest that risk may be under-estimated or inappropriately estimated if only the obvious sources are considered. Useful biomarkers of exposure will take account of exposures from all sources.

Epidemiologic studies that can or did not collect biomarkers on the entire study population may nevertheless be able to use biomarkers to assess exposure in a four-step procedure:

1. assess the association between exposure and biomarker in a small but high-quality study among human volunteers, if possible;
2. assess exposure in categories in the full study population, using a method that is able to rank subjects according to exposure;
3. measure biomarkers in a random sample of subjects from each category of exposure;
4. estimate (mean) exposure for subjects within each crude category based on biomarker-based exposure levels in the sampled individuals.

In this way, exposure levels are “calibrated” by the application of biomarkers.

5.2.1.2. Use of biomarkers that are representative for the health effect of interest. The issue of surrogate endpoints measured in biological samples using a biomarker instead of the disease endpoint is somewhat complex. However, it continues to attract attention because of its great potential in reducing the time required to establish links between exposure and disease. Some of the key issues surrounding the use and interpretation of surrogate endpoints in cancer epidemiology have been summarised by Schatzkin et al. (1997). The validity of a potential surrogate endpoint of cancer is determined primarily by the extent to which the marker is a necessary event on the causal pathway to cancer. In fact, the existence of a plausible major alternative causal pathway weakens inferences from that marker to cancer. Further interest in the role of surrogate endpoints arises from the possibility of using them in studies of chemoprevention. Vineis and Veglia (2001) have described the principle of “propagation of a mark” as crucial to establishing whether a biomarker is on the causal pathway to cancer and its usefulness as a surrogate marker of disease. Evidence that such biomarkers exist comes once again from studies on aflatoxin. Experimental animal studies established that reduction of liver DNA adduct formation for a given dose of aflatoxin using a drug, oltipraz, was predictive of a reduction or even
elimination of the risk of cancer in that organ. Recently it has been shown in volunteer studies that oltipraz is capable of effecting a similar change in humans and large intervention studies using this drug are under way (Kensler et al., 1999).

A recent example of the use of a biomarker as a surrogate for the disease endpoint relates to the role of aristolochic acid in kidney cancer.

Unfortunately, apart from these instances, there is still very limited evidence that unequivocally shows that early effect biomarkers are truly on the causal pathway to disease. There is a great need for properly designed prospective studies that document the predictive value of proposed effect biomarkers.

5.2.1.3. Use of biomarkers for genetic susceptibility in epidemiologic studies (see also Eisenbrand et al., 2002).

Genetic susceptibility is sometimes related to a major gene defect that is linked to a high disease penetrance. This type of genetic susceptibility, however, does in general not affect an individual’s susceptibility of a dietary (or any other) exposure and will therefore not be further considered in the context of this review. Another, very relevant, type of genetic susceptibility refers to the variants of a gene (polymorphism) in the human population that are associated with a change of the metabolic function of the gene product. It is estimated that for cancer risk, subtle gene–environment interactions such as those engendered by polymorphisms in enzymes involved in metabolising environmental carcinogens are likely to be of greater importance in populations than major cancer genes with low frequency such as the retinoblastoma gene, the gene for familial adenomatous polyposis coli or the breast cancer genes BRCA1 and BRCA2 (Moolgavkar et al., 1999c).

This equally applies to other diseases than cancer, although there is still only a limited amount of evidence of diet–gene interactions in cancer and other diseases such as heart disease. The status of research has recently been summarised by Daly (2001) and Bingham (2001). Another function of the use of markers of genetic susceptibility is to obtain a better insight into causal mechanisms (Hunter, 1999). Hunter suggested that if a genetic variant changes the disease risk in the way as predicted by their altered gene product it will be a strong argument that the gene product itself is causally involved in the disease process.

As discussed above, individuals may have increased susceptibility because they carry low penetrating alterations in genes (genetic polymorphisms), which occur frequently and are more indirectly related to the disease process. For cancer, these include the genetic polymorphisms for activating enzymes (which usually catalyse oxidation reactions, phase I enzymes). They may cause risk factors to have either enhanced or decreased impact in humans. Alternatively the genetic polymorphisms for phase II enzymes that usually catalyse conjugation reactions with few exceptions will determine the extent of endogenous chemoprotection against genotoxic risk factors in humans. The development of techniques such as polymerase chain reaction partially coupled to restriction fragment length polymorphism methods (RFLPs), enables precise identification of an individual’s genotype. Specifically, genes coding for the activating super enzyme family cytochrome P450-iso-enzymes have been identified in several allelic variants. These variations include MSP1 restriction sites in CYP1A1, indicative of a point mutation in exon 7. Allelic variants are also observed for CYP2E1, CYP2A6, the aromatic hydrocarbon receptor and for the super gene family glutathione S-transferases (GSTP1). For this enzyme frequent null genotypes (GSTT1*0, GSTM1*0) are also detected. Epoxide hydroxylase, which activates epoxide intermediates to more water-soluble trans-dihydro derivatives, occurs in different allelic variants, as possibly also uridine diphosphate glucuronosyl transferases (UDPGT). For nutritional toxicants such as heterocyclic amines, slow and fast acetylatyng genotypes based on different allelic variants of N-acetyltransferase 2 (NAT2) may be of specific importance (Smith et al., 1995; Bartsch and Hietanen, 1996; d’Errico et al., 1996). These genetic variants may be associated with either enhanced or decreased rates of metabolic conversion by the specific enzyme.

Depending on type of metabolic conversion (activation, deactivation), the result will be more or less genotoxic exposure and thus cancer risk. Epidemiologic studies aimed at finding an association with cancer risk and genetic polymorphisms have been carried out in patient groups afflicted with specific tumours. A recent review of the literature yields a meta-analysis of the multiple available studies (d’Errico et al., 1996), while other reviews such as (Smith et al., 1995; Bartsch and Hietanen, 1996) discuss specific aspects of individual cancers. Studies directed at assessing the impact of nutrition on differently susceptible individuals, although still comparatively rare, are becoming more frequent.

Future biomarker approaches will also consider modulating influences of the diet on biomarker responses reflecting genetic damage in some cases, making the cells more vulnerable to additional toxic compounds; while in other cases leading to less damage, making the cells more tolerant to additional exposure-related factors. Reported examples are higher levels of 5-HO- methyl uracil by high fat diets (Djuric et al., 1991) or modulation of malondialdehyde–DNA adducts by diets with different fatty acid composition (Fang et al., 1997), or reduction of intrinsic oxidative DNA damage by moderate wine consumption (Fenech et al., 1997). Recently, the adequate consumption of carotenoid-containing vegetable juices (Pool-Zobel et al., 1997) or
intake of vitamins as dairy supplements (Duthie et al., 1996) have been shown to reduce oxidative damage of lymphocyte DNA. In contrast, the comparison of vegetarian and non-vegetarian lifestyles did not reveal differences in genetic damage in lymphocytes, detected as micronuclei (Fenech and Rinaldi, 1995; Kim and Mason, 1996). Modulation of metabolising enzymes by dietary influences is expected to affect the impact of hazardous food compounds. Thus, the dietary induction of GSTs may be considered a protective mechanism in situations of exposure to hazardous chemicals that are deactivated by these enzymes.

5.2.2. Repeated exposure assessments

Cohort studies sometimes rely only on baseline exposure assessment whereas follow-up is extended to several years. In regard to change of exposure over time, such as a possible change in diet in younger persons, although not likely, the change could be monitored by means of repeated measurements in a subsample, and hence is not a major problem. Repeated exposure assessments are required for all cohort members if exposure is expected to vary over time. Future studies should provide richer exposure information (e.g. age-specific exposure histories). Cumulative lifetime exposure could be valuable for time-dependent exposure patterns.

5.2.3. Improvements in assessment of organ functions

Exposure studies in animals generally involve the sacrifice of the animal and subsequent dissection and analysis of separate organs. To compensate for these extensive analyses, a standard battery of measurements could be established for studies in humans including function tests for various organs such as liver, kidneys, lungs, heart, etc.

5.2.4. Composite public health measures

Currently, much attention is being focused on the development of “composite public health measures”, which are defined as measures that combine information on mortality, morbidity and reduction of “quality of life”. Assessment of risk can be refined, as the burden of the disease is also taken into account in a weighted manner. Examples are the healthy life expectancy (HLE), the disability adjusted life years (DALY) and the disability adjusted life expectancy (DALE). HLE is defined as the mean number of years persons will live in good health. DALYs are calculated as the sum of years of life lost by death and by disease. The latter component is calculated by multiplying the number of years lived with the disease by a disease-specific weight factor, which is a measure for the seriousness of the disease. DALYs are calculated separately for every disease and summed to obtain the “total burden of disease” (Hoeymans et al., 2000).

The above-mentioned measures, however, are still in the development stage and to date have scarcely been used in epidemiology.

5.2.5. Long-term effects of new developments in the food markets and post-launch monitoring (PLM)

The evaluation of the long-term effects of entirely new substances, or known substances used in much higher concentrations or in a completely different context may require new epidemiologic approaches in exposure measurements and endpoint determination. PLM encompasses the assessment of the amount and pattern of consumption in detail and determination of the nature and degree of expected and unexpected effects after the introduction of new food products. The need for and exact content of a PLM might be determined case-by-case and will depend on the concerns sometimes expressed during the evaluation of a product during the premarket phase. The design of such studies will be prospective and new methods of collecting dietary data as discussed in detail (see Kroes et al., 2002) might be applied [e.g. using combinations of methods using modern technology, such as European Article Number (EAN) codes]. The follow-up for diseases and other endpoints enables examination of intended effects as well as tracing for long-term unexpected effects. For such a follow-up record, linkage to routinely collected outcome data (recorded in registries) should be considered. Such approaches are already successfully applied in pharmacoepidemiology (Herings et al., 1999). For rare and specific adverse effects, such as allergic reactions, a case-control design might be more efficient.

Unexpected effects, rare and unspecific adverse effects can also be monitored by passive surveillance using standardised data collection. If appropriate, this information can be followed up by active studies. In its early days, PLM for aspartame (NutraSweet®) was carried out over 12 years to evaluate and document anecdotal reports from some customers alleging adverse effects from aspartame (Butchko et al., 1994). These reports were not only reviewed by the company but also by epidemiologists at the Centers for Disease Control (CDC) (Centers for Disease Control, 1984; Bradstock et al., 1986) and the US Food and Drug Administration (FDA, 1995; Tollefson, 1988; Tollefson et al., 1988). CDC and FDA concluded that the majority of symptoms reported were mild and common in the general population, and there were no specific clusters of complaints that suggest a causal relationship with aspartame. They further concluded that focused clinical studies would be the best way to address thoroughly the issues raised by the anecdotal reports; the results of this focused research did not show a causal relationship between aspartame and alleged adverse effects (Tschanz, 1996).

Potentials for PLM are strongest for readily identifiable substances (e.g. olestra, phytosterols/stanols in
margarines). PLM is more difficult, but not impossible, for ingredients of more generic nature such as aspartame. However, it seems obvious that long-term monitoring of health effects associated with new developments in the food market should be of overall scientific interest and not restricted to particular products. A broader perspective that addresses the public health concerns regarding the array of new products being introduced regularly and their possible long-term consequences (not observable in usual, small-scale pre-launch testing) would entail systematic longitudinal monitoring of populations. The actual conduct of such studies could take advantage of current infrastructures (where they exist), such as national surveys or panels or large ongoing cohort studies, with the introduction of new questions or measurements as needed. This type of approach avoids the difficulties of a case-by-case effort, and can be undertaken in a more neutral context.

5.3. Conclusions and recommendations

- Beyond the standard concerns about the quality of study designs, epidemiologists must pay particular attention to the presentation of exposure data.
- Validation and calibration studies are very helpful in establishing accurate exposure levels as well as correction of the dose–response association for measurement error.
- Linkage of external databases with substances in foods to dietary data in ongoing epidemiologic studies extends the application of these studies for the purpose of risk assessment in a very cost-effective manner.
- High collinearity between substances occurring naturally in foods limits the possibility to perform risk assessment for those separate substances; however, when risk assessment of mixed exposures is considered, this is one of the strengths of epidemiologic studies as compared to experimental studies.
- Further standardisation and application of Good Epidemiologic Practices to the dietary field is encouraged.
- The infrastructure of ongoing cohort studies can be exploited for risk assessment purposes with relatively low additional costs.
- Biomarkers of exposure can be used to improve or calibrate exposure assessment in new studies.
- Biomarkers of effect sometimes can be used as a surrogate for disease endpoints; there is, however, in general a great need for validated biomarkers that are strongly predictive of disease outcome.
- For new substances or known substances used in much higher concentrations or a different context, use of epidemiologic techniques in post-launch monitoring can provide data on unexpected adverse events.
- Collection of biological samples provides an opportunity to study the effect of genetic susceptibility on risk.
- Development of composite public health measures refines the assessment of risk by considering the burden of the disease.

6. Using evidence from epidemiologic studies in risk assessment

6.1. Individual studies: guidelines to classify epidemiologic studies for potential use in risk assessment

The lack of a systematic methodology has led to evaluation of epidemiologic studies on an ad hoc basis. There is a need for guidelines to classify individual epidemiologic studies of good quality as suitable for use in the different risk assessment phases. A classification of studies for use in the dose–response assessment of carcinogens (but with slight modifications also applicable to other health endpoints) was proposed by Hertz-Picciotto (1995). We present a slightly modified version in this review. In this framework of classification, studies can contribute to the dose–response assessment in three ways:

- **Category 1**: the study can be used to derive a dose–response relationship
- **Category 2**: the study is inadequate to confidently derive a dose–response relationship but can be used as a check on plausibility of an animal-based risk assessment.
- **Category 3**: the study cannot contribute to dose–response assessment, but still can play a role in hazard identification.

For the classification of studies into each of these three categories, a list of four criteria is used, addressing the validity (criteria 1 and 2) and utility of the study (criteria 3 and 4):

1. high overall quality (i.e. major biases in selection, follow-up, etc. can be ruled out);
2. no substantial uncontrolled confounding from other environmental exposures or lifestyle factors;
3. exposures that have been well characterised quantitatively and linked to the individuals in the study, and which are sufficiently variable;
4. evidence for a dose–response relationship between exposure and outcome.
The criteria and categories are summarised in Table 6.\footnote{In the former scheme of Hertz-Picciotto (1995), the first criterion was: a moderate to strong positive association present between the agent in question and the health outcome. However, in the weighted evidence approach as exemplified in evidence-based medicine, a valid study without biases and proper exposure assessment provides information whether or not an association is found.}

Some modification of the above-presented criteria is needed with respect to the evaluation of studies of food components associated with potential beneficial effects.

### 6.2. Combining (epidemiologic) evidence

Epidemiologic studies generally provide data on low doses of exposure, leading to low risks for disease or adverse effects. To enhance statistical power, combining results of individual studies is useful. This is also useful in obtaining the best estimate of an effect, for example a dose–response relationship. In addition, when different exposure ranges have been studied, data of separate studies can be combined into a dose–response relationship across a broader exposure range. Two approaches can be used: meta-analysis of published results and combined analysis by pooling of individual data.

#### 6.2.1. Evidence-based medicine

A new systematic approach of using scientific data was considered with the idea of evidence-based medicine. Most of the research questions currently considered by groups evaluating the literature according to the state of evidence are directed to medical questions and clinical practice. However, their theoretical concept might also be applicable to risk assessment and risk/benefit considerations. The US Preventive Services Task Force, which has been operating since 1984, has been particularly active. They proposed the following level of evidence associated with each study design (Briss et al., 2000) (Table 7).

Recently, more emphasis was directed to the level of evidence for population-based prevention. A new independent panel formed by NIH developed a draft for a Guide to Community Preventive Services, which will be available on the Internet in the year 2001. Although this guide addresses the preventive side of nutrition, their evaluation scheme for epidemiologic studies (Briss et al., 2000) (Table 7).
2000) might be important for quality assurance of existing evidence also for risk assessment. The state of the art regarding a high degree of evidence can be summarised as follows:

- High quality prospective study designs (intervention studies, prospective cohort studies)
- At least two independent studies of high quality which are internally consistent in results.

It is important to know that all considerations in the area of evidence-based medicine are based on evidence generated by human studies. Their framework of generating evidence favors the epidemiologic approach, in particular human intervention studies. Extrapolations from animal models to the human situation are not considered.

### 6.2.2. Meta-analysis of published results

Meta-analysis is a statistical procedure that integrates the results of several independent studies that are considered to be “combinable” (Egger et al., 1997), depending on their quality. Meta-analysis can be preceded by quality scoring based on criteria lists (Moher et al., 1995; Verhagen et al., 2000), then the meta-analysis can be restricted to high-quality studies or weighted according to quality scores (Boyd et al., 1993). When the magnitude of underlying risks is small and the results from individual studies are not conflicting, meta-analysis seems an attractive proposition both in etiological studies and in observational effectiveness research. Originally, meta-analysis was performed on randomised controlled trials, but the use of observational data is increasing. The number of published meta-analyses concerning observational studies in health has increased substantially during the past four decades (678 in 1955–1992, 525 in 1992–1995, and more than 400 in 1996 alone) (Stroup et al., 2000). Meta-analyses of observational studies present particular challenges because of inherent biases and differences in study designs, yet they may provide a tool for helping to understand and quantify sources of variability in results across studies (Stroup et al., 2000). In a workshop held in Atlanta (1997) a checklist was developed summarising recommendations for reporting meta-analyses of observational studies in epidemiology, including background, search strategies, methods, results, discussion, and conclusion (Table 8). Use of the checklist should improve the usefulness of meta-analyses for authors, reviewers, editors, readers and decision-makers (Stroup et al., 2000).

Major problems in meta-analysis include: publication bias, heterogeneity of results among studies, as well as variation in quality of studies, measures of exposure and response, presence of effect-modifiers, and control of confounders.

Heterogeneity, defined as variation among the results of individual trials beyond that expected from chance alone, is an important issue in a meta-analysis (Engels et al., 2000). Heterogeneity may indicate that trials evaluated different interventions or different populations. When heterogeneity is present, it may be inappropriate to combine separate study estimates into a single number, particularly using fixed effects methods that assume a common treatment effect. Random effects methods which can provide an attractive approach to summarising heterogeneous results, and model heterogeneity as a variation of individual study treatment effects around a population average effect, all can be applied to meta-analysis (Greenland, 1987). However, in this case, the population average effect may not be meaningful (Poole and Greenland, 1999) and may or may not be a good estimate from which to extrapolate to other populations, depending on the characteristics of these other populations. Meta-regression can be used to detect reasons for heterogeneity (Greenland, 1998).

### 6.2.3. Pooling of individual data from studies

For low-dose chronic exposures, a careful large-scale pooled analysis is probably the most useful tool for the direct estimation of risk and for testing the adequacy of extrapolations. However, pooling generally is more costly and time-consuming than meta-analysis and often not possible since original data should be accessible. Also, data combined should come from comparable studies (Moolgavkar et al., 1999b). An example of pooling of data is the “Pooling Project” where the primary data of seven prospective cohort studies in four countries, which meet specific criteria with respect to dietary assessment, are being analysed in a standardised manner (Hunter et al., 1996). One of the results published thus far is a linear dose–response association between total alcohol intake and risk of breast cancer, based on 4335 cases in a combined total of 322,647 women under study (Smith-Warner et al., 1998). Owing to the homogeneity of the study results and the large statistical power of the pooled results, it was possible to estimate the effect of drinking alcohol on the risk of breast cancer more precisely: the risk of breast cancer was increased by 9% per 10 g/day increase in alcohol consumption.

A step beyond pooling results from independent studies, are multicentre epidemiologic studies. An example is the EPIC-study on diet and cancer, which is being conducted in nine European countries and 22 centres among over 400,000 participants, using similar methods of data collection, which are nevertheless adapted to the specific cultures of the different populations involved. Calibration of the dietary data by 24-hour recalls with biomarkers is being conducted in a sample of the total cohort to achieve an absolute and comparable estimate of food consumption.
6.2.4. Publication bias and related problems

For an appropriate risk assessment, assessors should ideally have access to all results from studies on the agent of interest, be it human, animal or in vitro studies. In this respect it is important that all good quality studies on the particular topic are published, and publication should not be limited to statistically significant positive outcomes (called publication bias). In reality, publication bias exists. In particular, the smaller human studies with negative or null results will have a lower chance of getting published. Publication generally is dependent on the quality of the study and its presentation, but also on other factors dictating acceptability for publication. Apart from authors, journals should also pay more attention to high-quality studies presenting unexpected effects or weak associations lacking statistical significance. Many journals have little interest in such findings. Although some journals specifically stimulate publication of “null results”, it is clear from the way they do this that they are considered less valuable than studies with positive results (see text box). Several methods have been proposed to detect the existence of publication bias in a meta-analysis. Perhaps the most common is the funnel plot (Light and Pillemer, 1984) and related graphical methods for visually determining the existence of missing studies (Duval and Tweedie, 2000). In a funnel plot (Figs 9 and 10), each dot represents the measured study effect of a particular study.

Table 8
A proposed reporting checklist for authors, editors and reviewers of meta-analyses of observational studies (Stroup et al., 2000)

<table>
<thead>
<tr>
<th>Reporting of background should include:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Problem definition</td>
</tr>
<tr>
<td>• Hypothesis statement</td>
</tr>
<tr>
<td>• Description of study outcome(s)</td>
</tr>
<tr>
<td>• Type of exposure or intervention used</td>
</tr>
<tr>
<td>• Study population</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reporting of search strategy should include:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Qualification of searchers (e.g. librarians and investigators)</td>
</tr>
<tr>
<td>• Search strategy, including time period included in the synthesis and keywords</td>
</tr>
<tr>
<td>• Effort to include all available studies, including contact with authors</td>
</tr>
<tr>
<td>• Databases and registries searches</td>
</tr>
<tr>
<td>• Use of hand-searching (e.g. reference lists of obtained articles)</td>
</tr>
<tr>
<td>• List of citations located and those excluded, including justification</td>
</tr>
<tr>
<td>• Method of addressing articles published in languages other than English</td>
</tr>
<tr>
<td>• Method of handling abstracts and unpublished studies</td>
</tr>
<tr>
<td>• Description of any contact with authors</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reporting of methods should include</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested</td>
</tr>
<tr>
<td>• Rationale for the selection and coding of data (e.g. sound clinical principles or convenience)</td>
</tr>
<tr>
<td>• Documentation of how data were classified and coded (e.g. multiple rates, blinding and inter-rater reliability)</td>
</tr>
<tr>
<td>• Assessment of confounding (e.g. comparability of cases and controls in studies where appropriate)</td>
</tr>
<tr>
<td>• Assessment of study quality, including blinding of quality assessors; stratification or regression in possible predictors of study results</td>
</tr>
<tr>
<td>• Assessment of heterogeneity</td>
</tr>
<tr>
<td>• Description of statistical methods (e.g. complete description of fixed or random effects models, justification of whether the chosen models account for predictors of study results, dose-response models or cumulative meta-analysis) in sufficient detail to be replicated</td>
</tr>
<tr>
<td>• Provision of appropriate tables and graphics</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reporting of results should include:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Graphic summarising individual study estimates and overall estimate</td>
</tr>
<tr>
<td>• Table giving descriptive information for each study included</td>
</tr>
<tr>
<td>• Results of sensitivity testing (e.g. subgroup analysis)</td>
</tr>
<tr>
<td>• Indication of statistical uncertainty of findings</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reporting of discussion should include:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Quantitative assessment of bias (e.g. publication bias)</td>
</tr>
<tr>
<td>• Justification for exclusion (e.g. exclusion of non-English-language citations)</td>
</tr>
<tr>
<td>• Assessment of quality of included studies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reporting of conclusions should include:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Consideration of alternative explanations for observed results</td>
</tr>
<tr>
<td>• Generalisation of the conclusions (i.e. appropriate for the data presented and within the domain of the literature review)</td>
</tr>
<tr>
<td>• Guidelines for future research</td>
</tr>
<tr>
<td>• Disclosure of funding source</td>
</tr>
</tbody>
</table>

414  P. van den Brandt et al. / Food and Chemical Toxicology 40 (2002) 387–424
The larger the size (precision) of a study, the more probable it is that the measured study effect is close to the true association (the dark line in the figure). When all studies would have been published, irrespective of study results, the figure to the left shows study results and sample sizes indicating natural variation. However, if publication bias has taken place and studies with a positive effect are over-represented, the figure to the right will appear. The assumption is that, whether because of editorial policy or author inaction or other reasons, these results (which show no significance or perhaps the reverse effect) are the ones that might not be published. Such suppression will affect the estimate of the combined study outcome. There also exist several quantitative methods that estimate the number of missing studies and, by explicitly modelling the probability of publication, provide estimates of the effects of the missing studies on the overall effect size (Duval and Tweedie, 2000).

In a study by Dickersin et al. (1992) comparing published randomised clinical trials with completed, yet unpublished randomised trials by the same investigators, it was found that 55% of the published, but only 15% of the unpublished studies favored the new therapy being assessed. Other forms of publication bias are language bias (publications in English-language journals tend to have more significant results than papers from the same authors with the same quality in their native languages). Also, it has been shown that sources of funding appear to influence publication. All forms of publication bias have an important impact on literature reviews and certainly on the performance of meta-analyses. Publication bias could be avoided by creation of study registers and advance publication of research designs (Szklo and Nieto, 2000). Several medical journals are considering this. There is no reason to assume that publication bias is less prevalent in toxicological research than in epidemiologic research, also considering the time it takes to conduct a study.

Null results also should also be mentioned in abstracts to prevent hiding of publications during computerised literature research. “Abstract bias” lies in the words chosen for the abstract. Abstract bias is due to the tendency of authors to draw attention especially to positive or statistically significant results, certainly driven by the limited word counts set by the journals. As abstracts are often used to trace publications in computerised databases such as MEDLINE, it is of utmost importance that abstracts cover findings as completely as possible.

Taken from: Instructions of Authors, Cancer Epidemiology, Biomarkers and Prevention (July 2000):

‘Reports of null results in brief’.

“Original reports of null results of important a priori hypothesis tested with sufficient statistical power. These brief reports should be no more than 1 published page in length and must follow the specified format exactly. Please indicate in your covering letter that your submission is for the Null Results category. Note that authors of submissions to the Null Results in Brief category will be permitted to make only minor revisions; submissions that require substantive revisions will not be accepted. In addition, reviewers’ comments may not be provided to authors of rejected manuscripts.”

6.3. Conclusions

Epidemiologic studies of good quality can be classified according to their use for risk assessment using a classification framework.

Combining of evidence from studies of sufficient quality estimates the dose–response association with more precision (statistical stability) and takes into account studies with “null” results. The evidence should be based on the quality of individual studies, not on their outcome.
7. Application areas where epidemiologic studies can contribute to risk assessment (matrix)

A potential role of epidemiology (or more general: human data) exists for all stages of the process of risk assessment of food-related agents. The possibilities for contribution depend strongly on the (type of) food constituent. As mentioned in the general introduction, these have been categorised in this FOSIE project as low molecular weight chemicals, micronutrients and nutritional supplements, macronutrients, whole foods, novel foods and novel food ingredients, food processing (see Appendix A). The quality of the dietary intake measurement is an important determinant of the utility of the epidemiologic study in risk assessment (or risk/benefit assessment). The quality depends on the food/constituent category at hand. Thus, for ingredients that can only be measured poorly in an observational epidemiologic study, the contribution to risk assessment will be limited. The quality of intake data depends, among other factors, on the variability of the concentration within foods and on the availability of information on the levels of the constituent in foods.

Generally, epidemiologic intake data on low molecular weight chemicals such as flavourings, additives and contaminants will be of poor to moderate quality compared to intake data on micro- and macronutrients for which nutrient databases were originally developed. With the currently available data, epidemiology is therefore generally better suited for risk assessment of micro- and macronutrients than low molecular weight chemicals. However, when an effort is made to identify the important food sources of the chemical and concentration levels are measured in these foods, epidemiologic studies can be used in the risk assessment for that chemical [e.g. for the additives BHA and BHT (Botterweck et al., 2000)]. Of course, these epidemiologic data were only generated after introduction of the chemical on the market.

Because the intake measurement essentially uses foods as the directly measured items (as opposed to the indirectly measured nutrients or low molecular weight chemicals), epidemiologic studies can also contribute importantly to risk assessment of whole foods and novel foods. Randomised controlled trials are only considered when a potential benefit is to be expected from a constituent or (novel) food. Therefore, RCTs can be designed to estimate this effect. At the same time, they may also be extremely valuable for assessing risk, as has been shown by the beta-carotene trials in Finland and the USA that were designed because of an expected benefit. RCTs are more feasible for micronutrients (supplements) than for macronutrients or foods because it is easier to use placebos. Although epidemiology has contributed data on the risks associated with consuming processed foods (e.g. salting/pickling and stomach cancer; meat frying and colorectal cancer), the information on food processing, however, is currently rather crude, if available at all.

When valid biomarkers of exposure are available and can be used in population studies, epidemiologic studies employing these biomarkers can also be used in categories for which direct intake measurements are of poor to moderate quality. The earlier mentioned study on the contaminant aflatoxin and liver cancer is such an example (Ross et al., 1992).

Although cohort studies and RCTs could in principle be designed for evaluation of novel foods and novel food ingredients, their applicability will be much further increased when use could be made of short-term effect biomarkers that are truly validated in terms of disease outcome.

For each stage of the risk assessment process, the suitable types of epidemiologic study designs are available for each stage of the risk assessment process and are indicated in Table 9.

7.1. Hazard identification

The importance of epidemiologic data in the process of hazard identification is already well recognised. For some agents, hazards have been identified by epidemiologic studies before they were found in laboratory animals (for example, lung cancer caused by smoking). Hazards can be identified based on all types of epidemiologic studies, observational studies and randomised controlled trials. However, as noted, RCTs will only be performed when no serious adverse effects are expected or when exposure to a prevailing harmful substance can be removed (the oltipraz/aflatoxin example) and preferably when (removal of) exposure can be administered in a blinded way. This will be the case in studies on micronutrients and (pre-launch) with novel foods, food supplements or fortified foods. In new products or ingredients, PLM can be used to monitor the occurrence of unexpected hazards.

7.2. Hazard characterisation (dose–response assessment)

To establish the quantitative relationship between dose and response, epidemiology can make major contributions so long as both dose (exposure) and response (health outcome) are adequately quantified and biases are minimised. Generally, the same epidemiologic study designs can contribute to hazard characterisation as hazard identification, although quality of exposure measurement is more critical here. In hazard identification, a crude (dichotomous) measure may suffice. This limitation is indicated in Table 9 by cohort and case-control entries placed between parentheses (for low molecular weight chemicals and for food processing). RCTs are only useful for hazard characterisation when
more than one dose level is tested; this is often not the case with RCTs that have sufficient statistical power.

Exposure or dose here refers to the amount ingested by the individual. Alternatively, when the dose–response relationship is quantified using indicators such as biomarkers, these biomarker levels subsequently need to be translated back to ingested amounts for further use in risk characterisation. To characterise the relationship between ingested amount and internal (absorbed) dose, human volunteer studies on bioavailability across a range of exposure are extremely important. These should not only consider simple solutions of food constituents, but rather measure bioavailability of the constituents when consumed as part of food.

7.3. Exposure assessment

Exposure assessment provides an evaluation of the distribution of human exposure that is likely to occur. Cohort studies can provide useful data on population dietary exposure, if the study population is more or less representative of the population to which the risk assessment refers. In the same way, dietary data from population controls of case-control studies can be used. These studies are almost always directed at measuring the usual individual intake, rather than measuring days with extreme intakes. Food consumption surveys (without measuring disease outcome) contribute importantly to exposure assessment (see Kroes et al., 2002).

7.4. Risk characterization

It should be noted that in the process of risk characterisation, the first three steps of risk assessment are combined. Therefore, several study designs can be used together and hence this is not noted as such in Table 9.

### Strengths and weaknesses

- **Strengths:** epidemiologic studies have the advantage that they directly contribute data on risk or benefit in humans as the investigated species, and in the full intake range normally encountered by humans or envisaged when ingredient levels (as supplement or in novel foods) are deliberately increased. Because of their human origin, epidemiologic studies require less extrapolation. In addition, they could provide disease risk data on real (mixtures of) exposures on subjects in a full range of susceptibility.

- **Weaknesses:** The quality of exposure information limits the applicability of epidemiologic studies, particularly for low molecular weight chemicals and for food processing. Additionally, for diseases with long latency periods, epidemiologic studies on risks of novel foods may take too long to conduct, unless valid early effect biomarkers are employed.
7.6. Conclusions

Epidemiology can be used in the first three elements of the risk characterisation process, provided that the study has sufficient quality (validity) and does provide the information required and relevant for risk assessment of a dietary exposure. It can be applied for each of the distinguished food ingredient groups. A framework or decision-tree approach should be developed to fit the epidemiologic studies in the risk assessment procedure.

The main limitation is that epidemiology cannot be used and will not be available for screening of novel substances that do not have a history of use in humans. Only after market launch of foods that contain such substances can epidemiology play a major role in collecting more evidence on the safety of these substances.

8. Conclusions and research needs

8.1. Conclusions

The present chapter serves as an introduction to the epidemiologic approach, identifies areas of contribution of epidemiology to the risk assessment procedure, suggests ideas for tailoring the presentation of epidemiologic study results to the risk assessment procedures, and summarises the current status regarding combination and evaluation of epidemiologic studies.

The advantages of epidemiologic data for risk assessment are manifold, but above all it provides information of direct relevance. Epidemiologic data generally study the range of human exposures of interest, and the magnitude of error is likely to be less when human data are used (Hertz-Picciotto, 1995; Calderon, 2000b). Epidemiologic data include the genetic diversity, and variability in exposure and in other endogenous factors inherent in human populations.

While in the past the sparseness of epidemiologic data was a barrier to its use in risk assessment, currently, more high-quality human data are available and suitable for incorporation in the risk assessment process. As risk assessors feel comfortable with the standardised animal studies and extrapolation models, introduction of human data often leads to criticism and exaggeration of flaws in a study, due to misconceptions over the promise and practice of epidemiologic data (Calderon, 2000a). Thus, this present review has tried to put these misconceptions in the right perspective and has also tried to contribute practical solutions to most of the problems encountered when epidemiology is used for risk assessment.

Epidemiologic data will be especially valuable (Calderon, 2000a):

- When good exposure assessment data are available.
- When economic costs of a proposed regulation or action are very high, epidemiology can show whether the costs of regulation justify the health benefits.
- When humans are the most sensitive species for an agent.
- When sensitive human subpopulations are known which have a much higher risk.
- For evaluation of the effect of promulgated regulations, management strategies or policies; epidemiology can document reduction of exposure and disease in the population.
- To provide a sense of perspective to set priorities in the larger context of public health priorities. The role of one particular agent in causing a disease may in some cases be minor in comparison to other risk factors.
- When risks have to be offset against benefits, it is necessary to have human data available instead of relying on extrapolating both risks and benefits.

The availability of a good exposure measurement can be considered as the key factor in the usefulness of epidemiologic studies in the risk assessment process.

Epidemiologists should pay more attention to quality and presentation of exposure measurements. Epidemiologists also need to be made aware of the improvements in analysing and reporting their results, which is required to better enable their use for the purpose of risk assessment.

Based on the above-mentioned conclusions, a more vigorous effort should be made to promote the use of epidemiologic studies in the regular risk assessment process. This use should not be limited only to published studies but it should also stimulate efforts to answer new questions within the framework of ongoing epidemiologic studies. More work, such as the development of a framework and decision trees, is required to study the ways in which epidemiologic studies are best incorporated in the risk assessment process in the food area.

Last, but not least, communication and understanding between toxicologists and epidemiologists has to be improved to achieve an integrated approach to risk assessment in the food safety domain. Multi-disciplinary composition of risk assessment committees should help to improve this integration.

8.2. Research needs

Research needs address two areas of application: those that are needed to facilitate the use and conduct of epidemiologic studies for the purpose of risk assessment and those needed to integrate epidemiology with tox-
iology into the regular risk assessment process. They are ordered by priority.

1. Develop decision trees for dietary risk assessment, in which toxicological and epidemiologic research each have their role. Such decision trees may be incorporated in regulations for food safety evaluations.
2. Work towards a common terminology and approach (e.g., risk models, weight of evidence).
3. Development of short-term effect biomarkers that are highly predictive of risk of disease and that can be applied in human studies on a relatively large scale at low cost, without reducing response rates. The need for such established markers is great both with regard to cancer as well as for many other conditions.
4. Expansion of food composition tables with data on various chemicals in food (additives, contaminants, etc.).
5. Further development and testing of framework/guidelines to select epidemiologic studies suitable for risk assessment for the various categories of chemicals in food (additives, contaminants, nutrients, foods, etc.).
6. Investigate the opportunities for incorporating more probabilistic approaches into the risk characterisation process. The probabilistic approach, which relies on probability distributions for many of the ingredients of risk characterisation such as exposure, the shape of the dose–response curve, toxicokinetic and toxicodynamic parameters, human variability and precision of the data, is already more familiar to epidemiologists than toxicologists, particularly in Europe. The probabilistic approach should provide a more rational and realistic framework for risk characterisation and would facilitate the integration of toxicology and epidemiology into the framework (see also Kroes et al., 2002).
7. Study interactions regarding disease risk between nutrients and non-nutritive compounds taking into account human patterns of consumption.
8. Development and validation of methods to combine risks and benefits. As exposures can be protective for particular disease(s) while enhancing risk of other diseases, further development and use of composite public health measures will provide insight to optimal dose ranges of exposures.
9. Increased use of genetic susceptibility measures on a population basis and studying interaction between genetic susceptibility and dietary exposures regarding disease risk to refine risk assessment for subgroups of the population; may possibly reduce uncertainty factors that are currently used to account for human variability.
10. Study the bioavailability of various compounds in normal food matrix and combine these data with exposure information to better define dose.
11. Further development of PLM using modern food assessment techniques (EAN codes, etc.) and investigating the possibilities for record linkage to routinely collected data in registries.
12. Refinement of methods to calculate lifetime risk from epidemiologic studies conducted in cross-section of population in various exposure groups covering a relatively short part of life.
13. Establishment of the value of applying non-parametric methods to model empirically the association between exposure and disease in epidemiologic studies (e.g., spline regression).
14. Establish an acceptable and standardised set of (already existing) biomarkers for organ function in humans (e.g., liver and kidneys).

Appendix A. Categories of chemicals in food and diet in Europe (Source: FOSIE/ITGG/4/AP01)

- Low molecular weight chemicals (food additives, flavourings, substances used in production of foods, contaminants, pesticides and veterinary residues, natural toxicants)
- Micronutrients and nutritional supplements (vitamins, minerals, miscellaneous)
- Macronutrients (history of food use)
- Whole foods (history of food use)
- Novel foods (from non-GMO sources, no history of use)
- Food processing (traditional processes as cooking smoking drying; irradiation, novel processes)

Appendix B. The role of human volunteer studies in the hazard assessment of chemicals in food

Studies of new food chemicals and components in volunteers are not a substitute for toxicity studies in animals. The use of human volunteers is generally considered only for studies of tolerance, acceptability and palatability, or absorption, metabolism and kinetics, and then only if there are adequate toxicity data from studies in animals or prior use in humans on which to determine a schedule of administration which is likely to be safe for humans. If the design of the study allows data relevant to the determination of the safety of the product in humans to be obtained, this is an acceptable outcome.

It can be anticipated that there will be an increase in human volunteer studies in the future. This will occur as a consequence of the increased marketing of novel foods and new products specifically designed to optimise human bodily function (functional foods or...
nutraceuticals). Examples include: bacterial–culture-enriched dairy products, claimed to provide benefits by altering the microflora of the gastrointestinal tract; products containing phytosterol esters, claimed to reduce the concentrations of low density lipoprotein-cholesterol in blood; or products containing phytostrogens, claimed to be beneficial for post-menopausal women. The increasing fortification of foods with vitamins and essential minerals has also introduced a need for data that can be used for the assessment of upper safe levels of dietary intakes of these micronutrients.

In some situations it may be desirable to carry out studies with individuals for whom the food component is particularly designed, or in whom consumption is likely to be highest. For example, the influence of artificial sweeteners on glucose tolerance parameters in diabetics, or the possible influence of a common genetic polymorphism on metabolism, can be investigated in human volunteer trials.

Another important group of food components is the naturally occurring toxicants, such as the inherent plant toxins and mycotoxins. Some of these may have an impact on the safety assessment of novel foods. They are already present in the diets of many populations and the background levels present in the body may offer the opportunity to investigate their biokinetics, metabolism and tissue interactions, which may assist in the prediction of hazard and estimation of risk.

Volunteer studies are an essential part of pharmaceutical research, but to date their use in the area of non-pharmaceutical products has been limited. They are especially relevant to the study of the toxicokinetics of a compound. Comparison of such data with those from studies in laboratory animals can increase the accuracy in extrapolation from animals to humans. Such data may also reveal whether humans are likely to be more susceptible than any of the animal species tested. Moreover, some adverse effects, such as diplopia, dizziness, headache and nausea or inter-subject variability can only be evaluated in human tolerance studies. Similarly, establishing whether a new product has the expected beneficial effects in man ultimately requires human volunteer trials. For claim support and safety assessment, a (double-blind) randomised controlled design is the gold standard for such studies.

Human volunteer studies are also very important for biomarker development and validation, in particular of biomarkers of exposure (see section 5).

In order to calculate toxicokinetic parameters, it is important to prepare an adequate sampling schedule. A pilot study or a provisional physiologically based pharmacokinetic (PBPK) model, if available, may be very helpful in setting up the sampling strategy. A low quantification limit in plasma or other body fluids is important to ensure that the kinetics of a compound can be described reliably (Meulenbelt et al., 1998).

Radiolabelled compounds can be used in volunteer studies. However, the use of these will normally require special approval. For radiolabelled compounds the use of accelerator mass spectrometry should be considered, since this technique can reduce the administered radioactive dose to a level below that requiring specific regulatory approval in, for example, the UK (Barker and Garner, 1999).

Obviously, studies of teratogenicity and carcinogenicity and the use of clearly toxic doses are not acceptable in volunteer studies. Similarly, in healthy volunteer studies, although fluids such as saliva, blood and urine may be collected for analysis, tissues cannot be collected for research.

The evaluation of longer-term exposure is difficult in humans. In a human study the results will be influenced by unknown factors, since it will be difficult to keep the circumstances constant for a long period and compliance with the protocol may be compromised.

For some food components, such as vitamins and essential nutrients, the margins between intakes conferring benefits and those that might give rise to adverse effects may be quite small. This aspect needs to be borne in mind when designing human studies, for example by including where possible early biomarkers of exposure/effects. Although studies designed to characterise the beneficial effects of a substance may be unsuitable for the identification of the possible hazards, in the past, information obtained from human volunteer studies designed with only beneficial aspects in mind has often been very limited, with opportunities to assess other aspects relevant to safety, such as tolerance and toxicokinetics, being missed. Moreover, the use of tolerance studies as a final step between full toxicological testing and marketing has been surprisingly infrequent, as has the use of post-marketing surveillance and monitoring after product launches. It can be anticipated that, increasingly, approaches such as these will be used to enhance safety evaluations and to provide reassurance about novel foods.

Appendix C. Bias, confounding, and effect modification

As observational studies are not controlled experiments with a limited exposure to one substance, risk estimates can be influenced by various intervening factors. In the sections on study designs, several of them have been mentioned. Illustrated below is a compact overview of these factors, identified as bias, confounding and effect modification.

A.1. Bias

Bias is the result of a systematic error in either the design or conduct of a study (Szklo and Nieto, 2000).
The error can be made during the selection of study participants (selection bias) or during the measurement of exposure or effect (information bias). Both types of bias may affect retrospective case-control studies more than cohort studies. When selection bias is present, individuals have different probabilities to become included in the study, depending on the exposure and health outcome (Fig. 11) (Szklo and Nieto, 2000).

Information bias results from a systematic tendency for individuals selected for inclusion in the study to be erroneously placed in different exposure/outcome categories, thus leading to misclassification (Fig. 12). Information bias can be due to the ability to recall past exposure is dependent on the case-control (disease) status (recall bias). Objective markers of exposure or susceptibility are less prone to bias than direct responses from study subjects. Certain genetic markers, such as genetic polymorphisms, for example, constitute “exposures” that are not time dependent and can be measured after the disease has occurred, thus being less suscep-

tible to bias (assuming that the genetic marker is not related to survival). Interviewer bias is a form of information bias, which occurs when data collection in a case-control study is not masked with regard to the disease status of study participants. Observer bias (outcome identification bias) occurs when the ascertainment of outcome is not independent of the knowledge of the exposure status. Finally, respondent bias occurs when information on the outcome (in a cohort study) is obtained by participant response. Whenever possible, it should be confirmed by more objective means, such as a hospital chart view.

In general, due to bias the observed result will tend to be different from the true result. Bias generally can be avoided by using an appropriate study design and valid and reliable methods for data collection.

A.2. Confounding

Generally, a confounder is causally associated with the outcome and non-causally or causally associated with the exposure but is not an intermediate variable in the causal pathway between exposure and outcome (Szklo and Nieto, 2000). Confounding refers to a situation in which this variable is responsible for the entirety or part of the statistical association between the exposure and the outcome (Szklo and Nieto, 2000). This may lead to the appearance or strengthening of an association not due to a direct causal effect or in the apparent absence or weakening of a true association. The difference compared to bias is that a confounded association, though not causal, is real (Szklo and Nieto, 2000).

Identification of potential confounders is usually based on a priori knowledge of the dual association of the possible confounder with the exposure and the outcome. Presence of confounding can be verified by: checking whether associations with exposure and outcome exists, by studying the exposure/outcome relationship on different strata of the confounder (see “stratified analysis”), or by checking changes in associations once adjustment has taken place.

Modern statistical procedures such as multivariate analysis enable epidemiologists to deal with (adjust/ control for) confounders (Hertz-Picciotto, 1995). The basic idea underlying adjustment is to use some statistical model to estimate what the association between the exposure and the outcome would be, given a constant value or level of the suspected confounding variable(s) (Szklo and Nieto, 2000). Collinearity is used for the situation where the correlation between a confounder and the exposure of interest is so strong that adjustment becomes difficult, if not impossible (Szklo and Nieto, 2000). Residual confounding arises when a confounding factor cannot be measured with sufficient precision.

Although confounding has received much attention in the development of epidemiologic methods and has
been noted to be a major limitation to the use of epidemiologic studies in safety evaluations, it should be emphasized that confounding can often be corrected for in properly conducted studies, and that confounding often accounts for errors of only 10–50%, but rarely more than 2–3-fold (Hertz-Picciotto, 1995). This is in contrast to the large errors that can be made when interspecies extrapolation or high-to-low dose extrapolation has to be performed because of lack of human data.

A.3. Effect modification

Effect modification is present if the magnitude of the association between exposure and disease varies across the level of another variable (Szklo and Nieto, 2000). In the presence of effect modification, any extrapolation to another population must take into account the distribution of the effect-modifying factor. If the distribution differs from that of the source study population from which the potency is estimated, then stratum-specific potencies are required. An example of this approach comes from a risk assessment for ambient airborne arsenic. The potency for lung cancer at different levels of smoking was calculated due to the strong synergism found across a variety of occupational studies (Hertz-Picciotto et al., 1992). The data were presented for the state of California showing each potency, which enabled risk assessors for other populations with different smoking prevalences to utilise these data (Hertz-Picciotto et al., 1990).

References


