

Use of Epidemiological Data for Cancer Risk Assessment: Approaches, Concepts, and Issues

Vladimír Bencko*

Institute of Hygiene and Epidemiology, First Faculty of Medicine, Charles University in Prague, Czech Republic

Abstract: Traditional approaches and study designs in cancer epidemiology have not been very successful identifying and evaluating potential risks and/or protective factors associated with disease. Failure is often due to small study sample size and inadequate exposure information. In this paper, we discuss issues and approaches relevant to these two challenges.

Multicenter study design is proposed as a way to increase study size and to mitigate criticism of meta-analysis of independent studies. Individual studies from a multicenter study can be either pooled using original data, or combined by meta analysis of summarized results. A multicenter study of large cohort or case-control studies also offers an exciting opportunity to study the contribution of epigenetic events that may be associated with life-style and environmental risk factors for human health. Optimizing methods for exposure assessment and reduction of exposure misclassification represent difficult components in epidemiological studies. Biomarkers present a potentially useful approach for improving exposure estimates. An example is provided to demonstrate how biomarkers of exposures can provide valuable information in addition to traditional exposure measurements in epidemiological studies.

Finally, it is argued that risk assessment and the precautionary principle should not be viewed as conflicting paradigms but, rather, as a complementary approach according to the data available. Together they may be used for developing appropriate policies to address risks posed by exposure to carcinogens and a wide spectrum of other health hazards.

Keywords: Epidemiological data, risk assessment, precautionary principle, carcinogenicity, multicenter studies.

INTRODUCTION

The field of epidemiology has reached a crucial point with challenges and opportunities. On one hand, it seems that most of the major occupational carcinogens have already been identified. Many chemicals classified as carcinogens by the International Agency for Research on Cancer (IARC) were first evaluated in the workplace. In the last decade, occupational exposure to known human carcinogens has diminished in many countries as awareness of their hazards has increased [2, 3]. On the other hand, we are still confronted with a long list of substances for which epidemiological data are lacking or inconclusive. The estimated number of chemicals in commerce ranges from tens of thousands to over 140,000 [1]; for most of them, relevant toxicological information is needed to set regulatory standards [2, 3].

We are now at an important crossroad; advances in the interrelated disciplines on which health risk assessment depends hold promise for comprehensive understanding of the influence of environmental stressors on human health. The last decade has been marked with major developments in the field of cancer risk assessment. There have been remarkable advances in the broad area of cancer epidemiology. This includes research not only on human exposures to major cancer risk factors in environmental and

occupational settings, but also on lifestyle and nutrition related risks. Traditional approaches and study design in cancer epidemiology have not been successful in identifying and evaluating these potential risk and/or protective factors. Two main reasons for this failure are often due to insufficient study size, and inadequate exposure assessment. In this paper, we discuss issues and approaches relevant to these two challenges, and the new opportunity of using emerging genomics information in epidemiology studies.

INCREASING STUDY SIZE

An important characteristic of research in the last decade is the increasing number of collaborative studies involving various countries, and as a consequence, sample size is greatly increased. By increasing the sample size, the power of a study to identify significant associations between exposures and a disease endpoint is enhanced. For example, considering that lifetime prevalence of occupation-related exposures in the population is low (typically below 5 or 10%), and the associated risk can be small as well (e.g. relative risk of 2). The study sample size for a community-based study must be large in order to identify statistically significant associations. This is even more crucial if the exposure or disease outcome is misclassified in a study population. A sample size of at least 1000 cases and controls has been recommended for a community-based case-control study on cancer. However, it is often the case that one center or one country can not provide such numbers within a reasonable amount of time. Therefore, multicenter studies are an obvious solution. Besides increased power, multicenter studies can provide additional advantages that include greater exposure variation in the study population

*Address correspondence to this author at the Institute of Hygiene and Epidemiology, 1st Faculty of Medicine, Charles University in Prague, Studnickova 7, CZ 12800 Praha 2, Czech Republic;
Tel/Fax: +420 224919967; E-mail: vladimir.bencko@lf1.cuni.cz

which is useful for dose-response analysis, and an opportunity to study differences in exposure and disease patterns between countries.

Multicenter studies can be created in two different ways. Collaboration between centers can be initiated after completion of each individual study (retrospectively planned multicenter studies), or before the studies have actually started (prospectively planned multicenter studies). Meta- and pooled analyses are examples of retrospectively planned multi-center studies, and have been performed in many areas of epidemiology. Prospectively planned multicenter studies have only recently become possible since international organizations and institutions such as the European Commission started to offer funding for these costly operations. The advantage these studies offer is an identical protocol for data collection at each center involved and avoid loss of information at the stage of data pooling. From the Central and East European Countries (CEEC) perspective, an important example of this approach is the CEEC Multicenter Lung Cancer Study that was organized by IARC/WHO/Lyon, France, about 10 years ago, and was supported by EC Inco Copernicus FP4. It represented collaboration of centers in the Czech Republic, Hungary, Poland, Rumania, Russian Federation, and Slovakia [4-7]. Later, the same organizational principle was applied in the Kidney Cancer Study supported by NCI/Bethesda, USA [8, 9].

A positive finding and great advantage of multicenter studies is that the potentially serious problems associated with the use of meta-analysis in observational studies can now be substantially mitigated. The merit of applying meta-analysis to observational studies has been questioned and controversial [10, 11] because errors and biases can be easily introduced when studies with different designs, methods, and population characteristics are combined. It would not be a problem for pooled or meta analysis if the study design is comparable, and method for data collection is coordinated prior to the initiation of the study among different centers. There are many factors that make meta-analysis of individual studies less appropriate. Consider, for example, meta-analysis of odd ratios or relative risk estimates that require auxiliary information such as age, sex, smoking status, and sample size. If all auxiliary variables are homogenous across studies, it would be appropriate to conduct meta-analyses of these studies [12, 13]. In reality, these conditions have never been satisfied. Therefore, the multicenter study design offers an attractive alternative to the traditional single-center epidemiological study.

IMPROVING EXPOSURE ASSESSMENT

Optimizing methods for exposure assessment and reducing exposure misclassification represent the most difficult components of epidemiological studies. A perfect exposure assessment for long latency diseases such as cancer would consist of quantitative measurements of internal dose, over the whole exposure period, for each subject in the study. This is a utopia unlikely to be attained in the near future. The availability of group-based external exposure measurements at a few points in time is already a luxurious proposition for epidemiologists. More often one has to fall back on subjective methods of exposure assessment. The possibilities for exposure assessment largely depend on the design of the epidemiological study, with community-based

and industry-based studies imposing their own specific limitations. In community-based case-control studies, occupational chemical exposures of interest often cannot be measured directly, and have to be estimated retrospectively. As a consequence, exposure estimates are often based directly on the job information provided by the study subjects (or proxies), or inferred from the job information through job exposure matrices, or by expert assessment case-by-case. The subjectivity and the limited use of data-driven quantitative estimates of exposure used in case-control studies are considered important shortcomings that can lead to substantial exposure misclassification. How to improve retrospective exposure assessment methods has been of frequent debate. The difficulty in evaluating the validity of all retrospective exposure assessment methods in the absence of a gold standard is considered a major shortcoming. Reliability studies indicate that a considerable amount of misclassification can occur in all retrospective methods [14-16]. There is, however, an urgent need to quantify levels of misclassification expected from each method in order to anticipate on the attenuation of the resultant odds ratio (OR) estimates. A major area of improvement needed is the departure from crude exposure indicators such as never/ever exposed towards a more quantitative exposure assessment covering the entire exposure period. Quantitative exposure data that enable the investigation of the exposure-response relationship is an important criterion for causation. The availability of quantitative exposure data also facilitates valid comparison of risk and exposure-response relations among studies, countries and industries, and provides a solid base for risk assessment and standard setting.

A potentially useful approach for improving exposure estimate is to use biomarkers of exposures. Biomarkers of exposure have been used in health risk assessment for a long time but its potential use in risk assessment has not been fully exploited. The utility of biomarkers is greatly enhanced when body burden resulted from multiple source exposures, or when the source of exposure is difficult to identify (e.g. exposure to second hand smoke). There are situations where biomarkers of exposures could provide valuable information in addition to traditional measurement of exposures in epidemiological studies. For instance, when data obtained from studies of a community exposed to elevated arsenic contamination in drinking water is used for risk assessment, a contentious issue is: what is the total arsenic intake for a person? Assessment of exposure is often based on the measured concentrations in the drinking water, estimations of the amount of water consumed and used for cooking, and an estimate of dietary intake. Under these uncertainties, biological monitoring offers a useful solution as biomarkers of exposure providing data on the absorbed dose for each individual. Biomarkers may include arsenic in urine, blood, hair, and nails. However, suitability of the various biomarkers to serve as indicators of acute or chronic exposure to inorganic arsenic and the various factors needs to be carefully evaluated. Although the objective here is not to evaluate arsenic risk assessment, we use it as an example to illustrate the potential utility of biomarkers.

Tables 1 and 2 show how biomarkers may be used to estimate total arsenic exposure by combining biomarker data taken from Bencko and Symon [17] and information from a relatively well-conducted control study of total arsenic

intake and biomarkers from Uchino *et al.* [18]. It is interesting to observe that the estimated total arsenic intakes calculated separately from biomarkers of hair and urine are comparable. One could use the estimated total arsenic body burden coupled with the incidence of a health effect endpoint to assess the risk associated with the exposure. This would also serve to make results comparable for a given endpoint with other community-based studies.

EPIGENOMICS - NEW CHALLENGES AND OPPORTUNITIES

While the field of cancer genetics has experienced a great deal of attention among cancer researchers in the last few decades [2, 3], appreciation of cancer epigenetics is more

recent. The study of the role of epigenetic changes induced by environmental, dietary and lifestyle factors is rapidly emerging but still in its infancy. Little is known about the precise contribution of epigenetic mechanisms to human health effects induced by adverse environmental and dietary stimuli. While there is accumulating evidence showing that aberrant DNA methylation may result from adverse exposures to epimutagens, there is a paucity of evidence regarding the effects of stimuli causing heritable changes in epigenetic information stored in histones. This is because epigenomics is a new and largely unexplored field. Although it seems inevitable that perturbations in histone modifications are induced by dietary and environmental factors that contribute to the development of human disease

Table 1. Estimates of Total Inorganic Arsenic Intake from Hair Bio-Monitoring Data

Community ¹	Distance from Source (KM)	Arsenic in Hair ($\mu\text{g per g}$) X ²	Estimated Total Arsenic Intake ($\mu\text{g Per Day}$), Y ³
A	16	0.878	240.20
B	11	1.057	250.04
C	10	1.196	257.69
C ₁	10	1.176	256.59
D	7.5	3.562	387.82
E	4	2.621	336.06
F	1.5	3.186	367.14
F ₁	1.5	3.793	400.52
G	4	3.261	371.26
H	8	1.822	292.12
I	10	1.021	248.06
J	12	1.854	293.88
K	15	2.054	304.88
L	21	1.337	265.44
M	23	0.794	235.58
N	30	0.657	228.04
O	36	0.295	200.07

¹Results of repeated examination of the same group of children after a 5-year period.

²Data from Bencko and Symon, 1977 [17].

³Y=0.055X+191.907, taken from Uchino *et al.*, 2006 [18].

Table 2. Estimates of Total Inorganic Arsenic Intake from Urine Bio-Monitoring Data

Community	Distance from Source (KM)	Arsenic in Urine ($\mu\text{g Per Liter}$), X ¹	Estimated Total Arsenic Intake ($\mu\text{g Per Day}$), Y ²
A	16	7.8	246.8
B	11	20.1	251.0
D	7.5	20.1	251.0
E	4	24.1	252.3
F	1.5	18.9	250.6
G	4	25.3	252.9
O	36	8.2	246.9

¹Data from Bencko and Symon, 1977 [17].

²Y=0.341X+244.106, taken from Uchino *et al.*, 2006 [18].

such as cancers, a rigorous proof of such a relationship remains to be established [19].

Multicenter and large cohort or case-control studies offer some of the most exciting opportunities to study the contribution of epigenetic events induced by the diet and environment to human cancer. Such examples are the objectives of the European Prospective Investigation into Cancer and Nutrition: a large prospective cohort study designed to investigate the relationship between diet, lifestyle, and the incidence of cancer in 10 European countries [20]; and a case-control study on lung and upper-aerodigestive tract cancers in Central and Eastern Europe [21-23]. These multicenter studies boast a large sample size of several thousand subjects and represent a unique possibility to identify dietary/lifestyle practices and environmental stimuli which may exert risk and/or benefit effects through epigenomic changes.

Epigenetic alterations in comparison with genetic changes are reversible and are typically acquired in a gradual manner. These features offer an important potential opportunity for prevention strategies [19].

RISK ASSESSMENT

Risk assessment is an evolving science, and methods for conducting risk assessment are still in their infancy, although undergoing rapid development. Cessation lag and lingering effect are important aspects of risk assessment and potential applications to dose-response analysis in risk assessment [24, 25]. In addition to providing insight into biological mode of action, the concept of cessation lag is useful for economic benefit analysis. Concept of effect lingering can be used to analyze epidemiological data by uncovering the hidden biological implications related to disease endpoints, thereby advancing current efforts to characterize and reduce risk assessment uncertainties. Controversies abound concerning the appropriate methods and useful data, great uncertainties involved in extrapolating beyond the range of available data, underlying biases and other limitations of observational data [26-29], and political and societal implications of these analyses [3]. Skeptics have argued that risk assessment, at least as it is currently practiced, has not been a useful tool for addressing societal concerns about exposures to environmental and occupational hazards [30]. Their primary concern is that increasingly intense debates concerning risk assessments may come to be used as an excuse for delay in the development of appropriate regulatory and other responses to environmental and occupational hazards. For example, it has taken the U.S. EPA more than 20 years to finalize its risk assessment for exposure to diesel exhaust particulates [31].

A spirited debate has emerged over the use of the "precautionary principle" as an alternative basis for public health decision-making, and has recently been embodied in environmental legislation of the European Union [32]. The precautionary principle has been defined as the need to take some 'precautionary' measures to prevent threats to human health even when a cause-and-effect relation has not been fully established [33]. We all know the story of how John Snow convinced the authorities to remove the Broad Street pump well before the cause of the cholera epidemic in London was properly understood [2,3]. This classical

successful story underscores the importance of decision-making based on the precautionary principle in the face of uncertainties. It is conceivable that John Snow and authorities must have considered the consequence of removing and not removing the pump. Therefore, with careful consideration of the risks of applying and not applying the precautionary principle, risk assessment and the precautionary principle should not be viewed as conflicting paradigms but, rather, as complementary approaches for developing appropriate policies to address risks posed by exposure to carcinogens and other hazards. Identification and quantification of risks is clearly a useful tool for informed decision-making. Appropriately, this is also the underlying principle for the current U.S. EPA's default approach for cancer risk assessment. This is not a new principle for epidemiologists. Risk assessments are inherently uncertain and should, as the National Academy of Sciences [34] suggested, be viewed as an iterative process in need of continual improvements through research targeted to fill the gaps in our evidence based knowledge. Tremendous advances in our understanding of basic epigenetic mechanisms and rapidly developing new powerful technologies, such as those for sensitive and quantitative detection of epigenetic changes as well as for genome-wide analysis (epigenomics), hold great promise that these issues may be addressed in the near future [19].

CONCLUSIONS

Epidemiological data play a crucial role in the desirable evidence-based cancer risk assessment and classification of human carcinogens. With the emerging genomic and molecular data and the adoption of the multicenter study concept, there is a real hope that the foreseeable future will bring long awaited answers to the problems we encounter today, such as the impact of aberrant genetic and epigenetic interaction with environment and diet. New biological information could lead to discovery of new biomarkers and development of novel strategies for health risk analysis useful for disease prevention, a major public health priority in the 21st century.

ACKNOWLEDGEMENTS

The author acknowledges sincere thanks to Dr. Chao Chen of the US EPA for his constructive comments that help to improve structure and informative value of the paper which was written within activities supported by the League Against Cancer, Prague and by the Follow up Lung Cancer CEEC and ARCACE EC supported study P-5/DZ-30/99/2000 experience dissemination activities.

REFERENCES

- [1] Hartung T, Rovida C. Chemical regulators have overreached. *Nature* 2009; 460(7259): 1080-1.
- [2] Bencko V. Environmental epidemiology, present chances and challenges for future. *Cent Eur J Public Health* 2007; 15(Suppl): S6-S8.
- [3] Stayner LT, Boffetta P, Vainio H. Risk Assessment of Carcinogenic Hazards. In: Schottenfeld D, Fraumeni JF Jr, Eds. *Cancer Epidemiology and Prevention*, 3rd ed. Oxford University Press 2006: pp. 65-9.
- [4] Gemignani F, Landi S, Szeszenia-Dabrowska N. Development of lung cancer before the age of 50: the role of xenobiotic metabolizing genes. *Carcinogenesis* 2007; 28: 1287-93.

- [5] Hung R, Hashibe M, McKay J. Folate-related genes and the risk of tobacco-related cancers in Central Europe. *Carcinogenesis* 2007; 28: 1334-40.
- [6] Brennan P, Hsu C, Moullan N, *et al.* Effect of cruciferous vegetables on lung cancer in patients stratified by genetic status: a mendelian randomization approach. *Lancet* 2005; 366(4): 1558-60.
- [7] Brennan P, McKay J, Moore L, *et al.* Obesity and cancer: Mendelian randomization approach utilizing the FTO genotype. *I J Epidemiol* 2009; 38: 971-5.
- [8] Moore L, Brennan P, Karami S, Hung RJ, Hsu C, Boffetta P. Glutathione S-transferase polymorphisms, cruciferous vegetable intake and cancer risk in the Central and Eastern European Kidney Cancer Study. *Carcinogenesis* 2007; 28(9): 1960-4.
- [9] Karami S, Brennan P, Hung J, *et al.* Vitamin D receptor polymorphism and renal cancer risk in Central and Eastern Europe. *J Toxicol Environ Health* 2008; Part A, 71: 367-72.
- [10] Chalmers T. Problems induced by meta-analysis. *Stat Med* 1991; 10: 971-80.
- [11] Thompson S, Pocock S. Can meta-analysis be trusted? *Lancet* 1991; 338: 1127-30.
- [12] Bletner M, Sauerbret W, Schlehofer B, Scheuchenpflug T, Friedenreich C. Traditional reviews, meta-analyses and pooled analyses in epidemiology. *I J Epidemiol* 1999; 28: 1-9.
- [13] Stroup D, Berlin J, Morton S, *et al.* Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA* 2000; 283(15): 2008-12.
- [14] Benke G, Sim M, Forbes F, Salzberg M. Retrospective assessment of occupational exposure to chemicals in community-based studies: validity and repeatability of industrial hygiene panel ratings. *Int J Epidemiol* 1997; 26(3): 635-42.
- [15] Mannetje A, Fevotte J, Fletcher T, *et al.* Assessing exposure misclassification by expert assessment in multicenter occupational studies. *Epidemiology* 2003; 14(5): 585-92.
- [16] Blair A, Stewart P, Lubin J, Forastiere F. Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. *Am J Ind Med* 2006; 50(3): 199-207.
- [17] Bencko V, Symon K. Health aspects of burning coal with a high arsenic content---arsenic in hair, urine, and blood in children residing in a polluted area. *Environ Res* 1977; 13: 378-85.
- [18] Uchino T, Roychowdhury T, Ando M, Tokunaga H. Intake of arsenic from water, food composites and excretion through urine, hair from a studied population in West Bengal, India. *Food Chem Toxicol* 2006; 44: 455-61.
- [19] Herceg Z. Epigenetics and cancer: towards an evaluation of the impact of environmental and dietary factors. *Mutagenesis* 2007; 22(2): 91-103.
- [20] Riboli E, Kaaks R. The EPIC Project: rationale and study design. *European Prospective Investigation into Cancer and Nutrition. I J Epidemiol* 1997; 26(Suppl 1): S6-S14.
- [21] Hung R, Brennan P, Canzian F, *et al.* Large-scale investigation of base excision repair genetic polymorphisms and lung cancer risk in a multicenter study. *J Natl Cancer Inst* 2005; 97: 567-76.
- [22] Hung RJ, van der Hel O, Tavitigian SV, Brennan P, Boffetta P and Hashibe M. Perspectives on the molecular epidemiology of aerodigestive tract cancers. *Mutat Res* 2005; 592: 102-18.
- [23] Scelo G, Constantinescu V, Csiki I, *et al.* Occupational exposure to vinyl chloride, acrylonitrile and styrene and lung cancer risk (Europe). *Cancer Causes Control* 2004; 15: 445-52.
- [24] Chen C, Gibb H. Procedures for calculating cessation lag. *Regul Toxicol. Pharmacol* 2003; 38: 157-65.
- [25] Chen C. Lingering Effect: Epidemiological information useful for risk assessment. *Regul Toxicol Pharmacol* 2008; 52: 242-7.
- [26] Bencko V. Risk assessment and human exposure to endocrine disrupters. In: Jedrychowski WA, Perera FP, Maugeri U, Eds. *Molecular Epidemiology in Preventive Medicine. International Center for Studies and Research in Biomedicine in Luxemburg* 2003; 315-27.
- [27] Bencko V, Rameš J, Van den Berg M, Pleško I, Trnovec T. Human exposure to polyhalogenated hydrocarbons and incidence of selected malignancies. In: Donnley K, Cizmas L, Eds. *Environmental Health in Central and Eastern Europe. Springer, Dordrecht* 2006; 31-7.
- [28] Sasco A. Cancer and globalization. *Biomed Pharmacother* 2007; 62: 110-21.
- [29] Bencko V, Rames J, Ondrusova M, *et al.* Human exposure to polyhalogenated hydrocarbons and incidence of selected malignancies - Central European experience. *Neoplasma* 2009; 56(4): 353-6.
- [30] Silbergeld EK. Risk assessment: the perspective and experience of U.S. environmentalists. *Environ Health Perspect* 1993; 101: 100-4.
- [31] Stayner L. Protecting public health in the face of uncertain risks: the example of diesel exhaust. *Am J Public Health* 1999; 89: 991-3.
- [32] Commission of the European Communities. Communiqué on the Precautionary Principle. Brussels: European Commission 2000.
- [33] Kriebel D, Tickner J. Reenergizing public health through precaution. *Am J Public Health* 2001; 91: 1351-61.
- [34] NRC Report---Subcommittee on Arsenic in Drinking Water, National Research Council (NRC) 1999.

Received: May 1, 2010

Revised: June 26, 2010

Accepted: June 28, 2010

© Vladimír Bencko; Licensee *Bentham Open*.This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.