Biomarkers in cancer epidemiology: an integrative approach

Paolo Boffetta*

International Prevention Research Institute, 95 cours Lafayette, 69006 Lyon, France

*To whom correspondence should be addressed. Tel: +33 658386724; Fax: +33 472387126; Email: paolo.boffetta@i-pri.org

There are different reasons for the increase in the use of biomarkers in cancer epidemiology which is as follows: (i) the fact that the identification of new carcinogens, characterized by complex exposure circumstances and weak effects, has become increasingly difficult with traditional epidemiological approaches; (ii) the increasing understanding of mechanisms of carcinogenesis and (iii) technical developments in molecular biology and genetics. While a distinction is made between biomarkers of exposure, intermediate events, disease, outcome and susceptibility, their integration in a unique conceptual model is needed. The use of exposure biomarkers in cancer epidemiology aims at measuring the biologically relevant exposure more validly and precisely. In some instances, there is an obvious improvement in using an exposure biomarker, as in the case of urinary markers of aflatoxin and tobacco-specific nitrosamines. Intermediate (effect) biomarkers measure early—in general non-persistent—biological events that take place in the continuum between exposure and cancer development. These include cellular or tissue toxicity, chromosomal alterations, changes in DNA, RNA and protein expression and alterations in functions relevant to carcinogenesis (e.g. DNA repair, immunological response, etc.). The analysis of acquired *TP53 mutations is an example of the potentially important. Biomarkers should be validated and consideration of sources of bias and confounding in molecular epidemiology studies should be no less stringent than in other types of epidemiological studies. The overarching goal is the integration of different types of biomarkers to derive risk and outcome profiles for healthy individuals as well as patients.

Introduction

During the last decade, the use of biomarkers in cancer epidemiology has greatly increased. Several reasons explain this expansion. The identification of new carcinogens, characterized by complex exposure circumstances and weak effects, has become increasingly difficult with traditional epidemiological approaches. In parallel, increasing knowledge of mechanisms of carcinogenesis led to the proposal of models involving genetic and epigenetic events, as well as cellular and histological alterations. Furthermore, developments in molecular biology and genetics, based on the increasing throughput of automatic analytical equipments, allow the large-scale application of assays that would otherwise be very resource intensive. A systematic review of the contribution of known risk factors to the burden of cancer in France recently concluded that established carcinogens explain only a relatively small proportion of cancers (Table I), in particular among non-smokers, and stressed the limitations of current etiological understanding of carcinogenesis in humans (1). These conclusions are applicable to other high-income countries. The French project identified two important challenges in studies of suspected carcinogens: the complexity of exposure circumstances and the low risk entailed by these exposures.

Research on the role of diet as cause of cancer is a good illustration of these methodological issues. A 1997 review of the evidence linking dietary factors to cancer concluded that low intake of fruits and vegetables was associated with an increased risk of several types of cancer, notably of the digestive tract (2). These evaluations were updated 10 years later (3): on that occasion, the judgment was that the strength of the evidence of a cancer preventive effect of fruits and vegetables had weakened, as illustrated by Table II. The early evidence was derived primarily from case–control studies, which suffer from the retrospective nature of dietary assessment and, in many instances, by small sample size. During the last decade, a number of large-scale prospective investigations have been reported both in individual and in pooled analyses: in most cases, these new results failed to confirm the protective effect of fruits and vegetable. In addition to sample size and—in case–control studies—bias derived from retrospective design, dietary exposure misclassification remains the main challenge of future studies of diet and cancer (4).

The field of environmental carcinogenesis also illustrates the limitations of current epidemiological understanding of cancer. Only a few environmental pollutants have been definitely associated with human cancer (using the classification into Group 1 of the International Agency for Research on Cancer Monograph program (5) as criterion), as shown in Table III. The identification of an environmental pollutant as human carcinogen has been relatively easy when subgroups of the population at very high exposure were available for investigation [e.g. women in some regions of China exposed to high level of indoor air pollution from burning of coal for heating and cooking (8)] or when the evidence was supported by studies of high-dose occupational exposure circumstances, as in case of asbestos (9). Involuntary smoking represents one of the few environmental carcinogens for which conclusive evidence of an effect—albeit weak—has been obtained from epidemiological studies, with support of other lines of evidence (Table IV).

These challenges (etiological complexity and small magnitude of effects) require an increased attention to protection from random error and systematic error in cancer epidemiological studies. While the former type of error can be addressed by approaches based on consortia and networks (20), the application of biomarkers is often invoked for the latter. This review aims at critically evaluating the contribution of biomarkers, when applied to epidemiology, in particular to the identification of environmental and genetic determinants of cancer risk.

A distinction is typically made between markers of exposure, effects (intermediate events, disease), outcome and susceptibility. This distinction, however, is somewhat arbitrary, and any classification reflects the current understanding of a complex biological phenomenon such as carcinogenesis and the ability to measure events that are considered relevant to it. In fact, the integration of biomarkers, based on the increased understanding of the process of human carcinogenesis, and the development of relevant and valid biomarkers represents the main challenge to molecular cancer epidemiology.

This review is intended to provide a critical overview of the main types of biomarkers used in molecular epidemiology.

Biomarkers of exposure, and their application to cancer epidemiology

One reason for using biomarkers in cancer epidemiology is to measure the biologically relevant exposure more validly and precisely. In some instances, there is a clear improvement in using an exposure biomarker, but in other cases, an improvement in validity and precision of the measurement of the relevant agent is not straightforward. The performance of exposure biomarkers should be compared with that of other exposure assessment methods, such as medical records, questionnaires and environmental monitoring. Main concerns are the relevance of the biomarker to the exposure of interest, its specificity (e.g. chemicals often share common metabolites) and the characteristics of the assay, including sensitivity, kinetics, source of variability and effect modifiers. Most biomarker-based studies, of both prospective and retrospective design, rely on a single biological...
sample. This represents a drawback as compared with traditional exposure assessment based on, for example, questionnaires or interviews. With such approaches, historical reconstruction of variations in exposure is—at least in some cases—feasible.

One of the few examples of a validated biomarker of exposure that has substantially contributed to the elucidation of the environmental causes of human cancer is that of urinary adducts formed by aflatoxin (11). Its uniqueness can likely be explained by a combination of factors: (i) the strong potency of aflatoxin as human carcinogen, (ii) the availability of a relatively specific and sensitive biomarker highly correlated with biologically effective dose and (iii) the availability in several populations of individuals with very different exposure levels and patterns. When these requirements are not met, as it is the case of most other dietary factors, the contribution of exposure biomarkers to etiological research has been modest if not counterproductive, as in the case of β-carotene (21). There are, however, examples of exposure markers, which have contributed to a better understanding of human cancer as in the case of tobacco smoking.

In a few prospective studies, serum level of cotinine, the main metabolite of nicotine, has been analyzed with respect to subjective risk of lung cancer (22,23). These studies not only have replicated the strong association between tobacco smoking, measured by this biomarker of exposure, and lung cancer risk but also have provided evidence on some characteristics of tobacco-related lung carcinogenesis, which were not easily identifiable in traditional epidemiological studies. Figure 1 illustrates two of these aspects, based on the results of a prospective study of serum cotinine level and lung cancer risk (22). These results provided no evidence for a plateau of excess risk at high doses of smoking, which was observed in questionnaire-based studies and was interpreted in terms of saturation of metabolic pathways (24). Furthermore, these results provide evidence for a comparable carcinogenic effect of tobacco smoking in men and women, contrary again to the conclusions of some questionnaire-based studies (25).

4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol is the metabolite of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butane, a tobacco-specific nitrosamine. Urinary NNAL is a biomarker of tobacco smoking which has recently been shown to quantitatively predict the risk of lung cancer (23) (Figure 2). More interestingly, the level of this biomarker has been found to be lower in Chinese from Shanghai than in Chinese from Singapore (Table V), probably reflecting the content in tobacco-specific nitrosamines in the smoke from cigarettes sold in the two countries (26). This finding might contribute to explain the lower relative risk of lung cancer observed in Chinese smokers as compared with smokers from other countries (27).

Exposures relevant to cancer epidemiology are typically time-related variables. Moreover, both carcinogenesis models and empirical evidence strongly point toward the importance of induction and latency periods in cancer occurrence, the need to assess both duration and intensity of exposure and the need to assess changes in disease risk after cessation of exposure. The goal of biomarker-based exposure assessment remains therefore the reconstruction of a full history of exposure during relevant time period.

### Table I. Proportion of cancer deaths attributable to known carcinogens by gender and smoking status, France, 2000 (1)

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>Men</th>
<th></th>
<th></th>
<th>Women</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Smokers</td>
<td>Non-smokers</td>
<td>AP%</td>
<td>Smokers</td>
<td>Non-smokers</td>
<td>AP%</td>
</tr>
<tr>
<td>Tobacco</td>
<td>39.7</td>
<td>—</td>
<td>19.3</td>
<td>—</td>
<td>—</td>
<td>19.3</td>
</tr>
<tr>
<td>Alcohol</td>
<td>10.0</td>
<td>6.7</td>
<td>2.9</td>
<td>3.0</td>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td>Infectious agents</td>
<td>3.1</td>
<td>3.0</td>
<td>4.8</td>
<td>3.9</td>
<td></td>
<td>3.9</td>
</tr>
<tr>
<td>Occupation</td>
<td>4.0</td>
<td>1.9</td>
<td>0.7</td>
<td>0.3</td>
<td></td>
<td>0.3</td>
</tr>
<tr>
<td>Overweight and obesity</td>
<td>1.1</td>
<td>1.4</td>
<td>2.1</td>
<td>2.5</td>
<td></td>
<td>2.5</td>
</tr>
<tr>
<td>Lack of physical activity</td>
<td>0.4</td>
<td>0.7</td>
<td>2.8</td>
<td>3.5</td>
<td></td>
<td>3.5</td>
</tr>
<tr>
<td>Exogenous hormones</td>
<td>—</td>
<td>—</td>
<td>1.9</td>
<td>2.4</td>
<td></td>
<td>2.4</td>
</tr>
<tr>
<td>Ultraviolet light</td>
<td>0.5</td>
<td>0.9</td>
<td>0.7</td>
<td>0.9</td>
<td></td>
<td>0.9</td>
</tr>
<tr>
<td>Pollutants</td>
<td>0.1</td>
<td>0.05</td>
<td>0.5</td>
<td>0.1</td>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td>All the above&lt;sup&gt;a&lt;/sup&gt;</td>
<td>50.6</td>
<td>14.0</td>
<td>31.8</td>
<td>15.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AP%, attributable proportion percent.

<sup>a</sup>Because of overlap in exposure, the totals are smaller than the sum of the contributions of each risk factor (see (1) for details).

### Table II. Evaluations of World Cancer Research Fund/American Institute for Cancer Research 1997 and 2007 reports on the evidence of a protective effect of high intake of vegetables and fruits on specific cancers (2,3)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Convincing</td>
<td>Mouth, esophagus, lung, stomach, colon, rectum</td>
<td>—</td>
<td>Mouth, esophagus, lung, stomach, colon, rectum</td>
<td>—</td>
</tr>
<tr>
<td>Probable</td>
<td>Larynx, pancreas, breast, bladder</td>
<td>—</td>
<td>Larynx, pancreas, breast, bladder</td>
<td>—</td>
</tr>
<tr>
<td>Possible</td>
<td>Liver, cervix, ovary, endometrium, prostate, kidney, thyroid</td>
<td>Lung, colon, rectum, ovary, endometrium</td>
<td>Cervix, ovary, endometrium, thyroid</td>
<td>Pancreas, liver, colon, rectum</td>
</tr>
</tbody>
</table>
Although based on a relatively small number of cases, the pattern of mutations in non-smokers is very different from that in smokers, while it closely resembles the pattern found in non-tobacco-related cancers. Studies of tobacco carcinogens, and specifically polycyclic aromatic hydrocarbons such as benzo[\(a\)]pyrene in determining G to T transversions at critical hotspots of \(TP53\), provide a link between biological and epidemiological findings.

In other cases, however, biomarkers of effect have failed to provide additional evidence of a role of specific alterations in etiologically defined carcinogenic pathways, beyond merely indicating the effect of the environmental exposure. This is notably the case of patterns of promoter methylation leading to activation or suppression of genes involved in carcinogenesis. For example, promoter hypermethylation in \(p16\) has been associated with tobacco smoking (29), and recurrence...
of early stage lung cancer (30), although not with risk of developing the disease. In other instances, however, molecular characterization of cancer has led to the identification of subgroups of etiologically relevant cancers, as in the case of colon cancers classified according to microsatellite instability (31).

As in the case of exposure markers, the use of biomarkers to measure the outcome of an epidemiological study (typically cancer) has the aim to increase the validity of the measurement, that is to increase the specificity and the sensitivity in the definition of the outcome. For example, microarray-based techniques to measure the expression of a large number of genes have led to the discovery that breast cancers indistinguishable according to traditional histological classifications may show profoundly different patterns of genetic expression. Alterations (e.g., mutations, deletions, epigenetic modifications) in genes with a role in carcinogenesis or a characteristic cytogenetic alteration, rather than the tumor itself, might become the outcome of a molecular epidemiological study. Studies based on biomarkers of disease are best suited in a prospective design since the identification of early events relevant to carcinogenesis would hopefully impinge on preventive strategies. Disease biomarkers can also be used in retrospective designs, in which only diseased individuals are enrolled. In such case-only studies, comparisons are made among subgroups of cases with differences in the profile of genetic mutations (or other molecular characteristics). The time co-ordinates of early effect markers are crucial for their application in molecular epidemiology. While adequate knowledge of the natural history is lacking for most human neoplasms, models of carcinogenesis developed for various tumors provide a framework for the application of effect markers.

**Biomarkers of susceptibility**

Biomarkers of susceptibility and risk contribute to the identification of high-risk subgroups of the population, independent on whether they are associated with exposure or they are involved in a defined pathway or mechanism. They can, however, contribute to the understanding of etiology in terms of identification of exposure and mechanisms. In recent years, the application of genome-wide analyses of genetic variants in large-scale association studies has led to the identification of a relatively large number of variants conferring an increased (or decreased) risk of cancer (32). A complete review of susceptibility markers is outside the scope of this review, rather, the case of lung cancer, for which three susceptibility loci have been identified at 15q25 and 5p15 (33–38), is chosen as example. These findings have brought additional information on lung carcinogenesis.

The 15q25 region contains three cholinergic nicotine receptor genes (CHRNA3, CHRNA5 and CHRNA4), encoding nicotine receptors in neuronal and other tissues. Six studies have so far reported very consistent results for different polymorphisms on this locus (Table VI) (33–35,39–41). The variants are clearly linked to tobacco dependence and have been associated with other tobacco-related conditions, including peripheral arterial disease (33), chronic obstructive pulmonary disease (42) and head and neck cancer (34). Establishing whether the association between 15q25 variants and smoking behavior, shown in several of the association studies (Table VI) and in other populations (43), fully explains their strong association with lung cancer (about an 80% increase in risk for those who inherit two risk alleles) or mechanism. They can, however, contribute to the understanding of pathways underlying lung carcinogenesis in humans. Association studies of never-smokers would help to clarify this issue, and the current evidence is not fully consistent (Table VI).

A second locus identified in genome-wide association studies of lung cancer is at 5p15.33 and contains two interesting genes, the telomerase reverse transcriptase gene, TERT. Two different variants (rs402710 and rs2736100) that are not strongly associated with each other were reported to be independently associated with risk of lung cancer, and in particular with lung adenocarcinoma (36–38). A large proportion of human cancers, including lung cancer, show telomerase activity, indicating that regeneration of telomeres are an important step in human carcinogenesis (44). Activation of the TERT promoter leads to synthesis of the TERT protein and resulting telomerase.

### Table V. Urinary tobacco biomarkers in lung cancer cases from Shanghai and Singapore (23)

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Shanghai (N = 155)</th>
<th>Singapore (N = 91)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cotinine (ng/mg creatinine)</td>
<td>3033</td>
<td>2873</td>
</tr>
<tr>
<td>NNAL (pmol/mg creatinine)</td>
<td>0.23</td>
<td>0.89</td>
</tr>
</tbody>
</table>

NNAL, 4-(methylnitrosamo)-1-(3-pyridyl)-1-butanol.

### Fig. 3. Pattern of TP53 mutations in lung cancer (by smoking status) and non-tobacco-related cancers—analysis of the International Agency for Research on Cancer p53 database (derived from 28). (A) Lung cancer, excluding non-smokers; (B) lung cancer, non-smokers and (C) cancers other than lung and liver cancer.

### Table VI. Studies on 15q25 polymorphism and risk of lung cancer

<table>
<thead>
<tr>
<th>Reference Country</th>
<th>(33) Iceland, Netherlands, Spain</th>
<th>(34) Europe, North America</th>
<th>(35) USA, UK</th>
<th>(39) Japan</th>
<th>(40) China</th>
<th>(41) USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>N cases</td>
<td>1024</td>
<td>4435</td>
<td>2013</td>
<td>1250</td>
<td>3565</td>
<td>1058</td>
</tr>
<tr>
<td>Polymorphism</td>
<td>rs1051730</td>
<td>rs8034191</td>
<td>rs8034191</td>
<td>rs169690968</td>
<td>rs6495309</td>
<td>rs1051730</td>
</tr>
<tr>
<td>RR (95% CI)</td>
<td>1.31 (1.12, 1.44)</td>
<td>1.30 (1.23, 1.37)</td>
<td>1.32 (1.24, 1.41)</td>
<td>2.2 (1.5, 3.4)</td>
<td>1.43 (1.27, 1.61)</td>
<td>1.59 (1.16, 2.19)</td>
</tr>
<tr>
<td>Association with smoking behavior</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>NA</td>
</tr>
<tr>
<td>Effect in never smokers</td>
<td>NA</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>NA</td>
</tr>
</tbody>
</table>

CI, confidence interval; NA, not available; RR, relative risk per allele.
Molecular epidemiology investigations should be considered within the framework of epidemiological studies in general. Epidemiology aims at identifying determinants of cancer and quantifying their role, while taking into account sources of random and systematic error (bias and confounding), as well as factors that modify the effect of the determinants of interest. To a large extent, molecular epidemiological studies fit into the same framework: they represent epidemiological studies in which risk factors, outcomes, confounders or effect modifiers are measured with biomarkers.

Our understanding of the causes of human cancer remains relatively limited, and the main challenges to epidemiological investigations are represented by the need to study weak associations and complex mixtures. The application of biomarkers to molecular epidemiological studies has in some instances contributed to the elucidation of causes and mechanisms of human cancer. The development of biomarkers that would address several components of the carcinogenic process would represent an important additional contribution of molecular epidemiology. Figure 5 illustrates two such examples with respect to lung cancer. The role of the genetic variants in 15q25 in terms of both determinant of exposure (by modifying nicotine dependence) and determinant of risk (by acting on pathways of lung carcinogenesis independent from smoking) has been discussed above. In addition to its role as determinant of lung cancer risk, DNA repair capacity has been shown to correlate with response to chemotherapy (50). Furthermore, the integration of several types of biomarkers into risk models (51) represents an additional example of multilevel approach in molecular epidemiology.

If biomarkers are to offer new opportunities to overcome some of the limitations of epidemiology, then their added value over traditional approaches should be systematically assessed. Biomarkers should be validated and consideration of sources of bias and confounding in molecular epidemiology studies should be no less stringent than in other types of epidemiological studies. Similarly, other aspects of the study such as determination of required sample size, statistical analysis, reporting and interpretation of results should be approached with methodological rigor. The overarching goal is the integration of different types of biomarkers to derive risk and outcome profiles for healthy individuals as well as patients.

Acknowledgements

Conflict of Interest Statement: None declared.

References


Received October 7, 2009; revised October 24, 2009; accepted October 27, 2009.