BIOMARKERS, 2000, VOL. 5, NO. 4, 307–313

SHORT COMMUNICATION

The choice of controls in a case–control study on WBC–DNA adducts and metabolic polymorphisms

CARLOTTA SACERDOTE[1], MARCO PELUSO[2], ARMELLE MUNNIA[2], CHRISTIAN MALAVEILLE[3] and PAOLO VINEIS[1]*

[1] Unit of Cancer Epidemiology, Dipartimento di Scienze Biomediche e Oncologia Umana, via Santena 7, 10126 Torino, Italy. tel (11) 670 6525, fax (11) 670 6692, e-mail: paolo.vineis@unito.it

[2] Servizio di Oncologia Sperimentale, Istituto Nazionale per la Ricerca sul Cancro, Largo R. Benzi 10, Genova, 16132, Italy

[3] International Agency for Research on Cancer, Lyon, France

Received 4 January 1999, revised form accepted 16 April 1999

The choice of the control group is a key issue in case–control studies, particularly in studies of molecular epidemiology. We discuss the potential bias introduced by different options. To exemplify the consequences of different choices, we have analysed two sets of controls in the context of a case–control study on bladder cancer: 55 were patients with urological conditions (cystitis, prostate hypertrophy), while 49 had a miscellany of medical or surgical conditions. We measured DNA adducts in white blood cells (WBC) by ^{32}P-postlabelling and a series of metabolic polymorphisms (GSTM1, GSTT1, GSTP1, NAT2, NQO1). While no statistically significant differences were found for metabolic polymorphisms, the two series of controls showed different concentrations of DNA adducts, suggesting that conditions related to bladder cancer or intermediate steps leading to bladder cancer, such as chronic cystitis, may be associated with higher adduct levels. An association between DNA adduct levels and infection has been noted before in experimental animals: both in lung and in the skin, an inflammatory response increased the biologically effective doses of polycyclic aromatic hydrocarbons. An alternative explanation is confounding; in fact, after adjustment for the level of consumption of fruit and vegetables (but not for smoking) the difference between the two control groups was no longer statistically significant. In conclusion, the choice of controls in studies of molecular epidemiology has subtle methodological implications, including confounding of metabolic/molecular measurements by complex exposures such as diet.

Keywords: case–control study, DNA adducts, metabolic polymorphisms, molecular epidemiology.

Abbreviations: GST, glutathione-S-transferase; NAT, *N*-acetyltransferase; NQO1, NAD(P)H:quinone oxidoreductase; WBC, white blood cells.

Introduction

The main types of design in epidemiology are cohort studies and case–control (case–referent) studies, which may include incident or prevalent cases. In addition, case–case studies, nested case–control studies, case–cohort studies, transitional studies, and meta-analysis have been introduced more recently and their application depends on the hypothesis to be tested and the means available

* Corresponding author.

(Rothman *et al*. 1995).

In a case–control design, 'subjects are selected on the basis of whether they do (cases) or do not (controls) have a particular disease under study. The groups are then compared with respect to the proportion having a history of an exposure or characteristic of interest' (Hennekens and Buring 1987). For example, cases may be affected by lung cancer, controls may be a sample of the general population, and the exposure/characteristic may be tobacco smoking or the CYP1A1 polymorphism.

In case–control studies, ascertainment of exposures/characteristics is usually retrospective, i.e. information is collected about past and recent history of the subjects. While exposures like smoking vary over time, other characteristics are stable. A cross-sectional ascertainment of present smoking habits or a specific metabolic enzyme phenotype among cases and controls is not a completely valid choice, because most cases have probably changed their habits as a consequence of the disease, or their metabolic status has changed for the same reason. Another reason for invalid comparisons involving phenotypes is enzyme inducibility (d'Errico *et al*. 1996). Chronic diseases like cancer are attributable to the cumulative exposure over a long time period; for example, in the British doctors' study the risk of lung cancer increased with the second power of the daily dose and the fourth power of smoking duration (Doll and Peto 1978). Conversely, stable variables like gender, blood groups and the metabolic genotypes can be validly measured with a cross-sectional design (e.g. the genotype in lymphocyte DNA of cases and controls).

Also the measurement of integrated markers of internal dose (such as DNA adducts) is strongly influenced by the choice of controls. Controls with certain diseases, different from the one under investigation, can show higher or lower adduct levels compared with a sample of healthy subjects. This can happen for example as a consequence of the endogenous formation of electrophilic compounds, as in the case of inflammatory conditions.

We exemplify the theoretical considerations above with real data from the context of a case–control study. We have conducted a case–control study in which we have measured DNA adducts in white blood cells by the ^{32}P-postlabelling method. The hypotheses underlying the study were: whether the level of DNA adducts is influenced by smoking; whether the NAT2, GSTM1, GSTT1, GSTPi, NAD(P)H:quinone oxidoreductase and other genotypes influence the level of adducts; and whether adducts are related to the case–control status. Here we analyze two different sets of hospital controls, in order to evaluate the influence of the choice of controls over the metabolic/molecular measurements.

Subjects and methods

We have conducted a hospital-based case–control investigation involving four departments of urology in two hospitals (Gradenigo and S. Giovanni Battista) in Torino, Northern Italy.

Controls were recruited (a) in the same urology departments and included benign diagnoses, mainly prostatic hyperplasia and cystitis (all newly diagnosed) and (b) in medical and surgical departments, and included hernias, vasculopathies, diabetes, heart failure, asthma and other benign diseases. All cancers, liver or renal diseases and smoking-related conditions were excluded. Controls were men living in the Torino Metropolitan Area, aged 40–74.

Cases and controls were interviewed by a trained interviewer before therapy, with a standard questionnaire concerning a detailed history of tobacco smoking (including brands and tobacco type), a simplified 24-h recall for dietary habits, drug use and occupational history.

Table 1. Distribution of genetically-based metabolic polymorphisms in two sets of hospital controls (55 urological and 49 non-urologically).

Polymorphisms		Urological controls	Non-urological controls
GSTM1	wild	24 (43·6%)	27 (55·1%)
	null	31 (56·4%)	22 (44·9%)
$\chi^2 = 1{\cdot}36$; p-value = 0·24 OR = 0·61 (0·26–1·40)			
GSTT1	wild	45 (83.3%)	43 (87.8%)
	null	9 (16.7%)	6 (12·2%)
$\chi^2 = 0{\cdot}40$; p-value = 0·52 OR = 0·80 (0·24–2·61)			
GSTP1	1a/1a	24 (44·4%)	29 (56.9%)
	1a/1b	17 (31·5%)	13 (25.5%)
	1a/1c	8 (14·8%)	4 (7.8%)
	others	5 (9·3%)	5 (9.8%)
$\chi^2 = 2{\cdot}25$; p-value = 0·5 OR = 0·64 (0·29–1·39)			
NAT2	rapid	24 (44·4%)	22 (44·9%)
	slow	30 (55·6%)	27 (55·1%)
$\chi^2 = 0.002$; p-value = 0.96 OR = 0.92 (0·41–2·07)			
NQO1	MM	2 (3·7%)	1 (2·0%)
	WM	24 (44·4%)	19 (38·8%)
	WW	28 (51·8%)	29 (59·2%)
$\chi^2 = 0{\cdot}69$; p-value = 0·70 OR = 1·42 (0·63–3·18)			

OR = age-adjusted Odds Ratios: Cl = Confidence Intervals. ORs are computed for wild homozygous subjects vs other subjects.

Before therapy, and after informed consent, blood was collected from both cases and controls (40 ml) and immediately centrifuged. Buffy coats were separated from coded blood samples (8 ml) by centrifugation at 800×*g* for 45 min, followed by lysis of the contained red cells by suspension in 3 volume 0·17 ammonium chloride at 4 °C for 10 min and centrifugation at 800×*g* for 10 min. The pellet containing WBC was washed with ammonium chloride and stored at –80 °C. WBC DNA was isolated and purified from the stored cell pellets by enzymatic digestion of RNA and proteins followed by phenol/chloroform extractions. The levels of DNA adducts were determined using the previously reported method (Peluso *et al*. 1996, 1998). The reproducibility of the technique was verified by analysing WBC DNA samples with a second independent ^{32}P-postlabelling experiment (Peluso *et al*. 1998).

Three known slow acetylator alleles (NAT2-5A, NAT2-6A and NAT2-7A) were identified by PCR as described (Bell *et al*. 1993), with slight modifications. Rapid acetylator genotypes are wild type allele homo-/heterozygotes, slow acetylator genotypes are those with two slow acetylator alleles. GSTM1 null, GSTM1A, GSTM1B and GSTM1A,B polymorphisms at the glutathione-S-transferase GSTM1 locus were identified using the method described by Fryer *et al*. slightly modified (Fryer *et al*. 1993). By similar methods, also GSTT1 (Pemble *et al*. 1994), GSTP1 (Zimmiak *et al*. 1994), NAD(P)H:quinone oxidoreductase (Rosvold *et al*. 1995, Bartsch *et al*. 1998) polymorphisms were analysed.

All the phases of the study till to the final statistical analyses were blind.

We have computed means and medians of DNA adducts, while the χ^2-test was used to compare the distributions of metabolic polymorphisms. Age-adjusted Odds Ratios (OR) and the corresponding 95% Confidence Intervals have been computed for metabolic polymorphisms (Hennekens and Buring 1987). Multivariate analyses (linear models) were performed with the SAS package for a personal computer. The logarithm of DNA adducts has been used in such analyses. Logistic regression has been used with dichotomic levels (below/above median) for DNA adducts as dependent variable.

Results

Table 1 shows the distribution of five metabolic polymorphisms by type of controls. No statistically significant difference is evident between the two control groups, and in fact most genotypes show a balanced distribution.

Table 2. Mean and median values of WBC–DNA adducts ^{32}P-postlabelling, RAL×10^{-8}) in two sets of hospital controls.

	Urological controls	Non-urological controls
n	54 (*)	49
Mean	0.27	0.16
Median	0.17	0.01
Standard error	0.05	0.03

(*), one value missing.

The concentration of WBC–DNA adducts is clearly different in the two control groups, with considerably higher concentrations in the urological group of patients (table 2). This difference is only slightly affected by adjustment for some of the potential determinants of adduct levels (age, smoking), while it is apparently explained away by adjustment for fruit and vegetable consumption (table 3, multiple regression models). Fruit and vegetable consumption, therefore, seems to be a confounder of the association between DNA adducts and type of controls.

Discussion

Problems with the choice of controls

In theory, controls should be a representative sample of the source population which generated the cases; in practice, the choice of controls has been quite variable. Hospital controls are usually included in studies on metabolic polymorphisms because of their higher response rate and compliance to the protocol, and the easier access to biological samples, among other reasons (d'Errico *et al*. 1996).

Metabolic polymorphisms

(a) The use of one specific group of patients as controls is questionable, since one cannot rule out that the association which is observed (such as a higher proportion of rapid debrisoquine metabolizers in lung cancer patients) is due to an inverse association with the control disease; for example, chronic pulmonary disease, used as a control group in studies of lung cancer, could be more frequent among poor metabolizers of debrisoquine (d'Errico *et al*. 1996).

(b) A particular form of bias, related to the selection of hospitalized patients, is called Berkson's bias. In 1946, Joseph Berkson published a paper (Berkson 1946) in which he raised a particular doubt about the validity of epidemiological research within hospital settings. The underlying idea was that the relative prevalence of a certain disease in a group of patients who are hospitalized for another disease is inherently biased when compared with the population served by the hospital. This phenomenon—a particular type of selection bias—reflects the way in which the probabilities of hospitalization combine in patients with more than one disease. Berkson's argument applies in particular to hospital-based case–control studies in which one or more risk factors (especially medications) are studied in relation to the risk of a specific disease. People with multiple diseases or conditions are over-represented in the hospital population, and his over-representation affects the distribution of risk factors as well. Berkson's bias is relevant to molecular

Table 3. WBC–DNA adducts (^{32}P-postlabelling, RAL×10^{-8}; logarithm transformation) in two sets of hospital controls: linear regression models.

Independent variable	Regression coefficient[a]	*p*-value[a]	OR (95% Cl)[b]
Model I			
Age	0.004	0.07	1.06 (1.01–1.11)
Type of controls	0.027	0.04	1.40 (1.06–1.85)
Model II			
Age	0.0044	0.06	1.07 (1.01–1.12)
Type of controls	0.029	0.039	1.42 (1.06–1.89)
Number of cigarettes	0.0003	0.8	1.01 (0.98–1.04)
Model III			
Age	0.0048	0.07	1.09 (1.03–1.15)
Type of controls	0.18	0.46	1.20 (0.75–1.94)
Number of cigarettes	0.00079	0.68	1.03 (0.99–1.08)
Fruit and vegetables (daily portions)	0.0084	0.60	1.02 (0.75–1.38)

[a] Linear regression with log (DNA adducts) as dependent variable.
[b] Based on a logistic regression with DNA adducts below/above median as dependent variable.
OR = Odds Ratio for a level of DNA adducts greater than the median; 95% Cl = 95% confidence interval.

epidemiology. One can imagine at least three mechanisms by which Berkson's bias can occur. First, if a person is hospitalized for a specific reason, but has more than one pathological condition, it is possible that the concurrent disease is also associated with the genetic polymorphisms under investigation. Second, patients with a certain allele at the polymorphic locus under investigation can have adverse reactions to drugs and be hospitalized for this reason. Third, induction of an enzyme by treatment can influence the phenotypic indicator of genotype.

(c) Although the genotype is stable over time and should be less sensitive to the choice of controls, one cannot rule out that certain allelles influence survival from specific (frequent) diseases, and are therefore associated with ageing and time since diagnosis (Okkels *et al*. 1996). For example, Kelsey *et al*. have recently shown that GSTM1 allelism can be associated with survival from breast cancer (Kelsey *et al*. 1996).

DNA adducts

Also the ways in which the choice of controls can influence the concentration of DNA or protein adducts are numerous. Infectious diseases tend to induce the formation of endogenous electrophils, which can react with macromolecules. In addition, several diseases, e.g. of the gastrointestinal organs, can change the living habits of the patients, including the consumption of foods that interfere with adduct formation (like fruit and vegetables).

We have made a comparison between two control groups in the context of a case–control molecular study. One control group was represented by urological conditions including prostate hypertrophy, while the second comprised non-urological diseases. While the frequencies of several alleles relevant to metabolic polymorphisms did not vary in the two control groups, WBC–DNA adducts had a higher concentration among urological controls. The difference was rather important in absolute terms and statistically significant. Unfortunately, the

subgroups with specific diagnoses were too small to allow more detailed inferences. We have two alternative explanations for our observation. One is that infectious urological conditions or chronic infections associated with prostate hypertrophy explain the higher adduct levels among the urological controls. An association between DNA adduct levels and infection has been noted before in experimental animals: both in the lung and in the skin, an inflammatory response increased the biologically effective doses of polycyclic aromatic hydrocarbons (Albert *et al*. 1996, Borm *et al*. 1997). This phenomenon would be a further reason for an accurate choice of controls in the context of molecular epidemiology studies. A second potential explanation is confounding; in fact, when adjusting for the level of consumption of fruit and vegetables (but not for smoking) the difference between the two control groups is no longer statistically significant (table 2, multiple regression models).

In conclusion, the choice of controls is crucial in molecular epidemiology studies. We have exemplified how different groups of controls can produce different results; in particular, the association between inflammatory diseases and DNA adducts, and confounding by habits such as fruit and vegetable consumption, are worth noting.

As practical suggestions for control recruitment one can consider the following: (a) avoid recruitment of a single disease entity, but limit a single disease to no more than 10% of controls (as an empirical rule); (b) avoid recruitment of conditions clearly related to known metabolic polymorphisms, or to adverse reactions to drugs; (c) avoid recruitment of inflammatory conditions, which can alter the measurement of DNA aducts.

Acknowledgements

This study has been funded by the Associazione Italiana per le Ricerche sul Cancro and the Italian National Research Council (Grant NO. 98.03166.CT04).

References

Albert, R. E., Miller, M. L., Cody, T. E., Talaska, G., Underwood, P. and Andringa, A. 1996, Epidermal cytokinetics, DNA adducts, and dermal inflammation in the mouse skin in response to repeated benzo(*a*)pyrene exposures. *Toxicology and Applied Pharmacology*, **136**, 67–74.

Bartsch, H., Malaveille, C., Lowenfels, A. B., Maisonneuve, P., Hautefeuille , A., Boyle, P. and International Pancreatic Disease Study Group 1998, Genetic polymorphisms of *N*-acetyltransferases, glutathione-S-transferase M1 and NAD(P)H: quinone oxidoreductase in relation to malignant and benign pancreatic disease risk. *European Journal of Cancer Prevention*, **7**, 215–223.

Bell, D. A., Taylor, J. A., Butler, M. A., Stephens, E. A., Wiest, J., Brubaker, D. H., Kadlubar, F. F. and Lucier, G. W. 1993, Genotype/phenotype discordance for human *N*-acetyltransferase (NAT2) reveals a new slow-acetylator allele common in African–Americans. *Carcinogenesis*, **14**, 1689–1692.

Berkson, J. 1946, Limitations of the application of 4-fold tables to hospital data. *Biomet. Bull*., **2**, 47–53.

Borm, P. J., Knaapen, A. M., Schins, R. P., Godschalk, R. W. and Schooten, F. J. 1997, Neutrophils amplify the formation of DNA adducts by benzo(*a*)pyrene in lung target cells. *Environmental Health Perspectives*, **105** (suppl. 5), 1089–1093.

d'Errico, A., Taioli, E., Chen, X. and Vineis, P. 1996, Genetic metabolic polymorphisms and risk of cancer: a review of the literature. *Biomarkers*, **1**, 149–173.

Doll, R. and Peto, R. 1978, Cigarette smoking and bronchial carcinoma among regular cigarette smokers and life-long-non smokers. *Journal of Epidemiology and Community Health*, **32**, 303–313.

Fryer, A. A., Zhao, L., Alldersea, J., Pearson, W. R. and Strange, R. C. 1993, Use of site-directed

mutagenesis of allele-specific PCR primers to identify the GSTM1A, GSTM1B, GSTM1A,B and GSTM1 null polymorphisms of the glutathione-S-transferase GSTM1 locus. *Biochemical Journal*, **295**, 313–315.

HENNEKENS, C. H. and BURING, J. E. 1987, *Epidemiology in Medicine* (Boston: Little, Brown and Co.).

KELSEY, T. K., HANKINSON, S. E., COLDITZ, G. A. *et al*. 1997, Glutathione-S-transferase class mu deletion polymorphism and breast cancer: results from prevalent versus incident cases. *Cancer Epidemiology Biomarkers and Prevention*, **6**, 511–516.

OKKELS, H., SIGSGAARD, T., WOLF, H. and AUTRUP, H. 1996, Glutathione-S-transferase mu as a risk factor in bladder cancer. *Pharmacogenetics*, **3**, 251–256.

PELUSO, M., MERLO, F., MUNNIA, A. *et al*. 1996, ^{32}P-Postlabeling detection of DNA adducts in peripheral white blood cells of greenhouse floriculturists from Western Liguria, Italy. *Cancer Epidemiology, Biomarkers and Prevention*, **5**, 361–369.

PELUSO, M., AIROLDI, L., MUNNIA, A., MARTONE, T., CODA, R., MALAVEILLE, C., GIACOMELLI, G., TERRONE, C., CASETTA, G. and VINEIS, P. 1998, White blood cell DNA adducts, smoking, and NAT2 and GSTM1 genotypes in bladder cancer: a case–control study. *Cancer Epidemiology, Biomarkers and Prevention*, **7**, 341–346.

PEMBLE, S., SCHROEDER, K. R., SPENCER, S. R., MEYER, D. J., HALLIER, E., BOLT, H. M., KETTERER, B. and TAYLOR, J. B. 1994, Human glutathione S-transferase theta (GSTT1): cDNA cloning and the characterization of a genetic polymorphism. *Biochemical Journal*, **300**, 271–276.

ROSVOLD, E. A., MCGLYNN, K. A., LUSTBADER, E. D. and BUETOW, K. H. 1995, Identification of an NAD(P)H:quinone oxidoreductase polymorphism and its association with lung cancer and smoking. *Pharmacogenetics*, **5**, 199–206.

ROTHMAN, N., STEWART, W. F. and SCHULTE, P. A. 1995, Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories. *Cancer Epidemiology, Biomarkers and Prevention*, **4**, 301–311.

ZIMNIAK, P., NANDURI, B., PIKULA, S., BANDOROWIEZ-PIKULA, J., SINGHAL, S. S., SRIVASTAVA, S. K., AWASTHI, S. and AWASTHI, Y. C. 1994, Naturally-occurring human glutathoine S-transferase GSTP1-1 isoforms with isoleucine and valine in position 104 differ in enzymatic properties. *European Journal of Biochemistry*, **244**, 893–899.