Statistical Science
1988, Vol. 3, No. 3, 351–357

Statistical Aspects of Using Biologic Markers

Barry H. Margolin

Abstract. This expository paper surveys a variety of statistical issues pertinent to the design and analysis of studies involving biologic markers of human genotoxic exposure. Examples with cytogenetic and mutagenic end points are presented. One principal theme is the valuable interplay of ideas among statistical analyses for in vitro, in vivo and human assays for genetic toxicity; e.g., statistical analysis of sister chromatid exchanges in vitro is suggestive of an analytical approach to the study of sister chromatid exchanges in humans. Other topics discussed include (i) combining information from a series of studies of a common suspect genotoxic exposure and (ii) the utility of historical control information.

Key words and phrases: Ames Salmonella assay, genetic toxicology, historical control data, meta-analysis, Poisson dispersion test, sister chromatid exchange assay.

1. INTRODUCTION

If one were to ask most statisticians how the second half of the twentieth century will be described by future historians of science, they probably would label it the Age of Electronic Computers, their professional activities having been so heavily influenced by computers. This period, however, will more likely be known as the Age of Molecular Genetics. The latter description is not intended to deny the considerable societal importance of computers, but rather to stress the profound consequences of the knowledge explosion in molecular genetics, touching nearly every aspect of human existence, from the economy to human emotion, thought and health. With regard to human health, one area of science that has and will continue to benefit from advances in molecular genetics is genetic toxicology, i.e., the science that studies the induction of heritable damage to genetic material by sundry toxic agents of interest. Another term for this discipline is mutagenesis. The established relevance of this cellular change to certain forms of carcinogenesis, and possibly aging and birth defects as well (Ames, 1979), is reason enough to explain why interest in this discipline has risen exponentially in the last decade. Not surprisingly, statisticians have become aware of the challenges genetic toxicology presents and have begun to make notable contributions to the field. This expository paper briefly reviews aspects of statistical analyses for two primary in vitro (literally, in glass, i.e., in test tubes) genotoxicity assays, the Ames Salmonella and sister chromatid exchange (SCE) assays, and discusses their utility for analyses of studies involving human subjects. Additionally, two other topics of relevance to human studies, the utility of historical control data and the combining of information from a series of comparable studies, are addressed. Illustrative examples are drawn mostly from papers published in genetic toxicology journals, which may have escaped the attention of statisticians newly interested in the field.

Barry H. Margolin is Professor and Chairman, Department of Biostatistics, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599.

2. STATISTICAL ANALYSES FOR IN VITRO ASSAYS: UTILITY FOR HUMAN STUDIES

Studies to monitor human exposure to genotoxic chemicals have greatly increased in number during the last decade; for example, Ashby and Richardson (1985) were able to locate by computer search 70 human studies with a cytogenetic (chromosome level) end point that were published from 1980–1983, whereas in all of 1965–1979 they found only 33 such published studies. To date studies of genotoxicity involving human subjects have been distinguished from in vitro and in vivo (living multicellular organisms, typically mammals) studies in one major respect: human studies have been strictly observational, as opposed to controlled laboratory experiments in which the investigator assigns treatments to experimental

units or subjects. Despite this important difference, the analysis of genotoxicity studies of human subjects can benefit from consideration of the analysis of in vitro studies involving the same genetic end point. This will be illustrated by reference to studies of mutagenic and cytogenetic end points involving two of the most commonly used assays for human investigations.

2.1 The Ames Assay

The dominant in vitro assay for the study of genetic toxicity is the Ames Salmonella/microsome assay (Ames, McCann and Yamasaki, 1975), which is now in international use in nearly 3000 laboratories. The assay keys on histidine, an amino acid essential for growth. Based on the creation of histidine-deficient strains of Salmonella, this assay monitors mutation (change in DNA sequence) to histidine independence. The response observed is the number of visible, presumably histidine-independent (revertant) colonies on a Petri dish or plate. One widely used method for monitoring human subjects for exposure to mutagenic agents involves the collection and concentration of urine, which is then assayed for mutagenicity via the Salmonella/microsome assay. For a discussion of some of the advantages and disadvantages of this in vivo-in vitro approach to monitoring for exposure to mutagens, see Vainio, Sorsa and Falck (1984). Table 1 presents urine mutagenicity data for a single nonsmoking subject from a study of the effects of passive smoking (Dr. G. W. Collman, personal communication). Urine concentration is linked to dosage because a constant volume of the various urine-dilutant mixtures is pipetted onto test plates. Three replicate measurements with Salmonella strain TA1538 were made at each of four logarithmically spaced concentrations of the subject's urine sample; three control plates were measured as well. These data are presented graphically in Figure 1, where one characteristic of the dose-response curve frequently associated with the Ames assay is easily discerned: the downturn in response at high doses, sometimes, as in Figure 1, to levels at or below background.

For the standard in vitro uses of the Ames assay, this downturn in response is usually attributed to the increasing toxic effects of higher concentrations of the test chemical, i.e., "dead" microbes don't mutate. With in vivo-in vitro urine studies the toxicity could derive from a host of chemicals in a subject's urine; regardless of its source, however, the downturn in dose response adds a complexity to the design and statistical analysis of such studies.

In view of the possibility of downturns in dose response, urine mutagenicity studies must include multiple concentrations for each subject; single-concentration studies are open to sufficiently serious challenge that they should generally be avoided. The in vivo-in vitro urine investigations share many similarities with the standard in vitro Ames bacterial mutagenicity assay, but there are major differences. A primary distinction from a statistical viewpoint is that in the latter case, there is generally a single agent for which one attempts to assess the evidence for a mutagenic dose response. In the former case, however, each subject can exhibit his own dose response and the goal is to determine whether and how a particular set of factors affects the collection of observed subject dose-response curves. To this end, one can specify a relevant characteristic of the dose-response curve to be measured for each subject. Presumably this measurement should reflect the mutagenic potency of the subject's urine, adjusted for possible subject-specific toxicity. Various statistical analyses proposed for in vitro Ames Salmonella assay data strive to measure mutagenic potency, with an adjustment for any attendant toxicity (Margolin, Kaplan and Zeiger, 1981; Myers, Sexton, Southerland and Wolff, 1981; Stead, Hasselblad, Creason and Claxton, 1981; Bernstein, Kaldor, McCann and Pike, 1982). Each of these methods can be employed to analyze data such as in Table 1, and to produce a potency estimate. To illustrate,

TABLE 1
Urine mutagenicity data with Salmonella strain TA 1538 for a single nonsmoking subject[a]

	Revertant colonies per plate at urine concentration of:				
	0 ml/plate	1.45 ml/plate	2.90 ml/plate	5.80 ml/plate	11.60 ml/plate
	13	18	19	21	12
	16	18	22	24	15
	18	19	24	30	18
Mean	15.7	18.3	21.7	25.0	15.0
Fitted value with model (1)	15.3	19.0	21.8	24.5	15.0

[a] Revertant colonies per plate are listed in ascending order within each concentration. A constant volume of the five urine-dilutant mixtures was pipetted onto test plates. Source: Dr. G. W. Collman, personal communication.

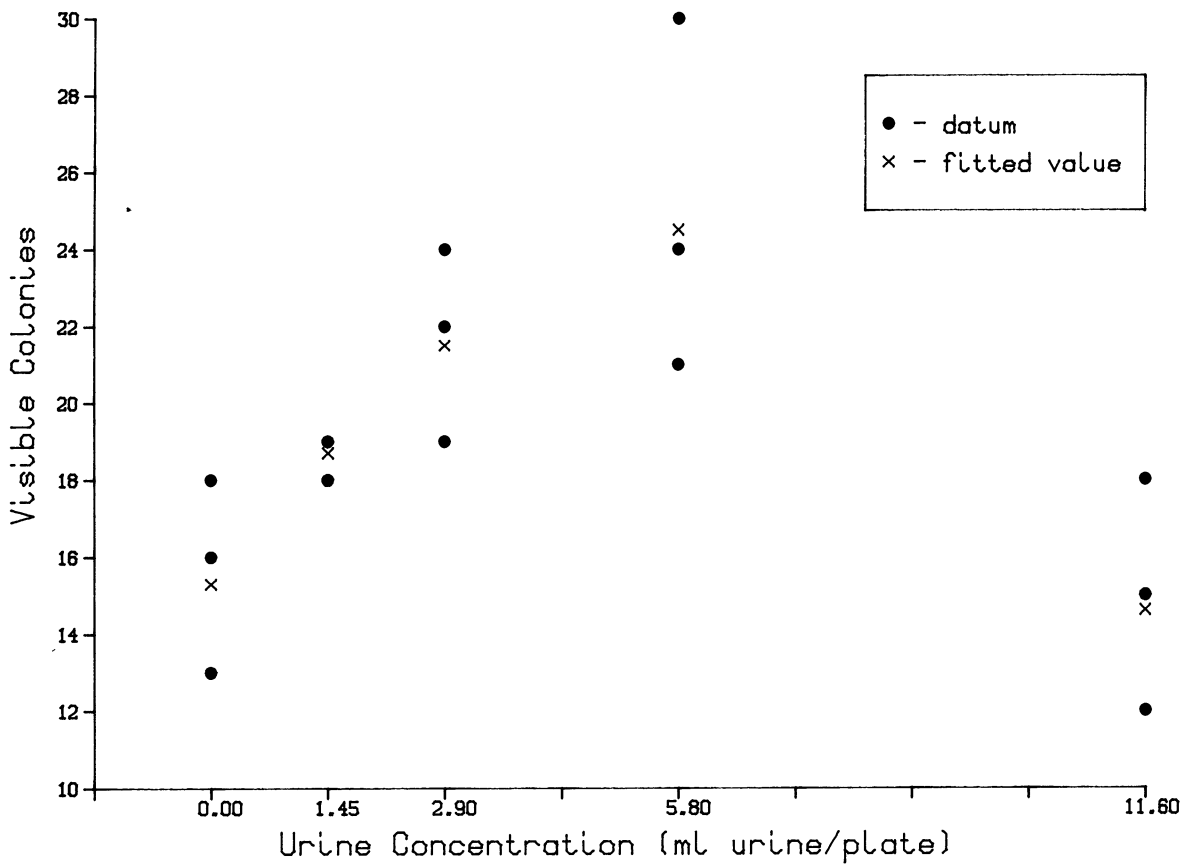

FIG. 1. *Plot of observed and fitted TA1538 plate counts versus urine concentration for a single nonsmoking subject. Data are in Table 1.*

one of the methods proposed by Margolin, Kaplan and Zeiger (1981) will be applied to the data in Table 1.

In this approach, P_D, the probability that a plated microbe yields a revertant colony after exposure to dose D of the test substance, is given by

$$(1) \quad P_D = \{1 - \exp[-(\alpha + \beta D)]\} \cdot [2\text{-}\exp(\gamma D)]_+,$$

where $[x]_+ = \max(0, x)$, $\exp[t] = e^t$, $\alpha > 0$, $\beta \geq 0$ and $\gamma \geq 0$.

The expression for P_D is seen to be a product of two terms, the first of which describes the mutagenic effect of the dose D and the second survival. The parameter β in (1) is the slope of the dose-response curve at zero dose after adjustment for toxicity, and, as such, its estimate provides a useful mutagenic potency measure for each subject. Fitting the model in (1) to the data in Table 1 via the method described by Margolin, Kim and Risko (1986) yields the estimate $\hat{\beta} = 3.6$ revertants/ml of urine. The fitted or predicted values from this model, assuming a plating of 10^8 microbes/plate, are included in Table 1 and Figure 1; the agreement is excellent, although this is not surprising given that three model parameters are used to describe five means.

In the standard in vitro use of an Ames assay, the question of whether there is evidence of mutagenicity can be formulated statistically as a test of significance for the hypothesis $\beta = 0$. In urine mutagenicity studies, as mentioned earlier, the question is not whether an individual subject's urine is mutagenic, but whether suspected mutagenic risk factors, e.g., passive smoking, lifestyle or occupational exposures, exhibit significant effects on the set of subject mutagenic potencies. One way to address this issue is first to compute an estimate $\hat{\beta}$ for each subject and then by standard statistical methods to analyze the set of $\hat{\beta}$ or $\log(1 + \hat{\beta})$ values for various factor effects. Covariates of interest, such as age, sex, race or variables related to exposure, are readily accommodated in this proposed approach, which is a decided improvement over the ad hoc and subjective analyses that have been commonly performed in urine mutagenicity studies.

2.2 Sister Chromatid Exchange

The two spiral filaments that constitute a chromosome are referred to as "chromatids." An SCE results from a reciprocal exchange of DNA between two sister chromatids. Although in and of itself, the exchange preserves the DNA content of the two chromatids, the SCE is indicative of a potentially damaging process that presumably produced DNA breakage and reunion (Latt, Allen, Bloom, Carrano, Falke, Kram, Schneider,

Schreck, Tice, Whitfield and Wolff, 1981). SCEs can be measured either in cultured cells, intact animals or humans; in all cases, the measurement of interest is a count representing the number of SCEs observed per cell. The statistical behavior of SCEs in cultured Chinese hamster ovary (CHO) cells (cultured cells) has been studied extensively by Margolin, Resnick, Rimpo, Archer, Galloway, Bloom and Zeiger, 1986). They present substantial empirical evidence that control cells from replicate cultures (flasks) created on the same day exhibit SCE sampling variability that is well approximated by a Poisson distribution. Their documentation of this behavior makes extensive use of the dispersion test for Poisson sampling, which is defined as follows (Snedecor and Cochran, 1967, page 232): If $X_1, \cdots, X_r$ represent SCE counts for r cells, then the dispersion test for Poisson sampling is based upon the ratio of the sample variance to sample mean, i.e.,

$$(2) \qquad H = \sum_{i=1}^{r} (X_i - \bar{X})^2/(r-1)\bar{X},$$

where $\bar{X} = \sum X_i/r$. The dispersion test compares the statistic $(r-1)H$ to a table of critical values for a χ^2 distribution with $r-1$ degrees of freedom to obtain an observed significance level. The demonstration that Poisson sampling is a tenable model for SCEs in control CHO cells has important implications for laboratory quality control, analysis of dose-response curves and protocol power investigations (Margolin, Resnick, Rimpo, Archer, Galloway, Bloom and Zeiger, 1986).

In human SCE studies employing peripheral lymphocytes (white blood cells), one might be inclined to proceed in parallel with the in vitro cultured cell development. There are, however, many authors who have reported that the Poisson model is not adequate to describe the sampling variability of SCEs within human subjects (see, e.g., Carrano and Moore, 1982; Hirsch, McGue and Cervenka, 1984; Husum, Wulf and Niebuhr, 1982; Yakovenko and Platonova, 1979). This departure from Poisson sampling in the direction of excess dispersion presumably reflects the heterogeneous nature of an individual's peripheral lymphocytes; the usual human cell scored for SCEs because of its ease of collection. This heterogeneity among cells of a common subject may be attributable to their varying age, type, sensitivity to DNA damage or other unknown considerations that contribute to SCE formation and accumulation.

Margolin and Shelby (1985) have proposed that the statistic H in (2) can function as "an index of heterogeneity of SCE cell counts within an individual [subject], suitably normalized to adjust for the individual's mean SCE level." Citing Moran (1973), Margolin and Shelby note that the statistic H is particularly sensitive to the presence of mixtures of Poisson distributions, an attractive model for human SCE counts from peripheral blood. It is also particularly sensitive to a long-tailed distribution that might include unusually high frequency cells (HFCs), whose definition and role in a statistical analysis have been discussed by Carrano and Moore (1982).

Margolin and Shelby (1985) proposed that an analysis of study subject H values be included as an adjunct to an analysis of subject mean values. To illustrate their proposal, they considered the study by Butler (1981) that reported SCE results for 32 normal healthy adults drawn from four different racial groups within the United States. Table 2 contains the mean and H values for these individuals; the individual SCE counts, the number of cells scored per subject and information regarding sex, age and smoking status are contained in Margolin and Shelby (1985). Butler (1981) found no significant differences in the mean SCE levels among the four racial groups via the Kruskal-Wallis test ($0.2 > p > 0.1$). Margolin and Shelby (1985) report, however, that the Kruskal-Wallis test does indicate a significant difference in racial H values ($p < 0.001$), with the Oriental group exhibiting elevated values. Indeed, the Oriental subjects account for a disproportionate number of cells with high SCE counts, e.g., seven of the eight Oriental subjects had at least one cell with an SCE count greater than 15 (ten cells in all), although none of the nine Caucasians, three of the eight Blacks (four cells in all) and one of the seven Native Americans (one cell) had cells with such scores. Margolin and Shelby (1985) discuss a possible artifactual explanation for the racial effect observed, namely that time of cell scoring and racial group were confounded. Nevertheless, whether the effect is an artifact of the data collection, this example illustrates the potential benefit of increased sensitivity derived from using the

TABLE 2
Means and H-values for four racial groups (Butler, 1981; Margolin and Shelby, 1985)

Black		Caucasian		Native American		Oriental	
Mean	H	Mean	H	Mean	H	Mean	H
8.44	1.09	8.27	0.58	8.50	0.45	9.84	1.63
7.40	0.62	8.20	0.52	9.48	0.75	9.40	1.29
8.15	0.91	8.25	0.65	8.65	0.85	8.20	1.39
7.36	0.97	8.14	0.45	8.16	0.50	8.24	1.28
7.60	0.75	9.00	0.56	8.83	0.61	9.20	0.66
8.04	1.02	8.10	0.44	7.76	0.63	8.55	2.06
9.04	2.63	7.20	0.58	8.63	0.93	8.52	1.24
9.76	1.18	8.32	0.79			8.12	1.30
		7.70	0.53				

heterogeneity index H as an additional response variable in human studies.

3. COMBINING INFORMATION FROM A SERIES OF COMPARABLE GENOTOXICITY STUDIES

One of the most intriguing and important types of scientific inference is the formal evaluation of information gathered from a series of comparable, but not truly replicate studies. In the social science literature, where this process of inference is known as meta-analysis (e.g., Light and Pillemer, 1984), there is no restriction to a series of studies that share a common response variable. To date, the biologic and physical sciences have been less adventuresome in this regard; the studies in a series being combined into one inference have typically involved a common end point. An example of meta-analysis in genetic toxicology is provided by the re-examination of the evidence for an SCE sex difference in humans conducted by Margolin and Shelby (1985). They considered a dozen studies from 1975–1983, located by a computerized literature search, that provided adequate information, separated by sex, on SCE levels for normal, healthy, nonsmoking, adult controls. Table 3 is derived from Table I of Margolin and Shelby (1985); the full details of data extraction and all references are in the footnotes to their Table I. Most of the authors referenced by Margolin and Shelby commented on the issue of a sex difference, generally reporting no significant difference found; only the reference labeled Hedner found a significant elevation of female SCE levels over males. Margolin and Shelby comment that "Although numerous authors have addressed the question of a sex difference in human SCE levels, the typical study ··· is small"; the median total sample size is 24. They further note that for the eleven studies in which the 95% confidence limits are computable, these intervals all straddle the value 0.52 SCE/cell, the median of the twelve mean sex differences in Table 3. A Wilcoxon signed rank test applied to these mean differences yields an observed significance level of 0.013, leading to the conclusion that among normal healthy adults, females score higher than males, with an average difference of approximately one-half SCE/cell. Via power considerations, Margolin and Shelby (1985) observe that if the true sex difference were 0.5 SCE/cell, given the sample sizes and inter-subject standard deviations common in the published studies, "the probability that at most one of these 12 studies would obtain the correct inference is a surprisingly large 0.32."

TABLE 3
Estimated SCE/cell differences for 12 human studies

Lead author of study	Sample sizes		Mean SCE sex difference
	Female	Male	
Butler	12	13	0.43
Cohen	12	12	0.75
de Arce	6	6	−0.8
Galloway	2	5	0.44
Hedner	12	23	1.39
Husum	27	33	0.39
Ikeda	5	6	0.4
Latt	8	5	2.18
Livingston	12	12	0.60
Morgan	17	17	0.68
Ohtsuru	2	4	0.66
Waksvik	28	14	−0.2

Clearly, the inference of a small SCE sex effect flew in the face of the then conventional wisdom that no such difference existed—a scientific myth created by a series of small, underpowered studies of a small effect. Fortunately, at the same time that the paper of Margolin and Shelby (1985) was in press, a huge control group from an occupational study of SCE levels in workers exposed to ethylene oxide appeared (Soper, Stolley, Galloway, Smith, Nicholas and Wolman, 1984). With 479 control individuals (269 men and 210 women) available, they were able to establish a statistically significant sex difference in SCE levels, in agreement both qualitatively and quantitatively with Margolin and Shelby's conclusions.

The existence of a sex difference in human SCE levels is of limited scientific importance. The converse myth of no difference, however, can definitely have negative consequences. This is exemplified by an incident involving a researcher who, just prior to the publication of the papers of Soper, Stolley, Galloway, Smith, Nichols and Wolman (1984) and Margolin and Shelby (1985), had her own SCE study's credibility challenged by a reviewer because her data exhibited a significant sex difference in SCE levels, a difference that everyone "knew" did not exist.

The concern regarding a series of small studies with negative findings reinforcing each other and producing a widely quoted myth of no effect assumes greater importance when one reviews the 113 human cytogenetics studies surveyed by Ashby and Richardson (1985). The sample sizes in these occupational studies are little different from those in Table 3, raising questions as to whether these studies had adequate sensitivity for their intended purposes.

4. COMMENTS ON THE UTILITY OF HISTORICAL CONTROL DATA

The role of historical control data in the analysis of human genotoxicity studies is in an early stage of evolution. In in vivo genotoxicity studies, this evolution has progressed much further; for example, historical control data play a dominant role in the

statistical analysis of the mouse morphologic specific locus assay (Russell, 1951), the primary test for mammalian germ cell mutagenicity. In this assay, the response observed is binary and the statistical analysis in common use is a Fisher exact test contrasting the response of the treated group with that of the aggregated historical and concurrent control groups (Selby and Olson, 1981). At this point in the assay's long history, the aggregated historical control data outweigh the concurrent control data to such an extent that the concurrent control is reduced to being solely a quality control check on the execution of the experiment.

Ashby and Richardson (1985) recommend the formation of a central repository for control data bases from human surveillance cytogenetics studies. With this goal in mind, it is important to consider criteria appropriate for judging the acceptability of historical control data from observational studies of genetic toxicity. The following criteria build upon discussions by Pocock (1976) of clinical trials and Margolin and Risko (1984) of laboratory experiments:

1. The historical data must have been gathered by the same research team that is conducting the current study.
2. The study protocol must have remained fixed throughout the period covering the historical studies and the current one. This includes the method of scoring the response.
3. The historical and concurrent control groups must be comparable with regard to age, sex, race, lifestyle factors, socioeconomic status, marital status and geographic location. Where doubt exists as to the comparability of control groups, the burden of proof rests upon the research team desirous of appealing to its historical control database.
4. There must be no detectable systematic differences in response between the various control groups, i.e., differences attributable to known covariates, such as time or technician.

As Margolin and Risko (1984) emphasize, studies forming a history "are ordered chronologically, so it is imperative that one demonstrate stationarity of the response ..." Margolin and Risko illustrate by an example involving lung tumors of female F344 rats from 70 historical control groups just how subtle time effects can be; because time may stand as a surrogate for an unknown influential factor, an assessment of time effects is crucial to the credible use of historical data.

These criteria and brief discussion are intended to focus attention and debate on ground rules for the establishment and use of control data bases. The brevity of treatment here does not adequately reflect the paramount importance of the underlying issues if the area of human genotoxicity studies is to avoid degenerating into methodologic chaos.

CONCLUDING REMARKS

It is highly likely that as the rest of this century unfolds, the interplay of environment and genetics on human health will increasingly be a matter for in-depth study. As the pace of development of new technologic probes for damage to human genetic structures accelerates, numerous challenging research questions will arise. Human monitoring for exposure to and damage from genotoxic agents will command increasing attention, and probably generate increasing controversy. Questions regarding the use of historical control data, adjustment for subject covariates and methods of combining information from a series of studies will come increasingly to the fore. The complex nature of this research strongly suggests the need for greater statistical and epidemiologic involvement in this effort.

ACKNOWLEDGMENTS

This paper was written while the author was Head, Statistical Methodology Section, National Institute of Environmental Health Sciences. Helpful discussions with R. Everson and R. Tice are gratefully acknowledged.

REFERENCES

AMES, B. N. (1979). Identifying environmental chemicals causing mutations and cancer. *Science* **204** 587–593.

AMES, B. N., MCCANN, J. and YAMASAKI, E. (1975). Methods for detecting carcinogens and mutagens with the Salmonella/microsome mutagenicity test. *Mutat. Res.* **31** 347–364.

ASHBY, J. and RICHARDSON, C. R. (1985). Tabulation and assessment of 113 human surveillance cytogenetic studies conducted between 1965 and 1984. *Mutat. Res.* **154** 111–133.

BERNSTEIN, L., KALDOR, J., MCCANN, J. and PIKE, M. C. (1982). An empirical approach to the statistical analysis of mutagenesis data from the Salmonella test. *Mutat. Res.* **97** 267–281.

BUTLER, M. G. (1981). Sister chromatid exchange in 4 human races. *Mutat. Res.* **91** 377–379.

CARRANO, A. V. and MOORE, D. H. (1982). The rationale and methodology for quantifying sister chromatid exchanges in humans. In *Mutagenicity: New Horizons in Genetic Toxicology* (J. A. Heddle, ed.) 267–304. Academic, New York.

HIRSCH, B., MCGUE, M. and CERVENKA, J. (1984). Characterization of the distribution of sister chromatid exchange frequencies: Implications for research design. *Hum. Genetics* **65** 280–286.

HUSUM, B., WULF, H. C. and NIEBUHR, E. (1982). Increased sister chromatid exchange frequency in lymphocytes in healthy cigarette smokers. *Hereditas* **96** 85–88.

LATT, S. A., ALLEN, J., BLOOM, S. E., CARRANO, A., FALKE, E., KRAM, D., SCHNEIDER, E., SCHRECK, R., TICE, R., WHITFIELD, B. and WOLFF, S. (1981). Sister-chromatid exchanges: A report of the Gene-Tox Program. *Mutat. Res.* **87** 17–62.

LIGHT, R. J. and PILLEMER, D. B. (1984). *Summing Up: The Science of Reviewing Research.* Harvard Univ. Press, Cambridge, Mass.

MARGOLIN, B. H., KAPLAN, N. and ZEIGER, E. (1981). Statistical analysis of the Ames Salmonella/microsome test. *Proc. Nat. Acad. Sci. U.S.A.* **78** 3779–3783.

MARGOLIN, B. H., KIM, B. S. and RISKO, K. J. (1986). The Ames Salmonella/microsome mutagenicity assay: Issues of inference and validation. Submitted for publication.

MARGOLIN, B. H., RESNICK, M. A., RIMPO, J. Y., ARCHER, P., GALLOWAY, S. M., BLOOM, A. D. and ZEIGER, E. (1986). Statistical analyses for in vitro cytogenetics assays using Chinese hamster ovary cells. *Environ. Mutagen.* **8** 183–204.

MARGOLIN, B. H. and RISKO, K. J. (1984). The use of historical data in laboratory studies. *Proc. Internat. Biometrics Conference* **12** 21–30.

MARGOLIN, B. H. and SHELBY, M. D. (1985). Sister chromatid exchanges: A reexamination of the evidence for sex and race differences in humans. *Environ. Mutagen.* **7** 63–72.

MORAN, P. A. P. (1973). Asymptotic properties of homogeneity tests. *Biometrika* **60** 79–85.

MYERS, L. E., SEXTON, N. H., SOUTHERLAND, L. I. and WOLFF, T. J. (1981). Regression analysis of Ames test data. *Environ. Mutagen.* **3** 575–586.

POCOCK, S. J. (1976). The combination of randomized and historical controls in clinical trials. *J. Chronic Dis.* **29** 175–188.

RUSSELL, W. L. (1951). X-ray-induced mutations in mice. *Cold Spring Harbor Symp. Quantitative Biology* **16** 327–336.

SELBY, P. B. and OLSON, W. H. (1981). Methods and criteria for deciding whether specific-locus mutation-rate data in mice indicate a positive, negative, or inconclusive result. *Mutat. Res.* **83** 403–418.

SNEDECOR, G. W. and COCHRAN, W. G. (1967). *Statistical Methods,* 6th ed. Iowa State Univ. Press, Ames, Ia.

SOPER, K. A., STOLLEY, P. D., GALLOWAY, S. M., SMITH, J. G., NICHOLS, W. W. and WOLMAN, S. R. (1984). Sister-chromatid exchange (SCE) report on control subjects in a study of occupationally exposed workers. *Mutat. Res.* **129** 77–88.

STEAD, A., HASSELBLAD, V., CREASON, J. and CLAXTON, L. (1981). Modeling the Ames test. *Mutat. Res.* **85** 13–27.

VAINIO, H., SORSA, M. and FALCK, K. (1984). Bacterial urinary assay in monitoring exposure to mutagens and carcinogens. In *Monitoring Human Exposure to Carcinogenic and Mutagen Agents (IARC Sci. Publ.)* (A. Berlin, M. Draper, K. Hemminki and H. Vainio, eds.) **59** 247–258. IARC, Lyon.

YAKOVENKO, K. N. and PLATONOVA, V. I. (1979). Spontaneous level of sister chromatid exchanges and their distribution in human cells. *Soviet Genetics* **15** 746–753.