

Application of Reliability Models to Studies of Biomarker Validation

Emanuela Taioli, Patrick Kinney, Anatoly Zhitkovich, Helen Fulton, Victoria Voitkun, Greg Cosma, Krystyna Frenkel, Paolo Toniolo, Seymour Garte, and Max Costa

Nelson Institute of Environmental Medicine, New York University Medical Center, New York, NY 10016 USA

We present a model of biomarker validation developed in our laboratory, the results of the validation study, and the impact of the estimation of the variance components on the design of future molecular epidemiologic studies. Four different biomarkers of exposure are illustrated: DNA-protein cross-link (DNA-PC), DNA-amino acid cross link (DNA-AA), metallothionein gene expression (MT), and autoantibodies to oxidized DNA bases (DNAox). The general scheme for the validation experiments involves n subjects measured on k occasions, with j replicate samples analyzed on each occasion. Multiple subjects, occasions, and replicates provide information on intersubject, intrasubject, and analytical measurement variability, respectively.

The analysis of variance showed a significant effect of batch variability for DNA-PC and MT gene expression, whereas DNAox showed a significant between-subject variability. Among the amino acids tested, cysteine and methionine showed a significant contribution of both batch and between-subject variability, threonine showed between-subject variability only, and tyrosine showed between-batch and between-subject variability. The total variance estimated through the experiment was used to calculate the minimum sample size required for a future epidemiologic study including the same biomarkers used for the reliability study. Such validation studies can detect the various components of variability of a biomarker and indicate needed improvements of the assay, along with possible use in field studies. **Key words:** biomarkers, molecular epidemiology, reliability. *Environ Health Perspect* 102: 306–309(1994)

Ever since biomarkers became a primary tool in epidemiologic studies, the importance of validating them before using them in field studies has been stressed (1,2). Although a few studies have addressed the issue of sensitivity and specificity of biomarkers (3–7), no studies have reported the results of experiments aimed at quantifying the sources of variability connected with the use of biomarkers in human subjects. Estimating the components of variability provides valuable data for use in the design of epidemiologic studies. However, no formal model to study biomarker reliability in humans has been proposed so far.

We have developed several biomarkers of exposure to carcinogenic metals, such as chromium and nickel. In this paper, we present and analyze data collected from

validation studies carried out on blood samples from healthy volunteers. Four different biomarkers of exposure are illustrated: DNA-protein cross-link (DNA-PC), DNA-amino acid cross-link (DNA-AA), metallothionein gene expression (MT), and autoantibodies to oxidized DNA bases (DNAox). We present a model of validation developed in our laboratory, the results of the validation study, and the impact of estimating the variance components on the design of future molecular epidemiologic studies.

Background

The major components of biomarker variability that affect the design of epidemiologic studies are variability between subjects (intersubject), variability within subjects over time (intrasubject), and variability due to assay measurement errors. The impact of these three categories of variability on the biomarker response can be represented by a linear model of the form:

$$Y_{ijk} = u + a_i + b_j + e_{ijk}$$

where Y_{ijk} is the biomarker response for subject i on day j and replicate measurement k ; u is the true population mean response; a_i is the offset in mean response for subject i (assumed to be normally distributed with mean = 0 and variance = s_a^2 ; s_a^2 represents the magnitude of intersubject variability); b_j is the offset in response on day j (assumed to be normally distributed with mean = 0 and variance = s_b^2 ; s_b^2 represents the magnitude of intersubject variability), and e_{ijk} is the assay measurement error (normally distributed with mean = 0 and variance = s_c^2).

Intersubject variability in biomarker responses may arise due to factors such as genetics, race, gender, and diet. Similarly, the biomarker response for a given subject may vary over time due to a change in diet, health status, variations in exposure to the compounds of interest, and variations in exposure to other compounds that influence the biomarker (e.g., tobacco smoke). Variability in laboratory measurements can have many sources, some of which will be specific to a given assay. Two general classes of laboratory errors are worth noting: those that occur between analytical batches and those that occur within batches. Analytical variability between batches pre-

sents a special problem in that batch variability may be superimposed on, and impossible to separate from, the inter- and intrasubject components of variability. Through proper study design, however, this problem can be minimized.

For a hypothetical molecular epidemiology study involving a comparison of biomarker levels in two populations that differ in level of exposure to some chemical, it is generally desirable to minimize the total intragroup variability. That variability represents the weighted sum of the three variance components defined above, with weights inversely related to the numbers of subjects, measurements per subject, and analytical replicates used in the study design. These design components can be optimized *a priori* using information from a carefully designed validation study of the type discussed below. Once the absolute and relative magnitudes of the three variance components are known, a study design can be determined that minimizes the total intragroup variance for a given level of resources.

Methods

The general scheme for the validation experiments, illustrated in Table 1, involves n subjects measured on k occasions, with j replicate samples analyzed on each occasion. Multiple subjects, occasions, and replicates provide information on intersubject, intrasubject, and analytical measurement variability, respectively. The specific experiment carried out for the different assays, described in detail below, included one or more components of this general scheme.

We analyzed data using analysis of variance methods. For the most general experimental design (illustrated in Table 1), we used two-way, random-effects analysis of variance, with subjects and occasions as the two main effects (8). This approach provides direct estimates of the three variance components of interest, where the variance estimates for the two main effects correspond to inter- and intrasubject variability, and the error variance corresponds to analytical measurement variability.

We applied the model described above to four biomarkers of exposure to metals. The validation experiments performed for

Address correspondence to E. Taioli, Department of Environmental Medicine, Epidemiology Program, New York University Medical Center, 341E 25th Street, New York, NY 10016 USA.

This study was supported by grants CA34588 and ES04895. The authors thank M. Kim for the useful suggestions and comments and D. Gray for typing.

Received 2 November 1993; accepted 28 January 1994.

Table 1. General scheme for the experiment of biomarkers validation^a

Subjects	Time (weeks)			
	1	2	3	k
1	1 _{1a,b,j}	1 _{2a,b,j}	1 _{3a,b,j}	1 _{ka,b,j}
2	2 _{1a,b,j}	2 _{2a,b,j}	2 _{3a,b,j}	2 _{ka,b,j}
3	3 _{1a,b,j}	3 _{2a,b,j}	3 _{3a,b,j}	3 _{ka,b,j}
n	n _{1a,b,j}	n _{2a,b,j}	n _{3a,b,j}	n _{ka,b,j}

^aa,b,j = replicates of the same specimen at each time.

each assay are described below. In addition, a brief description of the assays is reported here; more technical details are contained in specific publications cited throughout the text.

DNA-Protein Cross-link

The DNA-PC method detects the amount of proteins covalently linked to DNA, a process that has been implicated in many aspects of gene expression and inheritance. The procedure is applied to frozen white blood cells and is based on the fact that sodium dodecyl sulfate (SDS) binds to proteins but not to DNA. Addition of potassium chloride to SDS solution results in the formation of a potassium-SDS precipitate which is easily recovered by low-speed centrifugation. Binding of SDS to proteins cross-linked to DNA leads to selective precipitation of DNA containing a cross-linked protein upon addition of potassium chloride. The method is described in detail elsewhere (9).

We recruited five healthy, unexposed subjects for this study. Blood was drawn three times in three different weeks. Each time, the specimens were divided into three or four aliquots to obtain information on laboratory variability on the same subject. Thus, this experiment followed the scheme illustrated in Table 1, with $n = 5$, $k = 3$, and $j = 3$ or 4. We analyzed DNA-PC at each week, introducing a possible batch effect that was coincident with weeks.

Amino Acids Cross-linked to DNA

After cellular DNA was purified by standard proteinase with the phenol procedure, the residual amino acids associated with DNA were detected using ophthalaldehyde fluorescence. The methods will be published (10).

As in the previous experiment, five healthy, unexposed subjects provided blood samples on three separate weeks, with laboratory replicates on each occasion. We analyzed specimens each week, introducing a possible batch effect coincident with week. Although a total of 14 amino acids were analyzed, we present the results of the variability study on the 7 that have been demonstrated to increase in their association with DNA after exposure

to carcinogenic metals such as chromate (cysteine, histidine, tyrosine, threonine, methionine, glutamine, and glutamic acid) (10,11).

Metallothionein Gene Expression

The MT method is aimed at detecting the level of MT gene expression in peripheral lymphocytes isolated from whole venous blood and induced *in vitro* by treatment with 1 μ M CdCl₂ for 1, 3, and 6 hr. MTs are proteins that form complexes with toxic metals, and their synthesis is induced by exposure to metals. We determined gene induction by isolating mRNA from the cells followed by standard slot blot hybridization analysis using a human MT cDNA probe. The method is described in detail elsewhere (12).

Four healthy, unexposed subjects were included in the validation experiment, and repeated measurements were performed at four different weeks. We took only one measurement per subject at each week, precluding the estimation of analytical measurement variability from this experiment. Three separate measures of MT gene expression were considered: basal level of MT gene expression and cadmium-induced gene expression after 2 hr and at 6 hr of treatment *in vitro*. Variability for each of these measurements was calculated. Because the experiment required fresh lymphocytes, it was performed at each week, resulting in a potential batch effect that was considered in the statistical analysis.

Autoantibodies against Oxidized DNA Bases

The DNAox method measures the amount of serum autoantibodies that recognize oxidized DNA bases as a marker of inflammatory immune response to metals. Antigens consist of the riboside of 5-hydroxymethyl uracil (HMU) coupled to bovine serum albumin (HMdU-BSA) and mock-coupled BSA (M-BSA). Serum is diluted and incubated in the antigen-coated wells; wells are washed and then incubated with goat anti-human IgM labeled with horseradish peroxidase (HRPO). Addition of H₂O₂ and *o*-phenylene diamine for HRPO-mediated oxidation allows the development of color, which is measured at 492 nm in the microplate reader. More details are provided elsewhere (13).

Nine subjects who were part of a cohort study had blood drawn at three (n

= 3) or four ($n = 6$) different time points spanning 1–4 years. Blood samples were frozen upon collection and stored for later analysis in one batch. At the time of analysis, we analyzed duplicate specimens from each subject.

Sample Size Calculation

Reliability studies help to determine the sample size required in a future study. We hypothesized a classical study of comparison between two groups of subjects, one exposed and one nonexposed to a certain carcinogen. If a certain mean difference, d , between two groups is considered important and a significance level 0.05 and a 95% power is chosen, the number of subjects per group, n , is calculated as follows:

$$n = \frac{2(s_T^2 + s_e^2)(z_{\alpha/2} + z_\beta)^2}{d^2}$$

where $(s_T^2 + s_e^2)$ is an estimator of the variance that will characterize the variability of single measurements on those subjects, as provided by the reliability study (8).

Small pilot studies recently conducted by our group on subjects occupationally or environmentally exposed to carcinogenic metals were used to obtain an approximate value for d , the mean difference between the two groups. When data were not available, as for the DNA-AA assay, we considered a 50% increase of the value observed among the controls included in this study to be of interest for a future study.

Results

DNA-PC. The analysis of variance showed a significant effect of batch variability, but the intersubject variability was of borderline significance. A significant interaction between these two variables was also observed (Table 2). The interaction can be interpreted either as evidence that the batch effect varied across subjects or that there was additional week-to-week variability in DNA-PC beyond that due to the batch effect. This additional variability may represent biological variability over time in individual subject responses.

DNA-AA. Among the amino acids tested, cysteine and methionine showed a significant contribution of both batch and intersubject variability, but no interaction between them. Threonine showed intersubject variability only, and tyrosine

Table 2. Analysis of variance on the reliability study: DNA-protein cross-link

Variance component	Variance estimate	F(df)	p
Week	0.0545	13.68 (2,7)	<0.01
Between subject	0.0176	3.45 (4,7)	0.05 < p < 0.10
Week * subject	0.0110	2.33 (7,40)	0.045
Error	0.0317		

Table 3. Analysis of variance on the reliability study: amino acids cross-linked to DNA

Amino acid	Variance component	Variance estimate	F(df)	p
Cysteine	Week	0.0473	16.72 (2,23)	0.0001
	Between subject	0.0078	2.63 (4,23)	0.06
	Error	0.0280		
Methionine	Week	0.0044	5.11 (2,21)	0.02
	Between subject	0.0092	6.99 (4,21)	0.002
	Error	0.0059		
Threonine	Week	0	<1	ns
	Between subject	0.2010	4.69 (4,23)	0.0065
	Error	0.3188		
Tyrosine	Week	0.0413	2.44 (2,14)	>0.1
	Between subject	0.0193	1.47 (4,14)	>0.25
	Week* subject	0.0832	7.29 (5,14)	0.0015
	Error	0.0271		

ns, nonsignificant.

Table 4. Analysis of variance on the reliability study: metallothionein gene expression

	Variance component	Variance estimate	F(df)	p
Basal	Week	0.019	9.00 (3,4)	<0.05
	Between subject	-0.003	0.52 (4,9)	ns
	Error	0.021		
Induced/basal (2 hr)	Week	116.34	13.98 (3,4)	<0.05
	Between subject	2.87	1.34 (4,9)	ns
	Error	27.63		
Induced/basal (6 hr)	Week	40.95	3.48 (3,4)	ns
	Between subject	-8.19	0.75 (4,9)	ns
	Error	105.99		

ns, nonsignificant.

Table 5. Analysis of variance on the reliability study: DNA oxidative damage

Variance component	Variance estimate	F(df)	p
Between subject	626.97	103.89 (8,23)	<0.001
Within subject	14.37	3.12 (23,20)	<0.05
Error	11.07		

Table 6. Sample size calculation for a future epidemiologic study

Assay	Variance	d	Minimum sample size ^a
DNA-PC	0.0603	0.65% DNA (unpublished)	7
Cysteine	0.0766	0.35 nmol AA/mg DNA ^b	16
Methionine	0.0209	0.1 nmol AA/mg DNA ^b	54
Threonine	0.467	0.9 nmol AA/mg DNA ^b	29
Tyrosine	0.171	0.5 nmol AA/mg DNA ^b	18
Metallothionein (induced/basal at 6 hr)	155.13	3.5 (unpublished)	329
Metallothionein (basal)	0.043	0.23 (unpublished)	21
Antibodies against oxidized DNA bases	652.41	8.6 A ₄₉₂ /μl (14)	38

Abbreviations: DNA-PC, DNA-protein cross-linking; AA, amino acid.

^aFor each of the two groups.

^bSince preliminary data are not available, a 50% increase of the value observed among the controls is considered of interest.

showed interaction between batch and intersubject variability (Table 3). For histidine, glutamine, and glutamic acid, we did not detect any significant contribution of the different components to the variability of the assay.

MT. As shown in Table 4, the batch effect was more important than between intersubject variability in basal gene expression or induced/basal at 2 and 6 hr. To better understand whether the batch effect was due to laboratory or intrasubject variability, we conducted a small experiment

on one subject, whose lymphocytes were divided into five batches, treated individually with CdCl₂, and analyzed as independent samples in the same day by a technician who was not aware of the nature of the experiment. The coefficient of variation of the data was 0.18, showing a low level of variability due to laboratory error when time was not a factor. The variance of induced/basal level at 6 hr was 20.7. From Table 4, the combined effect of laboratory and intrasubject variability at 6 hr was 40.7. By comparing the two results,

we can hypothesize that half of that variance is due to laboratory error, half to intrasubject variability.

DNAox. Among the different components of variance, intersubject variability was significantly more important than intrasubject variability (Table 5). No interaction was present. The total variance estimated through the experiment described above was used to calculate the minimum sample size theoretically required for a future epidemiologic study including the same biomarkers used for the reliability study. Table 6 shows the total variance, the mean difference between the two groups considered biologically important, and the minimum sample size required to detect that difference.

The calculated sample sizes varied from 329 subjects for MT gene expression to 7 subjects for DNA-PC; a study involving the use of DNAox would have to include at least 38 subjects exposed and 38 controls. The other assays require intermediate values.

Discussion

We have presented a method to validate the use of some biomarkers of exposure in human populations by studying the different components of their variability. This step in biomarker validation gives information on whether the biomarker is suitable for epidemiologic studies. In fact, even in the presence of acceptable laboratory sensitivity and specificity of a given biomarker, the excess of intraindividual and laboratory variability observed in healthy volunteers might make its use in human studies not feasible. For example, in our experiment we observed a high level of variability in induced/basal levels of MT gene expression, mainly due to a temporal effect that might be ascribed to both intraindividual and batch variability. Unfortunately, efforts to reduce these sources of variability have been unsuccessful to date, perhaps due to the nature of the experiment, which involved fresh lymphocytes grown and stimulated *in vitro*. However, basal levels of MT gene expression exhibit a small amount of variability, therefore their use in epidemiologic studies should be considered more promising than the induced levels of gene expression.

DNA-PC showed a significant batch effect and a significant interaction between subjects and batch. The latter may be interpreted in two ways: as an indication that the batch effect varies across subjects, or that there is significant intrasubject variability over time due to biological factors. In the present context, none of these possibilities can be substantiated or rejected. However, to avoid the laboratory batch effect in future work, the assay has been

modified so that isolated lymphocytes are now frozen and the detection of DNA-PC is performed later, in one single experiment.

The examples reported show how the analysis of biomarker variability can be used by both the epidemiologist to design a new study, and the molecular biologist to technically improve the assay. A validation study such as the one we conducted can detect the amount of variability of the biomarker present in healthy subjects and presumably variability expected in a field study involving exposed populations and controls; this information can be used by the epidemiologist to calculate the sample size required for future studies. Calculating sample size is extremely helpful, especially when different biomarkers are tested in the same study. This happens more and more often in order to optimize the use of available resources and to improve the understanding of biological mechanisms (1). We may hypothesize that the biomarkers described in our paper will be included in a single epidemiologic study as markers of exposure to carcinogenic metals. In this case, a sample size that can be sufficient to detect a significant difference in DNA-PC between exposed and controls may not be adequate when using an assay showing a great amount of biological variability, such as induced levels MT gene expression.

In conclusion, we propose that the laboratory development of a new biomarker

should be followed by small-scale studies on healthy volunteers to assess intra- and interindividual variability and laboratory error. We demonstrated that the same design can be used by different laboratories, with small modifications. The results can be helpful to improve the biomarker reliability and to plan future epidemiologic studies.

REFERENCES

- Hulka BS. Epidemiological studies using biological markers: issues for epidemiologists. *Cancer Epidemiol Biomed Prev* 13:13-19 (1991).
- Schulte PA. Methodologic issues in the use of biologic markers in epidemiologic research. *Am J Epidemiol* 126:1006-1016(1987).
- Schatzkin A, Freedman LS, Schiffman MH, Dawsey SM. Validation of intermediate end points in cancer research. *J Natl Cancer Inst* 82:1746-1752(1990).
- Griffith J, Duncan RC, Hulka BS. Biochemical and biological markers: implications for epidemiologic studies. *Arch Environ Health* 44:375-381(1989).
- Hatch MC, Friedman-Jimenez G. Using reproductive effect markers to observe sub-clinical events, reduce misclassification, and explore mechanisms. *Environ Health Perspect* 90:255-259(1991).
- Friedman-Jimenez G. Nondifferential misclassification and the potential of biological markers to improve validity of studies in occupational epidemiology. Presented at the Ninth International Symposium on Epidemiology in Occupational Health, Cincinnati, Ohio, September 1992.
- Rothman N, Stewart WF, Caporaso NE, Hayes RB. Misclassification of genetic susceptibility biomarkers: implications for case-control studies and cross-population comparisons. *Cancer Epidemiol Biomed Prev* 2:299-303 (1993).
- Fleiss JL. *The design and analysis of clinical experiments*. New York:John Wiley and Sons,1986.
- Zhitkovich A, Costa M. A simple sensitive assay to detect DNA-protein crosslinks in intact cells and in vivo. *Carcinogenesis* 13:1485-1489(1992).
- Voitkun V, Zhitkovich A, Costa M. Complexing of amino-acids to DNA by chromate in intact cells. *Environ Health Perspect Suppl* (in press).
- Lin X, Zhuang Z, Costa M. Analysis of residual amino-acid DNA crosslinks induced in intact cells by nickel and chromium compounds. *Carcinogenesis* 13:1763-1768 (1992).
- Cosma G, Currie D, Squibb KS, Snyder CA, Garte JS. Detection of cadmium exposure in rats by induction of lymphocyte metallothionein gene expression. *J Toxicol Environ Health* 34:39-49(1991).
- Frenkel K, Karkoszka J, Kim E, Taioli E. Recognition of oxidized DNA bases by sera of patients with inflammatory diseases. *Free Rad Biol Med* 14:493-494(1993).
- Frenkel K, Karkoszka J, Baranski B, Cohen B, Cosma G, Jakubowski M, Taioli E, Toniolo P. Occupational and environmental exposure to Cd, Ni, and Cr modulates titers of anti-oxidized DNA base autoantibodies. *Environ Health Perspect Suppl* (in press).

Volume 102, Supplement 1, January 1994

Biostatistics in the Study of Toxicology

Environmental Health
perspectives
Supplements

Volume 102, Supplement 1, presents the proceedings of the International Biostatistics Conference on the Study of Toxicology held May 23-25, 1991, at Sanjo Kaikan, University of Tokyo, Japan. Because the importance of statistical methods in toxicology is recognized and methodologies are rapidly developing, the main objective of this meeting was to discuss the application of statistical methods for the evaluation of toxicological data. Sponsors for the conference are the Biometric Society, Japanese Region; Environmental Mutagen Society of Japan; National Institute of Hygienic Sciences, Japan; and the National Institute of Environmental Health Sciences, USA.

To order your copy, contact:
Government Printing Office
Washington, DC 20402
or call 202-512-2406.