Inter-rater Agreement in Assessing Occupational Exposure in a Case-control Study

by

© Mark S. Goldberg

A thesis submitted to the Faculty of
Graduate Studies and Research in
partial fulfillment of the requirements
for the degree of Master of Science.

Department of Epidemiology and Biostatistics
McGill University, Montreal

January 1985.

(i) **Abstract**

The accuracy of attributing exposure in occupational environments was assessed within the context of a case-control study (the Cancer Study) that investigated associations of different sites of cancer with a multitude of substances found in work situations. A team of chemists/engineers attributed exposure after reviewing job descriptions that were usually obtained by interviewing Cancer Study subjects. The present study investigated the validity and reliability of the chemists' exposure assessments through seven trials of inter-rater and intra-rater agreement. Comparisons were made 1) among the chemists and 2) between the chemists, considered singly or in groups, and external raters who possessed expert knowledge of industrial environments. In all trials, the agreement in identifying whether an exposure was present or absent was good to excellent (average Kappa $\pm$ S.D.; $\bar{\kappa}=0.59\pm0.07$). Agreement in judging the intensity of exposure was somewhat lower ($\bar{\kappa}=0.50\pm0.07$). The accuracy of attributing exposure by individual chemists and by chemists who used a consensus method was inferred to be quite good.

(ii) **Résumé**

L'exactitude de l'attribution d'exposition dans les environnements occupationnels a été évalué dans le contexte de l'étude cas-témoine qui examinait les associations entre différents sites de cancer et une multitude des substances que l'on trouve dans les milieux de travail. Les descriptions d'emploi obtenues sont revisées par une équipe de chimistes/ingénieurs utilisant une méthode de consensus pour attribuer l'exposition aux substances. La présente étude examine la validité et la fiabilité des evaluations des chimistes en utilisant sept essais de concordance inter-evaluation et intra-evaluation. Les comparisons de l'attribution des expositions ont été effectuées 1) parmi les chimistes et 2) entre les chimistes, considéré individuellement ou en groupes d'une part, et des évaluateurs externes possédent connaissances spécialisées des environnements industriels d'autre part. Dans tous les essais la concordance pour l'identification de la présence ou de l'absence l'exposition varie de bonne à excellente (moyen Kappa $\pm$ S.D; $\bar{\kappa}=0.59\pm0.07$). La concordance pour le jugement de l'intensité de l'exposition est quelque peu plus faible ($\bar{\kappa}=0.50\pm0.07$). Nous pouvons déduire que l'attribution de l'exposition par des chimistes individuelle et par des chimistes utilisant une méthode de consensus est assez bonne.

(iii) **Acknowledgements**

The author would like to thank Dr Jack Siemiatycki for suggesting the topic and for supervising the thesis. Dr Siemiatycki provided an enormous amount of logistic and technical support throughout the course of the work. Access to data, computer facilities, resource persons, and interviewers were provided. He tirelessly read the manuscripts and offered many invaluable suggestions regarding methodology, organization of the dissertation, and interpretation of the results.

This thesis could not have been completed without the help of Dr Michel Gérin. Along with Dr Siemiatycki, he designed four of the seven trials reported here. He helped to organize the three trials designed by the author by contacting the appropriate resource people. He coded exposure in four of the seven trials and assisted the external raters in two other trials. Dr Gérin thoughtfully read the manuscript and provided insight regarding the interpretation of the results.

Dr Jim Hanley provided statistical support. He read Chapters 3 and 6 and Appenix 5 and offered suggestions regarding the interpretaion of the statistical methods used herein.

The author would like to thank Drs Graham Gibbs, Gilles Lebeau and M. Jacques Guénette for permitting access to industrial sources of data. In particular, M. Guénette arranged interviews at two industrial factories. The author would also like to thank all of the chemical coders: Denis Bégin, Jean-Paul Boillot, J Doré, Michel Galopin, Dr Joseph Hubert, Howard

Kemper, Lucien Laroche, Denis Lessard, Christian Millet, Dr Peter Nichol, Roger Tremblay, and Peter Wrezsian.

The present study was conducted within the framework of a large case-control study based in Montreal (the Cancer Study). The author is indebted to the encouragement and support provided by many individuals associated with the Cancer Study. Lesley Richardson, who was manager of the Cancer Study, provided a great deal of technical support. Her encouragement and many suggestions are gratefully acknowledged. Denise Bourbonnais performed the interviews and selected job descriptions that were used in three of the trials. The procedure for using a statistical package to initially prepare the raw data for processing by a computer program was suggested by Ron Dewar. Special thanks are due to the computer operations group at l'Institut Armand-Frappier for their support. In particular, the author is indebted to Jean Pellerin for his technical assistance.

Since the development, exceution and completion of a thesis is usually a long and sometimes painful process, it is customary to thank one's cohabitor for her moral support, encouragement and patience throughout the process. In this instance, the author finds it difficult to find the appropriate adjectives to express his deepest gratitude to Nancy Mayo. Nancy spent an immeasurable amount of time on my thesis - time that could have been better spent completing her PhD thesis. She read and edited the text more times than I would like to think. Her valuable and critical suggestions greatly enhanced the quality of the work. Her patience in putting-up with my occassional bouts of sullen humour was remarkable.

The author gratefully acknowledges receipt of personal support from Les Fonds F.C.A.C. pour l'aide et le soutien à la recherche, l'Institut de recherche en santé et en sécurité du travail du Québec, and McGill University.

(iv) **Table of Contents**

Table of Contents, continued

Table of Contents, continued

Table of Contents, continued

Table of Contents, continued

Table of Contents, continued

Table of Contents, continued

Table of Contents, continued

Table of Contents, continued

Table of Contents, continued

Table of Contents, continued

(1) Introduction

One goal of epidemiologic research is to identify factors that are determinants of disease. This line of research is important if disease mechanisms are to be understood and if appropriate preventive steps are to be taken. It is crucial, therefore, to evaluate the accuracy of the epidemiologic data used to establish associations between disease and exposure.

Various mechanisms may invalidate the results of an investigation (cf, Kleinbaum et al., 1982). The internal validity of any epidemiological investigation requires, among other things, accurate ascertainment of outcome, such as symptoms, disease, or death, and of the factors that may be associated with outcome. This latter group of variables includes confounding factors and those factors that are under investigation; ie, the exposure variables (Last, 1983). Even though considerable effort is made in epidemiologic studies to obtain accurate measures of outcome and exposure, errors in measurement can still occur which may invalidate the results of the investigation.

The accuracy of the ascertainment of outcome and exposure partly depends on the design of the study. For example, the determination of disease outcome through retrospective cohort and case-control studies may be inaccurate because of changes in the classification of diseases or in diagnostic criteria. Similarly, the accuracy of measuring an exposure that occurred in the past will depend on many factors, such as the setting for the study and the accuracy of records and of information obtained from respondents.

Some data can be used to indicate whether a group of individuals having a common experience are at higher risk. For example, company records detailing worker employment may be highly accurate for placing workers in job categories. These records, however, may not provide adequate information for determining precise exposure to chemical agents. In general, the quality of exposure data will diminish as one looks further back into the past. This is particularly relevant for studies that investigate diseases which take a long period of time before expression, such as most cancers.

Two general types of errors arise when measuring exposure. The first is due to random fluctuations about the true exposure values. The second occurs when the measurements are systematically higher or lower than the true values. Both of these errors, operating singly or together, may alter the estimate of the association between the exposure and the disease. In case-control studies, for example, if the only errors in measuring exposure are random and if the error rates are identical in both diseased and non-diseased groups, the estimated excess risk will be shifted toward the null value (cf, Barron, 1977; Copeland et al., 1977; and Greenland, 1980). This is known as non-differential, random, misclassification bias, or simply, misclassification bias.

(1.1) Occupational Epidemiology

There is usually a great deal of uncertainty in identifying the precise agents that existed in occupational environments. Most occupational studies, therefore, have used job records to indicate whether persons employed in a particular job or industry were suffering adverse health effects as a result of their employment. This has been a profitable endeavour since there have been many reports of excess risk of disease in particular occupational groups. For example, excess risks for lung cancer have been found among steel foundry workers (Blot et al., 1983).

Use of job titles, however, may obscure an association if not all workers with a given title are exposed to the active agent (Hoar et al., 1980). Even if an excess risk of disease was detected for an occupational group, it would be difficult to infer which of the substances was the responsible agent (Siemiatycki et al., 1981a). In addition, the use of job titles may result in a loss of statistical power when occupations having a common exposure are not combined (ibid.). In recent years, there have been more attempts to go beyond job titles and to identify and quantify the extent of exposure to substances. This is a difficult task which requires novel methods (egs, Mancuso et al., 1977; Hoar et al., 1980; Beaumont and Weiss, 1981; Lubin et al., 1981; Siemiatycki et al., 1981b, 1982; and Delzell and Monson, 1982). The credibility of results based on these new methods depends on the accuracy of the exposure measurements. The assessment of the accuracy of one of these methods (Siemiatycki et al., 1981a) is the subject of this thesis.

(1.2) Evaluating the Accuracy of Measuring Instruments

The accuracy of any instrument may be evaluated in terms of two components. The first component is whether the instrument actually measures what it is intended to measure. This is known as validity. Precision, or repeatability, of the measuring instrument is the second component of accuracy. This latter component is also known as reliability.

Various types of validity have been defined (Abramson, 1979). Face validity refers to the belief of the investigator that the instrument under consideration evidently measures what it is intended to measure. If many other "experts" believe that the measure is credible then the measure is said to have consensus validity. Criterion validity, on the other hand, refers to the correlation of the measure under consideration with one that has higher face or consensus validity. Certainly, the best criterion is the true value of the variable. In the absence of knowledge of the truth, other measures which have been shown or are believed to have a high level of validity may be used as a comparison. Criterion validity can be further distinguished along temporal lines. Thus, concurrent validity refers to the correlation with another instrument when both measurements are made at the same time. Predictive validity, on the other hand, is determined by the ability of the instrument under evaluation to predict some other measure.

The determination of the criterion validity of an instrument that measures past occupational exposure is a difficult matter since another independent instrument, which has been shown or believed to be valid, must be used as a

comparison. Often, it is impossible to find even one instrument to measure past exposure. The ascertainment of the reliability of the measuring instrument may also be difficult since multiple measurements of exposure for all, or a subset, of the observations must be made.

The difficulty of evaluating validity has led some investigators to use, as a surrogate, the agreement obtained by comparing two or more instruments, neither of which could be considered a validated criterion. For example, in the case of evaluating X-ray films for pneumoconiosis (eg, Liddell, 1963), the comparison of the judgements of many raters (ie, *inter-rater agreement*) was used to indicate the validity of the approach.

(1.3) The "Cancer Study"

This thesis is concerned with the accuracy of the assessment of exposure in a multi-cancer site, multi-exposure, case-control study known as "Monitoring the Occupational Environment for Carcinogens" (hereafter, this will be referred to as the *Cancer Study*; Siemiatycki, 1979; Siemiatycki et al., 1981a; and Gérin et al., 1983b). The Cancer Study was designed to generate hypotheses concerning the potential association of 13 cancer sites with approximately 300 substances found in occupational environments. Male patients, positively diagnosed with cancer in Montreal hospitals, were entered into the Cancer Study if certain eligibility criteria were satisfied. Questionnaires were administered for the purpose of obtaining information concerning all prior occupations. For each occupation, specific information was sought concerning the exact tasks performed by the subject, materials and machines used, and the general working environment.

Using this data, a team of *internal raters* consisting of chemists, industrial hygienists, and engineers determined whether the subject, in each of his occupations, was exposed to any of the substances under consideration. (These internal raters will be referred to as *chemists* in this thesis.) A consensus approach to attributing exposure was used by the team of chemists. Level of exposure was also assessed using semi-quantitative scales.

(1.4) Objective of the Thesis

A series of trials were designed to investigate the accuracy of the chemists' assessment of exposure. (The use of the word "trial" in this thesis refers to an experiment designed to evaluate the accuracy of the chemical coding of the chemists.) The degree of inter-rater and intra-rater agreement in the attribution of exposure was assessed between individual chemists and between chemists, considered singly or in groups, and *external judges* who had expert knowledge of certain industrial environments.

This thesis reports on seven trials that were designed to meet the objective of the study. Not all of these trials were designed and implemented by the author. The principal investigators of the Cancer Study (Drs. Siemiatycki and Gérin) designed four of the trials while the author was responsible for designing and implementing the remaining three. All trials, however, have been analysed by the author.

(1.5) Outline of the Thesis

Chapter 2 reviews published epidemiological studies that have investigated the reliability, validity, and inter-rater agreement of ascertaining past exposure in occupational environments. In addition, a review of the medical and epidemiologic literature related to the ascertainment of agreement is also presented. Only studies that were directly relevant to this thesis will be discussed. Statistical methods used to evaluate agreement will be presented in Chapter 3.

The objective of this thesis and the means by which this objective were met are presented in Chapter 4. A brief description of the seven trials will also be presented.

Chapter 5 is concerned with describing the materials that were common to each of the trials. An overview and description of the relevant features of the Cancer Study is presented first. Following that, the professional and educational background of each of the chemists and the external raters are discussed.

Chapter 6 outlines the general methods used in the trials. Methods of coding exposure are explained first. This is then followed by an overview of the statistical procedures. Lastly, the computer program that was specifically written by the author for use in this thesis will be briefly described.

In Chapter 7 the results of each trial is reported. The trials are grouped

according to whether comparisons were being made: 1) between individual chemists, 2) between individual chemists and external judges, 3) between chemists, coding exposure by consensus, and external judges and 4) for a code/recode of identical files by the chemists using the consensus approach. The format of presentation is identical for each trial. A brief recapitulation of the purpose of the trial, a discussion of the materials and methods relevant to the trial, and the results are presented. The results of the seven trials are synthesized in the final sections of this chapter.

A discussion of the results and possibilities for further research are presented in Chapter 8.

A glossary of terms and abbreviations is set out in Appendix 1. Those terms that are particular to this thesis are defined in the glossary. The reader is cautioned, however, that not all technical terms are defined in the glossary or in the text. Reference can be made to Last, 1983, for those terms that are in common use. The contents of Appendices 2 through 6 are shown in the Table of Contents. Supplementary data related to the analysis of each trial will be found in Appendices 7 through 13.

(2.0) Review of the Epidemiologic Literature

(2.1) Introduction

Two types of epidemiologic studies were included in this review: a) those that measured reliability, validity, or inter-rater agreement of determining past occupational exposure; and b) those that studied inter-rater agreement in other areas of research that could be relevant to this thesis.

A review of the statistical methods used to ascertain agreement between raters judging qualitative data is presented in Chapter 3.

(2.2) Reliability and Validity of Occupational Exposure Data

Jarvholm et al., 1981, retrospectively studied the relation of oil mist to cancer morbidity for those individuals employed in the turning and grinding departments of a Swedish factory that manufactured bearing rings. Exposure to oil mist was determined for each individual in the cohort according to the calendar year of employment. Average exposure was estimated for those years prior to and subsequent to the installation of exhaust ventilation equipment on certain pieces of machinery. In the latter instance, exposure was estimated by measuring the density of oil mist when the ventilation equipment was operating. It was found that these measurements agreed with those recorded from previous surveys. For the time period prior to the introduction of exhaust equipment, exposure was ascertained by measuring ambient levels when the ventilation equipment was disabled. To validate these latter measurements, men with over 20 years of experience were

exposed to oil mist in a sealed room. The density of the mist in the room was about as high as that measured when the ventilation equipment was disabled. A questionnaire was then administered to determine whether the atmosphere in the room resembled past conditions. Ninety percent of these workers indicated that the air was mistier in the past than in the present.

In another study of the validity of past exposure data, Pershagen and Axelson, 1982, compared the results of two case-control studies at a Swedish copper smelter (Axelson et al., 1978; Pershagen, 1978). The objective of both of these studies was to determine the excess risk of lung cancer for workers exposed to inorganic arsenic. Cases and controls in both studies were derived from the same sampling frame using two sources of data. In the study by Pershagen (Study 1), cases and referents who resided in the area near the smelter were selected from a national mortality registry. In the study by Axelson and co-workers (Study 2), subjects were selected from a local parish register. Exposure to arsenic was assessed differently in the two studies. In Study 1, interviews of next-of-kin were employed to elicit information concerning the subject's employment and possible exposure to arsenic. In Study 2, company records and exposure measurements were utilized. The sensitivity (Se) and specificity (Sp) of responses to specific questions was assessed using company records as the criterion. The Se and Sp of responding to the question which asked whether the worker was employed at the smelter was about 98%. The values of Se and Sp for the question which inquired whether the worker was heavily exposed to arsenic was estimated to be 46% and 91%, respectively. The value of the odds ratio (OR) for exposure to high levels of arsenic was lower (OR=7.3) than the value of 10.1 found from using company records. The results of

this study indicate that questionnaires are useful in assessing past exposure, but their use may lead to an underestimation of excess risk.

Macaluso et al., 1983, conducted a case-control study of eight occupational exposures in relation to respiratory cancer. Detailed descriptions of each subject's occupations were obtained by interview. Two different systems for inferring past exposure were used and the relative risks derived from the two measurements were compared. The first system made use of an a priori job/exposure matrix (cf, Hoar et al., 1980; Hoar, 1983; and Hsieh et al., 1983) which had been designed by the investigators. The team was composed of chemists, occupational health physicians and epidemiologists. The second method evaluated exposure by the same team of investigators on a subject-by-subject basis.

Exposure was assigned on a five-point scale ranging from no to high exposure. An analysis of 698 job descriptions coded for asbestos and polycyclic aromatic hydrocarbons (PAH) was reported. The frequency of exposure was higher for the job/exposure matrix system - 39% vs 25% for asbestos and 54% vs 43% for PAH. For both exposures, use of the job/exposure matrix yielded a larger number of occupations coded in the "low" exposure categories. About the same number of occupations were coded as highly exposed to asbestos. The team of raters, however, coded 27% more occupations as being highly exposed to PAHs. Fair agreement, for attributing exposure to both asbestos and PAH on the full five-point scale, was observed for the two systems of measurement (average index of crude agreement, $\bar{P}_o$ = 60%; average Kappa($\bar{\kappa}$) = 0.34). As expected, there was better agreement when a collapsed two-point scale (categorized as exposed

and not exposed) was used ($\bar{p}_e$ = 79%; $\bar{r}$ = 0.56). Using the assessments of exposure by the panel of raters as a criterion, the sensitivity and specificity was evaluated for the job/exposure matrix. For asbestos, the Se and Sp was 85% and 77% (κ=0.53), respectively; for PAH's the Se was 88% and the Sp was 73% (κ=0.58). The exposures derived by the panel of raters consistently yielded higher ORs than did the job/exposure matrix. For any exposure to asbestos, the team estimated an OR of 1.9 versus 1.4 for the matrix system - for any exposure to PAH the odds ratios were 1.9 and 1.1. Better agreement (using κ) between the two measuring instruments was observed for exposure to PAH. Yet, the difference in the two odds ratios for exposure to PAH was larger than that observed for asbestos. This effect may be due to the manner in which the final exposure assessment for each individual was derived from the job titles. A more likely possibility is that agreement may not be correlated with differences in estimates of relative risk in a simple way. Other factors, such as the prevalence of exposure and the sensitivity and specificity of the measuring instrument, may play a considerable role in the relation between agreement and differences in risk ratios.

The validity of an exposure assignment scheme designed for routinely recording occupational exposure to potential carcinogens in a chemical plant was reported by Greenberg and Tamburro, 1981. Exposure to 22 chemicals used in the manufacture of synthetic rubber and plastics was under consideration. Exposure was asssigned to each worker on the basis of calendar year of employment, the area in the plant in which he worked, and the specific job function that was performed. Experienced senior employees from the company were chosen from each area of the plant to rate exposure

to these substances on a seven-point, rank-ordered scale. These judges met as a group and exposure assessments were made by consensus. Cumulative exposure to each substance was estimated by multiplying the exposure rank by the time period and then summing over all time periods. The validity of the measurements was indirectly inferred by examining exposure to vinyl chloride in four cases of hepatic angiosarcoma. It was postulated that since vinyl chloride was a known risk factor for hepatic angiosarcoma, that the average cumulative exposure based on the measurement system would be higher among cases than controls. Such differences were, in fact, observed for vinyl chloride and for two other substances that were used in the processing of vinyl chloride. The authors concluded that the exposure classification system was valid.

Soskolne, 1982, examined inter-rater and intra-rater agreement for assessing exposure to eight substances at a petrochemical plant in the U.S. A six-point, ordinal scale, was used to indicate the degree of exposure to each of these substances. The employment record of each subject was obtained from company records. Each job was described by the location in the plant, the specific occupation, and the calendar year of the job. (This was referred to as a "job title".) A panel of raters, consisting of an industrial hygienist and one of the researchers, attributed exposure to each job title for the purpose of obtaining estimates of excess risk. Subsequent to this, a 10% random sample of all job titles was taken so that inter-rater and intra-rater agreement could be studied. A second panel composed of five senior company supervisors having 28-40 years experience evaluated this subset of job titles. The industrial hygienist who participated in the original assessments of exposure also coded exposure to

these job titles.

The comparison of the exposure assessments between the first panel and the industrial hygienist, which approximately measured intra-rater agreement, showed little agreement above what would be expected by chance alone when the full six-point scale was used (average κ over all substances, $\bar{\kappa}$ = 0.13). Agreement improved somewhat when a collapsed three-point scale was utilized ($\bar{\kappa}$=0.26). Agreement between the first and second panels was also poor ($\bar{\kappa}$= 0.25).

Agreement improved substantially ($\bar{\kappa}$=0.6) when exposures which were rated "medium" in *either reading* were excluded from the analysis. The author concluded on this basis that a scale which excluded the medium exposure category might be preferable to a graded one. This conclusion is not justified since it is based on comparing exposures that were only judged high or low in either reading. For example, a job description would be excluded from the analysis if it was judged low in the first reading but medium in the second. Thus, a highly biased subset was chosen for comparison. It is therefore not surprising that the degree of agreement was substantially enhanced.

The poor level of agreement observed in these comparisons may have been due to the inexperience of the panels in attributing exposure or to an inconsistent use of the coding criteria. The paucity of detail provided in each job title could also have been a source of substantial variability. The conclusion that a scale which excluded the "medium" exposure level was preferable to a polytomous scale is not justified on the basis of the

analysis previously discussed. As expected, the results indicate that there was better agreement when the categories of the scale were combined in the analysis.

(2.3) Inter-rater Agreement in Other Areas of Epidemiology

Studies of agreement between raters who assessed a trait on a categorical scale have been reported in various fields of epidemiology and medicine. Many of the pioneering studies involved in assessing pneumoconiosis in miners through the use of X-ray films (cf, Liddell, 1961, 1963, 1972, 1974, 1977 and 1980; Liddell and Lindars, 1969; Rossiter, 1972; Felson et al., 1973; Copland et al., 1981; and Musch et al., 1984). There have also been numerous studies of psychiatric assessments of mental illness (egs, Spitzer et al., 1967; Fleiss et al., 1972; and Spitzer and Fleiss, 1974). Reviews of other works will be found in Fletcher and Oldham, 1964, and Koran, 1975a,b.

(2.3.1) The Development of a New Scale for Determining the Extent of Pneumoconiosis in Miners

The work on radiographic classification is relevant to this thesis because of the methods used to develop a more accurate scale for judging the extent of pneumoconiosis. In what follows, a description of this process will be traced by reviewing the work of Liddell, 1963.

The development of a new scale for classifying the extent of pneumoconiosis, as measured by the profusion of small opacities on chest

films, began in the late 1950s when the National Coal Board of Great Britain began monitoring more than one-half million coal miners with X-ray equipment. The goal of the surveillance program, as described by Liddell, was to detect the extent of pneumoconiosis in workers and to monitor the dust suppression program at each colliery. The last objective was met by determining whether pneumoconiosis in these miners was progressing or not. At that time, a four-point scale of the International Labour Organization (ILO) was in use. It was the belief of some researchers that the profusion of small opacities could be classified more accurately if a more continuous scale was used. Thus, a new twelve-point scale was introduced by Liddell which was eventually adopted by the International Labour Organization. (This scheme is now known as the UICC/Cincinnati classification system.)

In this scheme, the film would be classified into the major (or "formal") ILO category that was considered the more likely. The adjacent ILO category that had been given secondary consideration would also be indicated. A reading of 1/2, for example, indicated that the rater classified pneumoconiosis into category 1, but seriously considered category 2 as an alternative. Since the formal categories of the ILO scheme were retained in the experimental system, a comparison of the two systems could be made directly by collapsing the new scale into the four formal categories.

In 1963 Liddell undertook the first of a series of experiments in film reading. The objective of this experiment was to demonstrate that the experimental classification system, having 12 categories, was as reliable as the one sanctioned by the ILO. Validation of the new scale was achieved

through the assessment of inter-rater agreement. Two sets of films were used in the trials. The first set had been previously read by the raters using the ILO classification scheme only. These films were not re-read using the new system. Another set of films was read three times by the readers using the new system only.

Intra-observer agreement was measured for both sets of films and compared. For these purposes, the readings of the second set of films, which were read using the experimental system, were collapsed to the equivalent four-point scale of the ILO system. It was found that the degree of intra-observer agreement was higher when the new classification system was used.

Inter-rater agreement, using only the second set of films, was assessed in a variety of ways. Comparisons were made of the second reading of each judge with that obtained from a consensus of the three. The consensus version was derived in a joint session after the films had been read for a third time. The second method was to compare these assessments with the median value obtained from the three independent readings made just prior to the joint session. Finally, direct comparisons of the second reading of each observer were also made. In all cases, the experimental classification showed the least amount of inter-observer error.

From these experiments, Liddell was able to conclude that the new sixteen-point scale was more reliable than the ILO system. Confirmation of this conclusion was obtained in further experiments (op. cit.).

(3.0) A Review of Statistical Methods for Measuring Agreement Between Raters who are Judging Qualitative Data

In this chapter, the statistical methods used to describe and estimate the extent of agreement between raters who assess a qualitative variable are presented. The intent is to provide insight concerning the interpretation, the strengths and the weaknesses of the methods that will subsequently be used in this thesis. For a full account of the procedures used in measuring agreement for continuous and discrete data, the reader is directed to the many reviews which have appeared recently (eg, Landis and Koch, 1975, Wolfson, 1978, Fleiss, 1981, and Kramer and Feinstein, 1981).

(3.1) Measuring Agreement between Two Raters Judging an Attribute on a Two-Point Scale

The simplest situation arises when two raters assess a trait of n subjects on a scale consisting of two mutually exclusive categories (a dichotomous or two-point scale). The joint distribution of rater assessment, which is obtained by crosstabulating the judgements of both raters, is presented as a standard two-by-two contingency table (Table 3.1). The notation used in the table is that of Fleiss, 1981. The entry in each cell of the table refers to the proportion of subjects in the table with the appropriate combination of attributes, present or absent, as determined by the two raters. For example, the proportion of subjects scored "present" by both raters is denoted by p_{11}. The proportion of all subjects judged positive (negative) by Rater 1 is $p_{1.}$ ($p_{1.}$), and by Rater 2 is $p_{.1}$ ($p_{.1}$). These proportions, which are known as the *marginal proportions*, are obtained from

the table by summing the cells in the "present" (or "absent") category; thus, $p_{2.} = p_{21} + p_{22}$.

Table 3.1

Description of the Data Arising from Two Raters Classifying a Trait on a Two-Point Scale

		Rater - 2		
		Absent	Present	
Rater - 1	Absent	p_{11}	p_{12}	$p_{1.}$
	Present	p_{21}	p_{22}	$p_{2.}$
		$p_{.1}$	$p_{.2}$	1

The assessments of the trait by the two raters are said to be (design) independent if, for each subject, each rater classifies the trait without knowledge of the other rater's assessment. If, in addition, both raters employ different sets of criteria, then the observed agreement would be explainable by chance alone. If this is true, then the assessments are statistically independent. Thus, the expected value for any cell can be determined, using the multiplicative law of independent events, by simply multiplying the appropriate proportions in the margins. For example, the expected value of p_{11}, $E(p_{11})$, is calculated from the formula $E(p_{11}) = p_{1.} \times p_{.1}$.

The assessment of agreement (or disagreement) can be framed in terms of three questions. The first question asks whether there are differences between the observers in their estimates of the prevalence of the trait. If the answer is in the affirmative, then it is of interest to determine whether the differences are statistically significant. The second question inquires whether the extent of agreement observed in the table is due solely to the two raters randomly allocating subjects into these cells. This question is different from the first in that it addresses the problem of *testing* whether the agreement observed *in* the table (ie, in the p_{11} and p_{22} cells) is above what would be expected by chance. The third question, which is related to the second, asks whether the extent of concordance or discordance in the table can be measured using summary indices of agreement.

(3.1.1) Testing whether there are Differences between Raters in their Estimates of Prevalence

An estimate of the frequency, or prevalence, of the trait in the study population can be obtained for each observer; namely $p_{1.}$ for rater-1 and $p_{.1}$ for rater-2. These estimates can be compared, and their differences statistically tested using McNemar's test of non-independent samples (see Bennett, 1967). The test statistic, which is asymptotically distributed as a chi-square on one degree of freedom (df), is calculated as:

$$\chi^2 = \frac{n \times (p_{12} - p_{21})^2}{p_{12} + p_{21}} .$$

(3.1.2) Testing Whether Agreement is Greater than that Expected by Chance

It is instructive to display the hypothetical data of Table 3.1 in a different format, as shown in Table 3.2. Here, the two discordant cells, p_{12} and p_{21}, are combined into one. Observed and expected values for each cell can be obtained from Table 3.1. The expression, *pattern of concordance*, (Bennett, 1972; Siemiatycki et al., 1982) will be used to refer to this tabular presentation of the joint distribution of rater assessment.

Table 3.2

Observed and Expected Values of the Pattern of Concordance for n Subjects Categorized on a Two-Point Scale by Two Raters

No of Raters Attributing the Factor to be:		Observed	Expected
Present	Absent	Number	Number
2	0	$n \times p_{22}$	$n \times p_{2.} \times p_{.2}$
1	1	$n \times (p_{12} + p_{21})$	$n \times (p_{.1}\, p_{2.} + p_{.2}\, p_{1.})$
0	2	$n \times p_{11}$	$n \times p_{1.} \times p_{.1}$
	Total:	n	n

The format of the data presented in Table 3.2 suggests a comparison between the observed and expected numbers. Such a comparison can take several forms. For instance, the equality of the observed and expected values in the concordant cells (2-0 and 0-2) can be statistically tested. The null hypothesis for this test states that the sum of the two concordant cells is equal to the expected sum; i.e.,

$$H_0: p_{11} + p_{22} = E(p_{11} + p_{22}) = p_{1.} p_{.1} + p_{2.} p_{.2}.$$

The test can be performed by assuming that the sum of the two concordant cells, $p_0 = p_{11} + p_{22}$, is distributed as a binomial variable. (This sum is also known as the index of crude agreement.) The two independent parameters that specify the distribution are the probability of an event occurring in any given trial, p, and the number of trials, n. A trial is defined when both raters judge a randomly selected subject from the study population. An event is defined when both raters judge the trait either present or absent. The expected probability that the two raters will agree that the trait is either present or absent, p, is calculated from the formula $p = E(p_{11} + p_{22})$. The total probability of observing a value of p_0 greater than p, can be obtained by summing all values of the distribution greater than the observed p_0. The normal distribution can be used as an approximation to the binomial when the product of the number of subjects, n, and the index of crude agreement, $p_0 = (p_{11} + p_{22})$, is large. Since the square of any standard normal variable is distributed as a chi-square, then the usual chi-square test on 1 degree of freedom can also be used. The latter test is simpler to perform arithmetically. For the concordant and discordant cells, the difference of the observed and expected numbers is squared and then divided by the expected value. The required statistic is formed by summing these two values.

The extension to more than two raters follows by applying the fundamental theorems of probability to calculate the expected numbers. For three raters, the pattern of concordance will have four cells represented by the number of subjects in which all three raters agree that the trait is present (3-0) or absent (0-3), where two raters attribute the presence of the trait and one does not (2-1), and where one rater judges the trait present while the other two claim it is absent (1-2). Let p_i represent the total proportion of subjects judged present by rater i. The expected value for the 2-1 cell, for example, can be calculated by summing all three products having the form $p_i x\ p_j x\ (1-p_k)$, $i \neq j \neq k$. That is, the expected number is calculated as

$$p_1 \times p_2 \times (1-p_3) + p_1 \times p_3 \times (1-p_2) + p_2 \times p_3 \times (1-p_1).$$

As in the two rater situation, a statistical test can also be performed to determine whether the number of observations in the concordant cells is greater than that expected by chance.

(3.1.3) Indices Used to Measure Agreement

A number of indices are available which summarize the extent of agreement by considering the joint and marginal proportions of Table 3.1. These measures of agreement may be classified into two broad categories: a) indices that combine the values in the cells and the margins without referring to expected values and b) those that take into account agreement that would be expected by chance alone. These latter measures are referred to as *chance-corrected indices*.

(3.1.3.1) Summary Indices that do not Adjust for Chance Agreement

Table 3.3 lists eight measures that do not correct for expected agreement and three that take chance agreement into account. The simplest and most widely used measure is the index of crude agreement, or overall proportion of agreement, p_0, which simply sums the proportions in the two concordant cells. This index, therefore, measures the extent of agreement by equally weighting each of the concordant cells. It has been suggested by some authors (cf, Rogot and Goldberg, 1966) that chance agreement must be taken into account when using this measure. Thus, these authors developed the A_1 index whose main feature is that its expected value is equal to 0.5, regardless of the values of the marginal proportions. The proportion of specifc agreement, p_s, represents the conditional probability that a second rater attributes the trait present given that a randomly selected first observer also made the same attribution (Dice, 1945). This index may be used when the trait is judged to be absent more often than present. In effect, concordance is treated asymmetrically in that it is evaluated with respect to the value of the p_{22} cell, independent of the p_{11} cell. In a similar manner, the index p_s' tests concordance given that the trait is judged to be present more often than being absent. Kendall and Stuart, 1961, proposed the theta and phi indices which both measure the strength of the association between the ratings of two observers. The phi or tau-b index, which is proportional to the square root of the Pearson Chi-square, also reflects the amount of clustering about the concordant cells (Landis and Koch, 1975).

A measure, which has been used in cybernetic theory, has been introduced as

an agreement statistic by Liddell, 1963, and Liddell and Lindars, 1969. It is called the average amount of information transferred, I. Higher values of the statistic indicate that larger amounts of information are transferred from rater to rater which, in turn, indicates less inter-observer error. Thus, the inverse of I is a measure of the amount of inter-observer error. I is calculated from the formulae:

$$I = A - B$$

where $$A = \sum \left(p_{ij} \times \log_2(p_{ij}) \right)$$

and $$B = \sum \left(p_{\cdot i} \times \log_2(p_{\cdot i}) + p_{i\cdot} \times \log_2(p_{i\cdot}) \right)$$

such that $p_{ij} > 0$, $p_{i\cdot} > 0$ and $p_{\cdot i} > 0$.

Although not apparent from this formulation, the index is composed of a part that measures the maximum amount of information available for transmission and another part which measures the amount of information lost in the transmission from one observer to the next. In fact, the index may be written as the difference of these two quantities. The ratings of either observer can be taken as a source from which the information is transferred without disturbing the resulting value. Thus, the index is symmetric with regard to which rater is taken as the reference. Note also that the definition includes tables having more than two categories.

Each of the uncorrected indices may be used to evaluate agreement. Whether agreement is to be measured symmetrically about the concordant cells and whether marginal proportions should be accounted for is dependent on the problem at hand and, perhaps most of all, on the predilection of the investigator. It is important to be aware, however, that the use of each

index may confer a different interpretation of the data (see Fleiss, 1981, for an example).

Table 3.3

Indices of Agreement for Two Raters Judging an Attribute on a Two-Point Scale(1)

Index	Name	Formula(2)	Reference(3)
p_o	Index of Crude Agreement or Overall Proportion of Agreement	$p_{11} + p_{22}$	1
p_s	Proportion of Specific Agreement	p_{22} / p_s	2
p_s'	Proportion of Specific Disagreement	p_{11} / p_s'	2
A_1	Index of Adjusted Agreement	$\left[\frac{p_{ii} (p_{i.} + p_{.i})}{(4\ p_{i.}\ p_{.i})} \right]$	1
A_2	Alternate Index of Agreement	$(p_s + p_s')\ /\ 2$	1

Table 3.3, continued

Index	Name	Formula(2)	Reference(3)
λ_r	Lambda-r	$2 \times p_s - 1$	3
ϕ	Phi or tau-b	$(\chi^2 / n)^{1/2}$	4,5
I	Average Amount of Information Transferred	(see text)	6

Table 3.3, continued

Index	Name	Formula(2)	Reference(3)
Indices Corrected for Chance Agreement			
π	Pi	$\dfrac{4(P_{11}P_{22} - P_{11}P_{21}) - (P_{12} - P_{21})^2}{(P_{1.} + P_{.1})(P_{2.} + P_{.2})}$	7
κ	Kappa	$\dfrac{2(P_{11} + P_{22} - P_{12}P_{21})}{P_{1.}P_{.2} + P_{.1}P_{2.}}$	8
r_{11}	r-11	$\dfrac{2(P_{11}P_{22} - P_{12}P_{21})}{P_{1.}P_{2.} + P_{.1}P_{.2}}$	9

Table 3.3, continued

(1) See Table 3.1 for definitions of the symbols used.

(2) $p_a = (p_{1.} + p_{.2})/2;$

$p_a' = (p_{1.} + p_{.1})/2.$

(3) References are: 1) Rogot and Goldberg, 1966

2) Dice, 1945

3) Goodman and Kruskal, 1954

4) Kendall and Stuart, 1961

5) Kendall, 1955

6) Liddell, 1963,

and Liddell and Lindars, 1969

7) Scott, 1955

8) Cohen, 1960

9) Maxwell and Pilliner, 1968

(3.1.3.2) Summary Indices that Account for Chance Agreement

None of the above-mentioned indices account for the level of agreement that would be expected by chance when the raters' observations are statistically independent of each other. Rogot and Goldberg, 1966, suggested that all summary measures should be reported with the expected values. Another method is to incorporate the expected values directly into the index. In fact, when expected agreement is taken into account under the constraint of statistical independence, most of the uncorrected indices can be transformed into one identical estimator of agreement. The procedure involved in this transformation is to subtract the expected agreement, I_E, from that observed, I_O, and then to standardize to the maximum possible excess agreement $(1-I_E)$, by division. Thus, the transformation commonly used is:

$$M(I) = \frac{I_O - I_E}{1 - I_E} . \qquad (1)$$

M(I) has a maximum value of one and a minimum value of $-I_E / (1 - I_E)$.

Two of the three indices listed in Table (3.3) that account for chance can be derived using this transformation. The Kappa index of Cohen, 1960, is obtained when any of the uncorrected indices, except tau-b, are substituted for I_O and when the corresponding estimate of I_E is calculated assuming statistical independence. Scott's π index (1955) can be calculated when

the constraints of marginal homogeneity and statistical independence are simultaneously applied to the calculation of I_0 (Landis and Koch, 1975). The index of Maxwell and Pilliner, 1968, r_{11}, is derived from other considerations (see next section).

(3.2) The Interpretation of Chance Corrected Indices as Intra-class Correlation Coefficients

Standard analysis of variance (ANOVA) models can be used to measure agreement between raters who are judging either continuous or categorical variables. Either fixed or random effects models can be considered in this context (see Bartko, 1966, and Landis and Koch, 1975), but for purposes of illustration, a random effects model containing subject and observer effects will be considered. Following the notation of Landis and Koch, 1975, the number of observers randomly selected from a larger population of potential observers is represented by d. Each observer judges the trait, denoted by the variable Y, for each of n randomly selected subjects on a continuous or categorical scale. The standard model is:

$$Y_{ij} = \mu + S_i + D_j + e_{ij} \qquad (2)$$

where μ = is the overall mean;

S_i = is the subject effect for the ith subject being rated;

D_j = is the observer effect for the jth rater; and

e_{ij} = is the residual error.

To make statistical inferences from this model (equation 2) it is assumed

that the S_i are normally, independently and identically distributed with a mean of zero and variance σ_s^2 (ie, NID $(0,\sigma_s^2)$), the D_j are NID $(0,\sigma_d^2)$, and the e_{ij} are NID $(0,\sigma_e^2)$. Homoscedasticity is assumed for each term since the variance for each effect is considered constant across all levels. Normality is only required when tests of significance are computed. The S_i, D_j, and e_{ij} are also assumed to be mutually independent.

Consistent estimates of these components of variance are obtained from:

$$\hat{\sigma}_s^2 = \frac{MS_d - MS_e}{n}$$

$$\hat{\sigma}_d^2 = \frac{MS_s - MS_e}{d}$$

$$\hat{\sigma}_e^2 = MS_e$$

where MS_s, MS_d, and MS_e are the mean squares for the subjects, the observers, and the residual error, respectively. These components are given the following interpretation (Landis and Koch, 1975):

$\hat{\sigma}_s^2$: is an indicator of "within subject effect" or "between observer effect", since it estimates the variability of the Y_{ij} over all

raters.

$\hat{\sigma}_d^2$: is an indicator of the "within observer effect" or "between subject effect" since it estimates the variability in the average judgements over all subjects. Therefore, it represents inter-observer disagreement.

$\hat{\sigma}_e^2$: represents the variability unaccounted by the main effects.

A coefficient, rho(ρ), based on these components of variance is used to measure inter-observer variation. It is defined (Fisher, 1958; Bartko, 1966) as the ratio of the within-subject variability to the total variability; viz:

$$\rho = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_d^2 + \sigma_e^2} .$$

Bartko, 1966, presents criteria whereby the rho coefficient can be interpreted as a within subject intra-class correlation coefficient. The selected analysis of variance model must be appropriate to the situation and the assumptions implicit to the model must not be violated (ie, independence of each factor and homoscedasticity).

If these criteria are met then consistent estimates of the components of variance can be obtained. The rho coefficient can then be interpreted as a *correlation coefficient between observers since*

$$\rho = \frac{\mathrm{cov}(x_{ij}, x_{ik})}{\mathrm{var}(x_{ij}) \times \mathrm{var}(x_{ik})} .$$

A consistent estimate of ρ is obtained by replacing the variances in the formula defining rho with their corresponding estimates. Thus, for the above model,

$$\hat{\rho} = \frac{n \times (MS_s - MS_e)}{d \times MS_d + n \times MS_s + MS_e \times (n(d-1) - d)} .$$

This coefficient can be used to indicate the amount of inter-observer error. Small values of rho indicate that the sum of the variances due to the observers and due to intrinsic error is greater than that attributable to the subjects; i.e.:

$$\sigma_d^2 + \sigma_e^2 > \sigma_s^2 .$$

Thus, the observers can not consistently discriminate between subjects. On the other hand, a value of rho about equal to one implies that either the intrinsic error and the between-observer variability is a) small absolutely

or is b) small with respect to subject variability; i.e.:

$$\sigma_d^2 + \sigma_e^2 = 0$$

or

$$\sigma_d^2 + \sigma_e^2 < \sigma_s^2 \ .$$

This last inequality indicates that there may be difficulty in interpreting the value of rho as a measure of inter-observer bias. If, for example, the variability between subjects is very large compared to the other sources of variability, then a high value of rho can be obtained even though the amount of observer variation may be substantial. To ensure a correct interpretation of the intraclass coefficient, a judicious choice of subjects must be made so that there is not "excessive" subject variation.

The three measures that account for chance agreement (Table 3.3) are related to intraclass correlation coefficients derived from using different ANOVA models. Fleiss, 1975, demonstrated that Cohen's Kappa index is asymptotically equivalent to the intraclass correlation coefficient derived from the random-effects ANOVA model having subject and observer effects. Scott's pi index, which simultaneously assumes homogeneity in the marginal proportions and statistical independence, is asymptotically equivalent to that obtained from a model where the observer effects, S_i, are subsumed in the residual error term. The index of Maxwell and Pilliner, 1968, is developed from the fixed-effects ANOVA model using only observer and subject effects.

(3.3) The Kappa Index

As was previously discussed, the three indices in Table 3.3 that are corrected for chance agreement can be interpreted, asymptotically, as intraclass coefficients. The advantage of Kappa is that its corresponding intraclass correlation coefficient, unlike the index of Maxwell and Pilliner, corresponds to a random-effects model which includes observer and subject effects. Unlike Scott's Pi index, it does not assume marginal homogeneity.

The interpretation of these indices as intraclass correlation coefficients is useful in determining which measure is appropriate in a given situation. In particular, the use of Kappa may be questioned when the number of observers are thought to be fixed. For example, in this thesis, comparisons in attributing exposure are made between Cancer Study chemists. Use of the fixed-effects model may be preferred if these raters comprised the universe of possible raters or if a comparison among a particular subset of raters within the Study is required. On the other hand, if these raters are considered to be a sample of all possible chemists, and the results are to be generalized to the universe of potential raters, then the random-effects model would be appropriate. It is assumed in this thesis that the latter is the case.

(3.3.1) Qualitative Interpretation of Kappa

In an effort to provide a qualitative framework for interpreting Kappa, Landis and Koch, 1977a, have arbitrarily classified ranges of values of

Kappa obtained from large samples into qualitative degrees of agreement (Table 3.4). This classification scheme has not been validated in any research context.

There are two noteworthy problems when using Kappa as a measure of agreement. The first is the potential difficulty that may result if the non-linearity of its definition (equation 1) is also manifest in its interpretation. For example, it is not clear that an increment in agreement from $\kappa = 0.55$ to $\kappa = 0.60$ (9% increase) represents a greater improvement in "true" agreement (as measured by differences in sensitivity and specificity between the raters) than that obtained from an increase from $\kappa = 0.2$ to $\kappa = 0.4$ (100%). The second problem is that comparisons of values of Kappa from different populations may not be meaningful if the prevalences in the two populations differ (Thompson, 1982, and Walter, 1983). In general, therefore, the rather simplistic classification scheme of Landis and Koch may not be appropriate. In the absence of more satisfactory interpretive criteria, however, their scheme will be used in this thesis to qualitatively indicate the extent of agreement.

Table 3.4

Qualitative Interpretation of the Values of Kappa (1)

Kappa	Qualitative Interpretation of Agreement
0.81 - 1.00	Almost perfect
0.61 - 0.80	Excellent
0.41 - 0.60	Good
0.21 - 0.40	Fair
0.00 - 0.20	Slight
< 0	Poor

(1) Adapted from Landis and Koch, 1977a.

(3.3.2) Estimating the Variance of Kappa

Asymptotic estimates of the variance of Kappa have been derived (see Fleiss, 1981). Small sample variance estimates have not been reported although, theoretically, they can be estimated using resampling techniques such as the jackknife and the bootstrap (Miller, 1974; Efron, 1981; Efron, 1982; and Fleiss and Davies, 1982). It is usually assumed that for large samples Kappa is normally distributed but, to the author's knowledge, this has not been explicitly verified. For small sample sizes, Kappa can not be distributed normally since it is defined in a non-linear way, is finitely bounded at both ends of its range and is positively skewed. Fleiss and Davies, 1982, reported that the jackknife procedure functions quite well for samples having more than 50 subjects. Simulation of small sample variance estimates and comparison with resampling estimates are currently in progress for $n<50$ and will be reported elsewhere.

(3.3.3) Interpreting the Value of Kappa when there are a Small Number of Observations

The value of Kappa based on small sample sizes is quite unstable. This is illustrated with the following example. Suppose that the estimates of prevalence for two raters judging a trait were both low. Suppose also that very few subjects were available for observation. Consider the following

two contingency tables:

	R-2			R-2		
	-	+		-	+	
R-1 -	18	1	19	17	1	18
R-1 +	1	0	1	1	1	2
	19	1	20	18	2	20

$\kappa = -0.05$ $\kappa = 0.44$

There is excellent agreement in both tables using the index of crude agreement (p_0 = 90%). However, the two values of Kappa tell a different story: poor agreement is indicated in the first table ($\kappa = -0.05$) but good agreement is indicated in the second ($\kappa = 0.44$). The large differences between these two values is due entirely to the small changes in the marginal proportions that resulted when the p_{22} cell was changed from 0 to 1. In effect, the estimate of Kappa is related to the number of observations that are used to test agreement. This number, which can be regarded as the "effective sample size", is a function of the expected agreement obtained from the marginal proportions and the total sample size.

The highly discordant values of Kappa observed in the example presented above can be understood in terms of effective sample size. In the first table, the 18 observations in the absent category is about equal to that which would be expected by chance. Thus, only two observations are being

subjected to the "test" of agreement. The value of Kappa is about zero since zero out of the remaining two observations "pass" the test. In the second table, there are three observations to scrutinize for agreement; one out of the three passes, thus yielding a value of Kappa of about one third.

The manipulation of only one data point, under the constraint of a constant value for the index of crude agreement, had a profound effect on the value of Kappa. From this, it would appear that the variance of Kappa in either table should be very large. The variances, as derived from asymptotic theory, were computed for the two examples presented above. The estimates of variance in the first case is about zero but is much higher (0.46) in the second. Apparently, asymptotic theory is not appropriate for such small sample sizes. Without accurate estimates of variance, and the resultant confidence intervals, it is difficult to correctly interpret the statistic.

One further problem can occur when small samples are used to evaluate agreement. This difficulty is illustrated by considering two examples in which both observers are in complete agreement; viz,

		R-2				R-2		
		-	+			-	+	
R-1	-	0	0	0	-	1	0	1
	+	0	20	20	+	0	19	19
		0	20	20		1	19	20

$p_o = 1.0$ $\qquad$ $p_o = 1.0$

$p_e = 1.0$ $\qquad$ $p_e = 0.905$

$\kappa = 0/0 = \text{undefined}$ $\qquad$ $\kappa = 1.0$

The first panel of this example, in which the value of Kappa is undefined, illustrates that this index is not entirely suitable to situations for which perfect agreement is encountered. In this thesis, the problem of undefined values will be circumvented by setting κ to unity in all such circumstances.

(3.3.4) The Maximum Value of Kappa

The maximum value that Kappa can attain in any situation depends on the distribution of the marginal proportions. Thus, when two raters produce different marginal distributions, then, as Cohen, 1960, pointed out, the upper limit is less than unity. Cohen labeled this quantity as κ_{max} and showed that it can be calculated from the formulae:

$$\kappa_{max} = \frac{P_{0M} - P_E}{1 - P_E}$$

where

$$P_{0M} = \sum \min(p_{i.}, p_{.i}).$$

$$P_E = \sum (p_{i.} * p_{.i}).$$

The calculation of κ_{max} proceeds by replacing the concordant cells (p_{ii}) with the smallest marginal value corresponding to that value of i (ie, $p_{i.}$ or $p_{.i}$). The discordant cells are obtained by subtraction. Table 3.5 presents an example of the calculation of κ_{max}.

Table 3.5

Sample Calculation of the Maximum Value of Kappa

		R-2				R-2		
		-	+			-	+	
R-1	-	12	3	15	-	13	2	15
	+	1	4	5	+	0	5	5
		13	7	20		13	7	20

$\kappa = 0.53$ $\kappa_{max} = 0.77$

The ratio of κ to κ_{max} (hereinafter denoted as κ') represents the ratio of agreement to the maximum amount of agreement permitted by the marginal proportions. This quantity indicates the amount of agreement *within* the table after taking into account different marginal distributions produced by the raters. κ' is not a standardized index (equation 1) and, to the author's knowledge, it can not be interpreted as an intraclass correlation coefficient. κ_{max} has not been generalized to the case in which there are more than two observers.

(3.3.5) Extension of Kappa to Multiple Categories and Multiple Raters

Kappa has been extended to the case in which two observers rate a variable having m mutually exclusive categories. Let i represent the ith row and j the jth column of a mxm contingency table created by crosstabulating the two observers' ratings (Table 3.6). The general formula for Kappa is:

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

where

$$P_o = \sum P_{ii}$$

and

$$P_e = \sum (P_{i.} \times P_{.i}) .$$

The seriousness of each disagreement in the table can be taken into account through the use of weights (Cohen, 1968). This is the most general formulation for κ and is referred to as "weighted Kappa". Large sample variances have been obtained for the weighted and unweighted versions of Kappa (Fleiss, 1981).

An ANOVA model for multiple raters judging a trait on a dichotomous variable has been used to generalize Kappa. The model contains subject and observer effects and uses the definition of the intraclass correlation coefficient to derive an estimator for Kappa. In the general situation where d raters judge a variable containing m categories, the data can be

displayed as in Table 3.7. The variable x_{ij} represents the number of judgements by the raters into category "j" for subject "i". Kappa can be calculated using the formula (Fleiss, 1981):

$$\kappa = 1 - \frac{\left(n m^2 - \Sigma x_{ij}^2 \right)}{\left(nm(m-1) \Sigma \bar{p}_j \bar{q}_j \right)}$$

where

x_{ij} = is the number of judgements into category "j" for subject "i";

n = the number of subjects;

m = the number of categories;

$\bar{p}_j = \Sigma x_{ij} / n$;

and

$\bar{q}_j = 1 - \bar{p}_j$.

A summary statistic can also be calculated for a set of independent estimates of Kappa (Fleiss, 1981; Chapters 10 and 13). Each value of Kappa is weighted by the inverse of its estimated variance. A summation is then made over all values and then is normalized by dividing by the sum of the inverse of the variance. A test for homogeneity of the individual values of Kappa can also be calculated.

Further extensions to other more complicated situations have been considered by Landis and Koch, 1977a,b,c, Kraemer, 1980, Thomas et al., 1981, and Hubert and Golledge, 1983.

Table 3.6

Description of the Data for Two Observers Rating a Variable having m Mutually Exclusive Categories

		Rater - 1						
		1	2	3	4	.....	m	
Rater-2	1	P_{11}	P_{12}	P_{13}	P_{14}	.....	P_{1m}	$P_{1.}$
	2	P_{21}	P_{22}	P_{23}	P_{24}	.....	P_{2m}	$P_{2.}$
	3	.	.	.	.	.....	.	.
	.	.	.	.	.	.....	.	.
	m	P_{m1}	P_{m2}	P_{m3}	P_{m4}	.....	P_{mm}	$P_{m.}$
		$P_{.1}$	$P_{.2}$	$P_{.3}$	$P_{.4}$	.....	$P_{.m}$	1

Table 3.7

Description of the Data for More than Two Raters Assigning Values to a Variable having m Categories (1)

		Number of Ratings into Category:					
		1	2	3	4	.....	m
	1	x_{11}	x_{12}	x_{13}	x_{14}	.....	x_{1m}
	2	x_{21}	x_{22}	x_{23}	x_{24}	.....	x_{2m}
Subject:	3	.	.	.	.	.....	.
	.	.	.	.	.	.....	.
	n	x_{n1}	x_{n2}	x_{n3}	x_{n4}	.....	x_{nm}
		$x_{.1}$	$x_{.2}$	$x_{.3}$	$x_{.4}$	.....	$x_{.m}$

(1) x_{ij} is the number of judgements by the d raters into category "j" for subject "i".

(3.4) Other Measures of Agreement for Multiple Observers

(3.4.1) Tests of Homogeneity for the Marginal Proportions for More than Two Observers

It has been proposed that Cochran's Q index can be used to test the homogeneity of the marginal proportions among two or more raters (see Siegal, 1956 and Fleiss, 1965). The statistic is asymptotically distributed as a chi-square with d-1 degrees of freedom. To calculate Q, the data should be displayed as in Table 3.8. The quantities are defined as follows: Y_{ij} are the ratings (0,1) given to the trait for subject i by observer j; t_i represents the number of ratings where a value of "1" was given for subject i. (Thus, t_i is just the sum of the Y_{ij} over all observers, j=1,d, for each subject, i.) Q is calculated from the formula (Fleiss, 1965):

$$Q = \frac{(D-1)\left(D \left[Y_{i.}^2 - Y_{..}^2 \right] \right)}{DY_{..} - \left[t_j^2 \right]} .$$

Other tests have been suggested by Bennett, 1967, 1968, 1972.

(3.4.2) The Mean Majority Agreement Index (MMAI)

Various other indices of agreement have been proposed for multiple raters. Armitage et al., 1966, proposed three related indices of agreement for dichotomous data; i.e.: the Mean Majority Agreement Index (MMAI), the Mean Pair Disagreement Index (MPDI), and the Standard Deviation Agreement Index (SDAI). All three indices are uncorrected for expected agreement.

Only the MMAI will be described here. Majority agreement, for any one subject, is defined if more than half the observers judge the trait to be absent (or present). Consider, once again, the hypothetical data in Table 3.8. Let $p_i = t_i / d$. Thus, majority agreement is obtained if $|p_i| > 0.5$. The Majority Agreement Index, MAI, is defined as $MAI_i = |2\ p_i - 1\ |$. For example, if 3 of 5 observers score the same for a subject then $p_i = 3/5$ and $MAI_i = 1/5$.

The MMAI is just the arithmetic mean of the MAI over all subjects; viz:

$$MMAI = \sum MAI_i / n.$$

MMAI can attain a maximum value of one but its minimum is dependent on whether the number of observers is an even or odd number. In the case of two raters the MMAI is identical to the index of crude agreement.

Table 3.8

Description of the Data for Two or More Observers Rating a Trait on a Two-point Scale for n Subjects (1)

		Observer						
		1	2	3	4	....	D	
Subjects	1	Y_{11}	Y_{12}	Y_{13}	Y_{14}	....	Y_{1D}	t_1
	2	Y_{21}	Y_{22}	Y_{23}	Y_{24}	....	Y_{2D}	t_2
	3	.	.	.	.	....	.	.
	.	.	.	.	.	....	.	.
	n	Y_{n1}	Y_{n2}	Y_{n3}	Y_{n4}	....	Y_{nD}	t_n
		$Y_{1.}$	$Y_{2.}$	$Y_{3.}$	$Y_{4.}$	....	$Y_{n.}$	$Y_{..}$

(1) Refer to the text for the definition of these quantities.

(3.5) Summary

Various methods have been revieved that summarize agreement between observers judging qualitative data. The analysis of inter-rater agreement was distinguished along two broad lines. The first dealt with testing the equality of the marginal proportions. The second considered methods that would indicate the extent of agreement. Data display techniques (eg, "the pattern of concordance") and summary measures of agreement were discussed. Two types of summary measures of agreement were presented; those that did and did not correct for chance agreement.

For the purpose of this thesis, it was felt that no single method would suffice to portray inter-rater agreement; nor was it practical to use all possible methods. As a consequence, the following four complementary methods were used: 1) statistical tests for the equality of the marginal proportions, 2) patterns of concordance, 3) the index of crude agreement and MMAI and 4) values of Kappa were estimated between raters.

(4.0) Objective and Overview of the Trials

The Cancer Study was designed to generate hypotheses concerning associations of various sites of cancer with occupational exposures. Exposure to substances were ascertained by a team of chemists who reviewed detailed job descriptions usually obtained by interviewing subjects.

(4.1) Objective

The objective of the present study was to evaluate the accuracy of the chemists' assessments of exposure.

(4.2) Overview of the Trials

Seven trials (Table 4.1) were used to evaluate the extent of agreement in attributing exposure between different raters who were given the same job descriptions to code. The seven trials fall into the following four categories, depending on the nature of the raters being compared:

(1) the raters were individual chemists from the Cancer Study independently coding the same files;

(2) some raters were chemists from the Cancer Study, coding exposure independently of each other, and others were knowledgeable persons from industry;

(3) some raters were chemists coding exposure as a group, using a

consensus method, and others were knowledgeable persons from industry; and

(4) the same group of Cancer Study chemists, using a consensus method to attribute exposure, was given a set of files at two different times.

The trials were named, where applicable, according to the type of job description sampled. Some trials compared the assessment of exposure for job descriptions that were classified by industry (the Paint Manufacturing, Metal Industry, Chemical Manufacturing and Rubber trials) or occupation (Welding Trade Trial). Other trials (the General Comparison and Code/Recode trials) used two samples of job descriptions taken from the Cancer Study. The trials took place over the last four years (1981-1984).

Table 4.1

Catalogue of Trials

Trial	Purpose	Year of Trial	Number of Jobs (1)	Number of Raters: External	Number of Raters: Chemists
Paint Manufacturing Industry	To compare the coding of 2 Cancer Study Chemists with an external rater.	1981	5	1	2
Welding Trade	To compare the coding of 2 Cancer Study chemists to a group of welders and engineers from industry and academia.	1982	18	3(2)	2
Metal Industry	To compare consensus coding of Cancer Study chemists with industrial hygienists employed at a community health department.	1984	7*	2	4

Table 4.1, continued

Trial	Purpose	Year of Trial	Number of Jobs (1)	Number of Raters: External	Number of Raters: Chemists
Chemical Manufacturing	To compare coding of an industrial hygienist to a consensus coding by Cancer Study chemists.	1984	5+	1	4
Rubber Industry	To compare coding among 3 Cancer Study Chemists.	1981	15	0	3
General Comparison	To compare exposure assessments of 4 Cancer Study chemists who reviewed job descriptions from the Cancer Study.	1981	20	0	4

Table 4.1, continued

Trial	Main Purpose	Year of Trial	Number of Jobs (1)	Number of Raters: External	Number of Raters: Chemists
Code/ Recode	To compare consensus coding for job descriptions that were rated by the Cancer Study Chemist Team at two different times.	1984	23	0	2

(1) All job descriptions were taken from interviews of subjects in the Cancer Study, except for those indicated with a "*" in which interviews of non-diseased persons were obtained at industrial locations in Montreal and with a "+" in which job descriptions were fabricated from company records.
(2) The three external raters attributed exposure as a team.

(5.0) Materials

(5.1) A Description of the Study: "Monitoring the Occupational Environment for Carcinogens"

The work for this thesis took place within the context of this study (referred to as the *Cancer Study*). Only the relevant features of the Cancer Study are described below.

The Cancer Study was designed to discover associations between different sites of cancer and substances found in occupational environments (Siemiatycki, 1979, Siemiatycki et al., 1981a,b, 1982, 1983a,b; and Gérin et al., 1983, 1984). A case-control approach was taken. Patients diagnosed with cancer were eligible for admission in the Cancer Study if certain criteria were satisfied. A team of chemists/engineers, using detailed job descriptions usually obtained by interview, evaluated life-time exposure to about 300 substances. For the purpose of obtaining estimates of excess risk, cases for each site could be compared with community controls and/or with subjects having any of the other types of cancer (Thomas et al., 1984a, 1984b).

(5.1.1) **Selection of Cases**

Patients were eligible for registration in the Cancer Study as cases if they were: 1) male; 2) between the ages of 35 and 74 years in the calendar years 1979 to 1984; 3) newly diagnosed, at a Montreal hospital, with any of 13 primary sites of cancer (stomach, esophagus, colon, pancreas, liver, testes, penis, kidney, lung, bladder, prostate, lymphatic tissues, and skin or eye melanoma) which were positively confirmed by histological examination; and 4) domicile in the greater Montreal area at the time of diagnosis. Names of eligible cases were obtained from records held in pathology departments in participating hospitals.

(5.1.2) **Interview Procedures for Subjects**

A two part questionnaire/interview (*QI*) was used to elicit relevant lifetime information for those subjects who could be interviewed. A self-administered questionnaire was mailed to the subject or to his next-of-kin if, for any reason, an interview could not be performed. In the first part of the QI, an outline of the subject's lifetime work history was obtained by the use of a self-administered questionnaire (*SAQI*). Personal data and relevant information for each occupation was obtained in the second part of the QI. The accuracy of reported occupations has been verified by Baumgarten et al., 1983.

Detailed descriptions of each subject's occupations were obtained. (These are referred to as *job descriptions*.) Information was requested regarding the specific job functions, materials and machines used, and the general

working environment. Specific questions have been developed for various types of jobs (egs, welders, painters, farmers; see Appendix 2 for some questionnaires in current use). The questions were open-ended so that the interviewers were able to probe as deeply as possible. The objective was to form as clear a mental picture as possible of the environment of each of the subject's occupations. The interviewers were encouraged, when necessary, to consult with the chemists prior to conducting the interview so that possible ambiguities in interpreting the job descriptions could be ironed-out in advance.

(5.1.3) Attributing Exposure from Reported Job Descriptions

For each reported occupation, a team of chemists attributed exposure from a checklist containing approximately 300 chemical and physical agents (Gérin et al., 1983, 1984). Team members were trained chemists, engineers or industrial hygienists familiar with industrial processes and with commercial and industrial environments in the greater Montreal area. (As previously noted, these raters will be referred to as *chemists*.) Information concerning exposure was taken from the literature (see Price, 1982), from discussions with industry experts, personal experience, and from occupational health surveys (eg, DSC, 1983).

For each job, one of the chemists determined the Standard Industrial (SIC, 1970) and occupation codes (OCC, 1971). Next, exposure to each of the substances listed on the chemical coding form was determined. The type and extent of exposure was indicated using four semi-quantitative scales: 1) average dose or *concentration*, 2) *frequency* of exposure during a normal

workweek, 3) type of exposure (*contact*), and 4) the degree of certainty that the exposure actually occurred (*level of confidence*).

At least one other team member reviewed each file. The chemist who first assigned exposure to a subject's job description is referred to as the *primary coder*. The chemist, or chemists, who subsequently reviewed the primary coder's assessment is referred to as the *secondary coder*. The complete process will be referred to as the *consensus coding*.

A listing of all exposure assessments was periodically produced. The purpose of the list was to allow the chemist to reference past exposure assessments for similar types of jobs. The ability to consult previous assessments helped to ensure that the coding was uniform over time. This type of list is referred to as an *exposure audit trail*. Other devices, such as periodically reviewing all job descriptions, were also used to standardize the coding process.

(5.1.4) Variables Used to Distinguish the Type and Level of Exposure

As indicated above, four variables, referred to collectively as the *exposure variables*, were used to describe the type and level of exposure to each substance on the exposure coding form. These variables are described below in more detail (Table 5.1). (See Appendix 4 for a further explanation of how these variables are utilized.)

Concentration refers to the average dose to which the subject was exposed. The scale does not make reference to any absolute scale of measurement (eg, molecules per unit volume of air) or to accepted maximum permissible doses. Criteria have been developed for coding specific substances and classes of substances on the chemical checklist. These criteria were used to help ensure that each substance would be consistently coded over all job descriptions.

Level of confidence indicated the certainty that the exposure actually had occurred. The rater's confidence of the assessment depended on her/his ability to correctly interpret the job description. This, in turn, depended on the quality of the interview and of the chemist's knowledge of the job situation. A value of "possible exposure", for example, implied that the chemist was uncertain that the exposure had occurred, even if high values of frequency and concentration were assigned. Such a situation could occur, for example, if the chemist was undecided whether either of two substances were used in a process, but was fairly certain that the worker was highly exposed to one of them. Thus, both substances would be coded with a "possible" level of confidence.

Table 5.1

Description of the Four Variables Used to Indicate the Type and Level of Exposure for Each Substance on the Exposure Checklist

Exposure Variable	Description	Scale Used
Contact	Type of exposure	0 - no exposure 1 - respiratory 2 - cutaneous 3 - 1 and 2
Frequency	Average fraction of time where contact occurred	0 - no exposure 1 - <5% of the time 2 - 5 to 30% 3 - >30%

Table 5.1, continued

Exposure Variable	Description	Scale Used
Concentration	Average dose (1)	0 - no exposure 1 - "low" 2 - "medium" 3 - "high"
Level of Confidence	Certainty of exposure assessment (1)	0 - no exposure 1 - possible exposure or "low" 2 - probable exposure or "medium" 3 - almost certain exposure or "high" (2)

(1) Refer to the text and Appendix 4 for a discussion.

(2) The term "any" is used to refer to exposure at levels 1, 2, or 3.

(5.1.5) Potential Sources of Error in Attributing Exposure from Job Descriptions

It is difficult to accurately attribute exposure from reviewing job descriptions. Even if a chemist has considerable knowledge and coding experience, there are at least three possible sources of error that can occur. The first is the difficulty of accurately interpreting the job description. For example, errors in correctly identifying materials used at a plant may result if the job description is ambiguous. The second is incomplete knowledge of industrial processes and materials. The third is the lack of consistency of applying the coding criteria to the exposure variables.

(5.2) Selection and Description of Raters

Raters from the Cancer Study (Table 5.2) were chosen from chemists who were employed at the time of each trial. Chemists 1 and 4 were the most experienced raters having been involved in the Cancer Study since its inception. In particular, Chemist 1 was one of the principal investigators and Chemist 4 was responsible for initially reviewing almost all new files (ie, primary coding). Chemists 2, 3, and 5 were employed mostly as secondary coders. Chemist 6 was involved in the design of the Cancer Study but did not have any day-to-day experience coding exposure.

Table 5.2

Description of the Professional Experience of the Chemists who Participated in the Agreement Trials

Chemist	Professional Title/ Principle Work on Chemist Coding Team(1)	Expertise: Coding Experience	Expertise: Prior to Cancer Study
Chemist 1	Chemist(PhD)/principal investigator	1979-84	
Chemist 2	Chemist/ 2nd coder	1980-82	
Chemist 3	Chemist/ 2nd coder	1980-83	Paint mfg
Chemist 4	Engineer/ 1st coder	1979-84	Pulp and paper, minerals
Chemist 5	Chemist/ 2nd coder	1982-84	Paint mfg
Chemist 6	Chemist(PhD)/ special projects (2)	-----	

(1) This refers to whether the chemist worked as a primary coder (1st coder), as a secondary coder (2nd coder), or in special coding projects more than 50% of the time.

(2) This chemist had no extensive experience in coding exposure but was involved in designing the Cancer Study.

The external judges were selected because of their presumed detailed knowledge of processes and substances that were used in various work situations. The external raters were either experienced industrial chemists, industrial hygienists, or were actively involved in monitoring industrial environments. Table 5.3 presents a brief portrait of each of the external raters' professional experience.

Table 5.4 lists the raters who participated in each of the trials.

Table 5.3

Description of the Professional Experience of the External Judges who Participated in the Agreement Trials

External Judge(s)		Profession	Experience
Paint Manufacturing Trial			
Judge 1		Industrial hygienist at a large Canadian paint manufacturing company.	about 10 years
Welding Trade Trial			
Judges 2:	a)	Engineer involved in welding research and education.	about 10 years
	b)	Welder from trade association.	about 10 years
	c)	Active in training of welders in industry.	>30 years

Table 5.3, continued

External Judges	Profession	Experience
Metal Industry Trial		
Judges 3	Two industrial hygienists employed at the DSC Sacre-Coeur.	about 5 years each
Chemical Manufacturing Trial		
Judge 4	Industrial hygienist at a large Canadian chemical manufacturing company.	about 7 years

Table 5.4

Raters who Participated in each of the Trials

Trial	Chemist(s)	External Judge(s)
Paint Manufacturing	Chemists 1 and 2	Judge 1
Welding Trade	Chemists 1 and 3	Judges 2
Chemical Mfg	Chemists 4, 5 (1)	Judge 3
Metal Industry	Chemists 4, 5 (1)	Judges 4
Rubber Industry	Chemists 1,2,3 (2)	--------
General Comparison	Chemists 1,2,4,6	--------
Code/Recode	Chemists 4,5 (3)	--------

(1) This was a consensus coding performed by these chemists.

(2) A Final Coding, produced by one of the chemists after reviewing the three assessments of exposure, was also available for analysis.

(3) This was a trial in which job descriptions were coded at two different times by these chemists using the consensus approach.

(6.0) Methods

In this Chapter, the methods common to all trials will be described. Specific details for each trial will be found in the Materials and Methods sections in Chapter 7.

(6.1) Coding Procedures

In each trial, all raters independently assessed exposure after reviewing an identical set of job descriptions. In five of the seven trials, job descriptions were selected from the files of the Cancer Study. Only those job descriptions which had been obtained by interviewing subjects were eligible. In the remaining two trials, job descriptions were fabricated from company records (the Chemical Manufacturing Trial) and by interviewing workers at two industrial plants (the Metal Industry Trial). The specific criteria for selecting job descriptions will be discussed in Chapter 7.

The number of substances used in each trial depended on which coding form was being used at that time in the Cancer Study. Three different exposure coding forms, containing 172, 270, and 300 substances, were used in the trials.

In order that all raters interpreted the coding criteria in a similar way, some external judges were given training in the use of the coding system. Other external raters were assisted in their exposure assessments by a senior chemist (Chemist 1), who did not otherwise participate in the specific trial, and the author. In no instance did the chemist or the

author suggest to the raters which substances should be attributed.

All raters had access to the same literature, although only the chemists were able to reference all data available in Cancer Study records (eg, the exposure audit trail).

(6.2) Statistical Methods

Various indices, measures and data display techniques were used to determine the degree of agreement among the raters. Two methods of analysis were performed: a) agreement was measured over all substances and job descriptions taken together; and b) agreement was assessed for each substance separately.

(6.2.1) Structure of the Data Sets

The data sets generated from the agreement trials were similar in structure. In general, there were d raters who attributed exposure to s substances for a sample of n job descriptions. The four exposure variables were scored by each rater on different four-point scales (each coded: 0, 1, 2, and 3). Each data record, therefore, contained an identification code for each substance, a variable that identified the job description, and values for each of the four exposure variables attributed by each of the d raters (Table 6.1).

Table 6.1

General Structure of the Data Sets Used in the Analysis of each Trial

Substance Code (1)	Job No	Rater Assessment for the Exposure Variables(2): Contact R1 R2...Rd	Concentration R1 R2...Rd	Frequency R1 R2...Rd	Level of Confidence R1 R2...Rd
xxxxxx	1	c c... c	c c... c	c c... c	c c... c
xxxxxx	2	c c... c	c c... c	c c... c	c c... c
xxxxxx	3	c c... c	c c... c	c c... c	c c... c
xxxxxx	.	. .... .	. .... .	. .... .	. .... .
xxxxxx	n	c c... c	c c... c	c c... c	c c... c
.....	.	. .... .	. .... .	. .... .	. .... .
.....	.	. .... .	. .... .	. .... .	. .... .
yyyyyy	1	c c... c	c c... c	c c... c	c c... c
yyyyyy	2	c c... c	c c... c	c c... c	c c... c
yyyyyy	3	c c... c	c c... c	c c... c	c c... c
yyyyyy	.	. .... .	. .... .	. .... .	. .... .
yyyyyy	n	c c... c	c c... c	c c... c	c c... c

Table 6.1, continued

(1) Only two of the s substances are shown here.

(2) The letter "c" is used to indicate the values 0, 1, 2, 3 that were given to each of the exposure variables by any of the raters, R1, R2,..., Rd.

(6.2.2) Measures of Agreement Adopted for this Thesis

Certain summary measures and data display techniques were used to assess agreement. The observed and expected numbers of exposure for each pattern of concordance was calculated. In addition, two summary indices were selected as measures of agreement between pairs of raters. The index of crude agreement (or overall proportion of agreement), p_o, which does not correct for chance agreement, was chosen because 1) it summarizes, in a symmetric way, the extent of agreement in a contingency table by summing the concordant cells, 2) it is easy to interpret, 3) it can be used for contingency tables containing more than two categories, and 4) it is compatible with other uncorrected indices, such as the Mean Majority Agreement Index (MMAI). The Kappa index, which takes chance agreement into account, was selected as the second summary index. The strengths and weaknesses of this widely used measure of agreement have been discussed in Chapter 3.

Agreement among all raters, considered as a group, was assessed using the MMAI and generalized Kappa indices. These were chosen because of their similarity with the indices used in the pairwise analyses.

(Ψ:2.3) Assessment of Inter-Rater Agreement using the Collapsed Table Method

The objective of this first analysis was to estimate an "average" level of agreement over all substances on the exposure coding sheet. One method, which had been employed elsewhere (Siemiatycki et al., 1982), combined the attribution of exposure over all substances and job descriptions. This method of assessing agreement is referred to as the *collapsed table method*.

For two raters, the method consists of crosstabulating the exposure assessments for each substance separately. The contingency tables are then combined into one by adding the values of the corresponding cells. The resulting contingency table, consisting of n x s observations, is referred to as the *collapsed table*. The judgements for more than two raters can be combined in a similar way.

For any job description, a large number of substances were coded as non-exposed by all raters. If s' of the s substances had a non-zero prevalence of exposure, as estimated by any of the raters, then there would be n x (s-s') observations coded as non-exposed. This subset can not be excluded from the analysis simply because all judges attributed no exposure to these substances. There is valuable information in the fact that a certain number of substances were judged absent by all of the raters. Therefore, the number of observations used in the analysis is equal to the number of job descriptions times the number of substances on the exposure checklist (n x s).

(6.2.3.1) Procedures Used for Summarizing Agreement

The first step was to obtain each rater's estimate of the frequency, or prevalence, of exposure at each level of confidence. The denominator for these calculations is equal to the total number of exposures, n x s. The equality of these estimates was tested among all raters using Cochran's Q statistic. McNemar's Matched Pairs Test was used to test equality between each pair of raters.

The second step involved grouping the judgements of each rater into two categories: exposed, at one or more levels of confidence, and exposed, or not, at all other levels. Exposure at the high, medium or high, or at any level of confidence was considered. The observed and expected numbers for each pattern of concordance was calculated for the various categorizations of exposure. The observed numbers were obtained from the cells of the contingency tables and the expected numbers were calculated from the marginal proportions using the methods described in Chapter 3. In addition to these tables the Mean Majority Agreement Index (MMAI) and the generalized Kappa (κ) index were also calculated.

The third step was to assess the extent of agreement between all possible pairs of raters based on the above scheme of categorizing exposure. Two-by-two contingency tables were created for each pair of raters and certain agreement statistics (p_o, κ and κ') were calculated.

For each exposure variable (contact, frequency, concentration, and level of confidence), agreement was calculated between each pair of raters. The

index of crude agreement, κ, and κ' were calculated from contingency tables that utilized the original four-point scales of measurement. Unweighted estimates of Kappa were used in these calculations since no meaningful weighting scheme could be devised. Agreement for a composite index of exposure, previously used for analysing exposure/site associations (Siemiatycki et al., 1983a), was also assessed. This index was obtained by combining frequency, concentration and level of confidence multiplicatively. (This index will be referred to as the *synthetic index*.) Level of confidence was used to weight the index in terms of its reliability of assessment. The index had only 11 distinct categories since each of the three variables was coded 0, 1, 2, or 3. Certain values of the synthetic index can be arrived at in more than one way. For example, the value 12 can be obtained from 2x2x3 (frequency x concentration x level of confidence), from 2x3x2, or from 3x2x2. Therefore, the exposure to a substance that was coded with a high degree of confidence and medium levels of concentration and frequency was assumed to be identical to a substance that was coded with a medium level of confidence, a high frequency of occurrence, and a medium concentration level. For the purpose of analysing inter-rater agreement, a new variable was created by arbitrarily categorizing the synthetic index into three intervals: 0-0, 1-6, and 8-27 (ie, no, "medium", and "high" exposure levels).

(6.2.3.2) **Statistical Problems Associated with the Analysis**

The purpose of this analysis was to obtain a summary measure of agreement. This was achieved, as described above, by combining each of the s contingency tables into one "collapsed" table by summing the values for each cell. It might appear that the original unit of observation (ie, the job description) had been replaced by the "exposure". By pooling the data in this way, however, a method has been developed which approximately measured the extent of agreement averaged over all substances under consideration.

In Appendix 5, it is shown that the index of crude agreement, averaged over a set of contingency tables, is identical to that calculated from the collapsed table. A heuristic proof is given that shows that the average *expected* value for this index and the average value of Kappa are approximately equal to that obtained from the collapsed table. From these considerations, it can be tentatively concluded that this method provided an approximate means of averaging agreement over all substances under consideration. It must be emphasized, however, that this conclusion has not been rigorously verified.

In addition, the confidence intervals of the estimates of Kappa derived from the collapsed table may not be interpretable in the usual statistical sense. Nevertheless, the confidence intervals are presented as a means of indicating the variability of the estimates.

A second problem with this analysis was that the assessment of exposure of

one substance by a rater may have been correlated with his/her assessment of other substances. For example, it might be expected that most raters would not be able to distinguish whether amphibole or crysotile asbestos mineral forms were present in the environment of a subject working in an insulation manufacturing plant. Thus, the raters would attribute equal levels of exposure to both substances. In the analysis of agreement between two such observers, this inability to discriminate between the two forms of asbestos would probably inflate most measures of agreement.

The correlation of exposure between some substances may affect the interpretation of Kappa. It will be recalled from Chapter 3 that Kappa may be interpretated as an intra-class correlation coefficient derived from a random-effects ANOVA model. One of the fundamental assumptions of this model is that the variables used to measure exposure should be mutually independent. Thus, the correlation of the assessments of exposure may affect this interpretation of Kappa.

One solution to this problem is to group all correlated substances into categories and then assess the agreement within each category. This is unattainable in practice because there is no reliable, prior information that could predict which substances are correlated. The problem is more easily remedied, however, by investigating agreement for each separate substance on the chemical checklist.

The violation of the assumption of constant within-rater and within-subject variance in the ANOVA models is another concern. The structure of the data used in each analyses was such that there were s' substances that had been

attributed present by at least one of the raters and there were (s-s') substances with completely null exposure information. The variance for this latter subset was identically zero whereas, for the former, the variances had some finite value. Therefore, homoscedasticity was violated when all s substances were considered simultaneously. Thus, Kappa can only be interpreted as an index that measures agreement after correcting for chance. In the case of assessing agreement for more than two observers, however, Kappa may not be interpretable as an index of agreement, since it was calculated directly from the ANOVA model (see Fleiss, 1981).

(6.2.4) Evaluation of Agreement for Substances on the Exposure Checklist

The second mode of analysis considered, separately for each substance, the extent of agreement between each pair of raters. The contingency tables that were used to produce these statistics had only two categories: exposed, at any level of confidence, and not exposed. An analysis was performed for each pair of raters. The index of crude agreement, κ, and κ' were calculated for each substance.

These tables were composed of small samples ($n<23$). Since estimates of Kappa obtained from tables having small samples may be highly unstable (see Chapter 3), more reliance was placed on the index of crude agreement as a measure of agreement. A qualitative scheme for interpretating agreement based on this index was used in this thesis (Table 6.2).

Table 6.2

Qualitative Interpretation of the Values of the Index of Crude Agreement for Contingency Tables having a Small Number of Observations

P_o	Qualitative Interpretation of Agreement
≥ 90%	Excellent
70 - 90%	Good
<70%	Poor

(6.3) Computer Program

A computer program was written in the FORTRAN language to facilitate the analysis. Values of P_o, κ (weighted and unweighted), asymptotic and jackknifed estimates of variance of κ, 95% confidence intervals of κ, and values of κ_{max} for pairs of raters are calculated. The program also calculates an arithmetic average of Kappa estimated for any number of tables and an average based on weighting each individual value by the reciprocal of its estimated variance. Asymptotic or jackknifed variances can be used in the calculations. Tests of homogeneity across strata can also obtained in the latter two cases (see Fleiss, 1981). The program is

driven from an SPSS (Nie et al., 1974; Hull and Nie, 1981) output file obtained by using the CROSSTABS procedure, with Option 10 specified. As a result, there is considerable flexibility in preparing data for analysis. Other measures of association and agreement calculated by SPSS can also be obtained, along with listings of the contingency tables. The KAPPA program is run in the second job step using the SPSS output file as data. Program options are implemented through the use of a LABELS file. FORTRAN NAMELIST parameters and labels for each Kappa table are specified in this file. Each output page is titled, date-stamped and labeled by the names of the two raters and the strata. This procedure of processing data is similar to the one reported by Dewar and Siemiatycki, 1984. A copy of the program is presented in Appendix 5. Further documentation and an up-to-date listing is available from the author.

(7.0) The Trials

(7.1) Agreement Trials Between Cancer Study Chemists

In two trials (ie, the General Comparison Trial and the Rubber Industry Trial) inter-rater agreement was assessed among individual chemists.

(7.1.1) The General Comparison Trial

(7.1.1.1) Materials and Methods

Five work histories from the Cancer Study were randomly selected from a set of files that had not yet been subjected to the consensus coding process. The five workers had held an aggregate of 20 different jobs. These jobs comprised the material for this trial. Time periods of employment, and occupation and industry classifications for each of these job descriptions are presented in Table A.7.1.

Four chemists from the Cancer Study (Chemists 1, 2, 4, and 6) independently imputed exposure using an exposure coding form containing 172 substances. Table 5.2 describes the professional experience of these raters.

(7.1.2.2) Results

Exposure was attributed for 105 of the 172 substances on the exposure coding form. The proportion of exposures ascribed at each level of confidence is presented in Table 7.1.1. On average, approximately seven exposures per job description were coded. Differences in the attribution of exposure between chemists, at the high level of confidence, were not

great. Nevertheless, there were significant differences at this level of confidence between Chemist 1 and Chemists 2 and 6 (Table A.7.2). Most of the differences in the estimates of prevalence were due to Chemists 1 and 6 attributing slightly more medium confidence level exposures.

Table 7.1.1

Proportion of Exposures Ascribed at each Level of Confidence in the General Comparison Trial (1)

Rater	Proportion of Exposures Attributed Present at Confidence Level: Low %	Medium %	High %	Any %
Chemist 1	0.4	1.7	2.9	5.0
Chemist 2	0.3	0.6	2.3	3.2
Chemist 4	0.0	0.2	2.6	2.8
Chemist 6	0.2	1.8	2.4	4.4
Average:	0.2	1.1	2.6	3.9

(1) The denominator used to calculate the percentages is equal to the total number of exposures considered by a rater. This is equal to 20 job descriptions X 172 substances (3440 exposures).

Observed and expected numbers of exposures for each pattern of concordance among the four raters is presented in Table 7.1.2. For each comparison, the observed number of exposures is significantly greater than would be expected by chance alone (Chi-square > 140; $p << 0.001$). Very good agreement among the four raters (Table 7.1.3) was also indicated by the high values of the Mean Majority Agreement Index (MMAI ≥ 0.96) and Kappa ($\kappa \geq 0.52$). Agreement was excellent at the high level of confidence (average over all raters: $p_o = 98\%$; $\kappa = 0.6$; Table 7.1.4) and was slightly lower at the medium or high level (average $\kappa = 0.54$) and at any level of confidence (average $\kappa = 0.52$). Agreement was good (average $\kappa = 0.44$; Table A.7.3) for the four exposure variables (ie, contact, concentration, frequency, and level of confidence) and the synthetic index; the lowest agreement was observed for the scale defining concentration ($\kappa = 0.40$).

Table 7.1.2

Pattern of Concordance by Level of Confidence Among Four Raters Ascribing Exposure in the General Comparison Trial (1)

Number of Raters Attributing Exposure:		Number of Exposures Observed and Expected at Confidence Level:					
		High		Medium or High		Any	
Present	Absent	Observed	Expected	Observed	Expected	Observed	Expected
4	0	37	0.0	46	0.0	46	0.0
3	1	24	0.2	32	0.7	36	0.6
2	2	26	12.7	42	24.6	43	25.1
1	3	79	325.1	137	475.5	152	448.9
0	4	3274	3102.0	3183	2939.2	3163	2965.4
Total (2):		3440	3440	3440	3440	3440	3440
Chi-Square (3):		143.3		196.3		145.0	

(1) The expected number of exposures represents the distribution of agreement calculated under the assumption of statistical independence. Agreement was calculated by categorizing the judgements of each rater into two categories: exposed, at 1 or more levels of confidence, and exposed, or not, at all other levels.
(2) The total number of exposures, 3440, is equal to 20 job descriptions x 172 substances.
(3) This is a test, on 1 df, of the null hypothesis that the observed numbers in the concordant cells (4-0 and 0-4) are equal to that expected by chance alone.

Table 7.1.3

Overall Agreement Among all Raters for Exposure Categorized at Three Levels of Confidence in the General Comparison Trial (1)

Index	Agreement Indices for Exposure Categorized at Confidence Level:		
	High	Medium or High	Any
MMAI	0.98	0.96	0.96
Kappa	0.60	0.53	0.52
95% C.I.	(.59, .61)	(.52, .55)	(.51, .53)

(1) A total of 3040 exposures = 20 job descriptions x 172 exposures were used in the calculations. Agreement was calculated by categorizing the judgements of each rater into two categories: exposed, at one or more levels of confidence, and exposed, or not, at all other levels.

Table 7.1.4

Pairwise Agreement Using the Collapsed Table Method for Exposure Categorized at Two Levels of Confidence in the General Comparison Trial(1)

Rater Pair	n_{22} (2)	p_0 %	κ	95% C. I. for κ	κ'
		High Confidence Level			
Chem 6 - Chem 1	54	97.8	0.58	0.49 - 0.67	0.64
Chem 6 - Chem 4	48	97.8	0.55	0.46 - 0.64	0.57
Chem 6 - Chem 2	48	98.1	0.58	0.49 - 0.68	0.59
Chem 1 - Chem 4	57	97.8	0.59	0.51 - 0.68	0.63
Chem 1 - Chem 2	57	98.1	0.63	0.54 - 0.71	0.72
Chem 4 - Chem 2	56	98.4	0.66	0.57 - 0.74	0.70
Average(3):	--	98.0	0.60	------	0.64
Average(4):	--	98.0	0.63	------	0.68
		Any Confidence Level			
Chem 6 - Chem 1	85	97.5	0.50	0.43 - 0.57	0.54
Chem 6 - Chem 4	64	97.6	0.50	0.42 - 0.58	0.65
Chem 6 - Chem 2	69	96.4	0.51	0.44 - 0.59	0.61
Chem 1 - Chem 4	68	96.1	0.49	0.41 - 0.56	0.70
Chem 1 - Chem 2	74	96.1	0.50	0.43 - 0.58	0.65
Chem 4 - Chem 2	67	97.9	0.64	0.56 - 0.72	0.69
Average(3):	--	96.9	0.52	------	0.64
Average(4):	--	96.7	0.54	------	0.68

Table 7.1.4, continued

(1) The total number of exposures, 3040, equals 20 job descriptions x 172 substances. Agreement was calculated by categorizing the judgements of each rater into two levels: exposed, at one or more levels of confidence, and exposed, or not, at all other levels.

(2) This represents the number of exposures attributed present by both raters.

(3) The average includes comparisons with Chemist 6.

(4) The average excludes comparisons with Chemist 6.

A separate analysis of agreement was performed for each of the 105 substances attributed by any of the four chemists. Detailed results are reported in Tables A.7.4 and A.7.5. Inter-rater agreement, averaged over all pairwise comparisons, is tabulated over four ranges of exposure prevalence which were also averaged over the chemists' assessments (Table 7.1.5). The values of both summary measures of agreement were fairly constant over the three non-zero prevalence ranges (average κ = 0.50 and p_o = 94.2%). However, there was a trend toward poorer agreement as the average prevalence of exposure increased (from p_o = 95.2% to p_o = 83.6%). The average value of Kappa for the 105 substances (κ = 0.50) was almost identical to that obtained using the collapsed table method over all 172 substances (κ = 0.52). There was excellent agreement for 91 of the 105 substances (ie, $p_o \geq 90\%$; Table 7.1.6). There was good agreement for 14 substances ($70\% \leq p_o < 90\%$) and none for which agreement was poor ($p_o < 70\%$).

Table 7.1.5

Average Pairwise Agreement for Exposure Categorized at Any Level of Confidence as a Function of Average Prevalence of Exposure in the General Comparison Trial

Average Prevalence(1)		Number of Substances in Prevalence Range		Average Agreement (2)	
Range %	Mean %	Number	% of Total	P_e %	κ
0 - 0	0.0	67	40.0	100.0	1.00
> 0 - 10	4.1	87	50.6	95.2	0.48
>10 - 20	14.7	14	8.1	90.8	0.61
>20	24.7	4	2.3	83.6	0.47
Totals (3):	6.3	105	61.1	94.2	0.50
(4):	4.3	172	100.0	96.4(6)	0.70
(5):	4.3	172	100.0	96.9(6)	0.52

Table 7.1.5, continued

(1) Based on the assessments of exposure to 20 job descriptions and averaged over the four assessments of exposure.

(2) Values of p_0 and κ for each substance were obtained by averaging agreement over all pairs of raters.

(3) Averaged over the 105 substances in which exposure was coded present by at least one rater.

(4) Averaged over all 172 substances on the exposure checklist. For the 67 (172-105) substances for which exposure was not coded, values of p_0 and κ were both set to unity.

(5) Averaged over the assessments of exposure using the collapsed table method (Table 7.1.4).

(6) The difference in these two numbers was due solely to rounding errors.

Table 7.1.6

Distribution of the Extent of Agreement for Individual Substances which were Attributed Present at Any Level of Confidence in the General Comparison Trial (1)

Degree of Agreement	Number of Substances		Average Agreement	
	Number	% of Total	p_o %	κ
Excellent	91	52.9	95.8	0.52
Good	14	8.1	83.9	0.37
Poor	0	0.0	----	----
Totals (2):	105	61.0	94.2	0.50
(3):	172	100.0	96.5(5)	0.70
(4):	172	100.0	96.9(5)	0.52

Table 7.1.6, continued

(1) Agreement for each individual substance, averaged over all four assessments of exposure, was categorized as follows:

P_o	
≥90%	Excellent
70-90%	Good
<70%	Poor

(2) Averaged over the 105 substances in which exposure was coded present in at least one job description by one of the four chemists.

(3) Averaged over all 172 substances on the exposure checklist. For the 67 (172-105) substances for which exposure was not coded in any job description, values of p_o and κ were set to unity.

(4) Averaged over the six pairs of raters using the collapsed table method (Table 7.1.4).

(5) The difference in these two numbers was due solely to rounding errors.

(7.1.2) The Rubber Industry Trial

(7.1.2.1) Materials and Methods

Fifteen job descriptions were selected from Cancer Study subjects who had been employed in the rubber industry. The industries in which these jobs were classified were distributed as follows: 80% among rubber products industries, 13% in the wholesale of motor vehicles and products, and 7% in publishing and printing. The specific occupations, industries and time periods of employment for each of these job descriptions are listed in Table A.8.1.

The three participating chemists from the Cancer Study (Chemists 1, 2, and 3) attributed exposure from a checklist containing 270 substances. Table 5.2 describes the professional experience of the raters. In addition, a fourth assessment of exposure (referred to as the *Final Coding*) was derived after a round-table meeting of the three chemists. At that time, a general consensus was reached with regard to assigning exposure to specific substances. Chemist 3 then reviewed each job description and re-evaluated exposure assignments.

(7.1.2.2) **Results**

Thirty-five of the 270 substances on the exposure coding form were attributed in at least one job description in any of the four codings. Twenty-nine substances were originally coded by the three chemists; an additional six were added in the Final Coding. The frequency of exposure at each level of confidence is exhibited in Table 7.1.7. In the original three assessments, the average proportion of exposure judged present, at any level of confidence, was 3.4%. This is roughly equivalent to attributing nine exposures per job description. The differences in the estimates of prevalence at the high level of confidence were small. There were significant differences ($p<0.01$), however, in these estimates at the medium or high, or at any level of confidence (Table A.8.2). This was partly due to Chemists 1 and 2 attributing a higher number of exposures at the medium level of confidence. There were many more exposures attributed in the Final Coding than in the original three.

A list of substances in which exposure was attributed in at least three of the 15 job descriptions is displayed in Table 7.1.8. (See Table A.8.4 for the complete list). Substances that were highly prevalent were rubber (average exposure prevalence = 100%), carbon black (82%), solvents (69%), pyrolysis and combustion fumes (67%), sulphur (58%), and aromatic napthas (56%). Large differences in estimates of prevalence between the raters were observed for some substances. Chemist 3 attributed no exposure for titanium dioxide and aromatic hydrocarbons and for the three solvents: benzene, toluene, and xylene.

This rater's estimates of prevalence for the general class of solvents was, however, in line with the other two raters. The estimates of prevalence from the Final Coding are also displayed in Table 7.1.8. Compared to the average of the coding of the three chemists, 14 of the 19 substances listed in this table were given higher estimates in the Final version while only three were given lower estimates. The differences in estimates varied from a few percentage points up to 100% (average difference of 20%).

Table 7.1.7

Proportion of Exposures Ascribed at Each Level of Confidence in the Rubber Industry Trial (1)

Rater	Proportion of Exposures Attributed Present at Confidence Level: Low %	Medium %	High %	Any %
Chemist 1	0.0	1.6	2.4	4.0
Chemist 2	0.3	1.0	2.3	3.6
Chemist 3	0.0	0.4	2.2	2.6
Average:	0.1	1.0	2.3	3.4
Final Coding	0.0	1.8	3.6	5.4

(1) The denominator used to calculate the percentages is equal to the total number of exposures considered by a rater. This is equal to 15 job descriptions X 270 substances (4050 exposures).

Table 7.1.8

Prevalence of Exposure for Selected Substances Ascribed at Any Level of Confidence in the Rubber Industry Trial (1)

Substance	Prevalence of Exposure as Attributed by:				
	Chem 1	Chem 2	Chem 3	Average of three chemists	Final
	%	%	%	%	%
Rubber	100	100	100	100	100
Carbon Black	87	80	80	82	100
Solvents	73	80	53	69	87
Pyrolysis & Combustion Fumes	80	80	40	67	80
Sulphur Dust	53	73	47	58	80
Aromatic Napthas	53	80	33	56	73
Aromatic hydrocarbons	67	80	0	49	7
Aromatic Amines	59	27	53	44	100
Aliphatic hydrocarbons	67	33	20	40	0

Table 7.1.8, continued

Substance	Prevalence of Exposure as Attributed by: Chem 1	Chem 2	Chem 3	Average of three chemists	Final
	%	%	%	%	%
Asbestos (Chrysotile)	53	40	27	40	60
Talc	33	47	13	31	53
Clay Dust	27	20	40	29	27
Benzene	53	20	0	24	53
Toluene	53	13	0	22	53
Xylene	53	13	0	22	53
Mineral Spirits	20	0	47	22	60
Adhesives	20	33	33	22	27
Lead Compounds	27	7	27	20	20
Titanium Dioxide	40	20	0	20	33

(1) Percentages are based on assessments of exposure to 15 job descriptions. Substances were selected for inclusion in this table if the average prevalence of the three chemist's codings was greater than 20% (ie, exposure was coded present in at least 3 job descriptions). See Table A.8.4 for the complete list. The judgements of each rater were categorized into exposed, at any level of confidence, and not exposed.

The pattern of concordance among the three chemists is presented in Table 7.1.9 for exposure classified by three levels of confidence. For each comparison, the observed number of exposures was significantly larger than would be expected by chance (Chi-square > 95.5). Very good agreement among the three chemists was also indicated by the high values of the Mean Majority Agreement Index (MMAI ≥ 0.97) and Kappa ($\kappa \geq 0.57$; Table 7.1.10). For each pair of raters, agreement was assessed at three composite levels of confidence. Between all pairs of raters, very good agreement was observed at the high, medium or high, or at any level of confidence (average κ = 0.58; average p_o = 97.7%; Table 7.1.11). Inter-rater agreement was generally good for all four exposure variables and the synthetic index (average κ = 0.49; Table A.8.3). The least amount of agreement was observed for concentration (average κ = 0.41).

Table 7.1.9

Pattern of Concordance by Level of Confidence Among Three Raters Ascribing Exposure in the Rubber Industry Trial (1)

Number of Raters Attributing Exposure:		Number of Exposures Observed and Expected at Confidence Level:					
		High		Medium or High		Any	
Present	Absent	Observed	Expected	Observed	Expected	Observed	Expected
3	0	44	0.1	58	0.1	61	0.2
2	1	32	6.3	67	12.9	70	13.8
1	2	85	266.7	96	337.0	94	393.0
0	3	3889	3776.9	3830	3660.0	3825	3643.0
	Total (2):	4050	4050	4050	4050	4050	4050
	Chi-square (3):	95.5		114.0		161.1	

(1) The expected number of exposures represents the distribution of agreement calculated under the assumption of statistical independence. Agreement was calculated by categorizing the judgements of each rater into two categories:- exposed, at one or more levels of confidence, and exposed, or not, at all other levels.
(2) The total number of exposures, 4050, is equal to 15 job descriptions x 270 substances.
(3) This is a test, on 1 df, of the null hypothesis that the observed numbers in the concordant cells (3-0 and 0-3) are equal to that expected by chance alone.

Table 7.1.10

Overall Agreement among Three Chemists for Exposure Categorized at Three Levels of Confidence in the Rubber Industry Trial(1)

Index	Agreement Indices for Exposure Categorized at Confidence Level: High	Medium or High	Any
MMAI	0.98	0.97	0.97
Kappa	0.57	0.61	0.59
95% C.I.	(.56, .59)	(.59, .63)	(.58, .61)

(1) The Final Coding is not included in these results. A total of 4050 exposures = 15 job descriptions x 270 exposures were used in the calculations. Agreement was calculated by categorizing the judgements of each rater into two categories: exposed, at one or more levels of confidence, and exposed, or not, at all other levels.

Table 7.1.11

Pairwise Agreement Using the Collapsed Table Method for Exposure Categorized at Two Levels of Confidence in the Rubber Industry Trial (1)

Rater Pair	n_{22} (2)	P_o %	κ	95% C. I. for κ	κ'
High Confidence Level					
Chem 1 - Chem 2	57	98.1	0.59	0.50 - 0.67	0.61
Chem 1 - Chem 3	59	98.3	0.62	0.54 - 0.70	0.65
Chem 2 - Chem 3	48	97.9	0.51	0.42 - 0.60	0.52
Final - Chem 1	74	97.7	0.60	0.52 - 0.67	0.75
Final - Chem 2	74	97.8	0.61	0.54 - 0.69	0.78
Final - Chem 3	76	98.0	0.64	0.57 - 0.71	0.84
Average(3):	--	98.1	0.57	-----	0.59
Average(4):	--	97.8	0.62	-----	0.79
Any Confidence Level					
Chem 1 - Chem 2	104	97.4	0.65	0.59 - 0.72	0.68
Chem 1 - Chem 3	79	97.3	0.57	0.50 - 0.65	0.74
Chem 2 - Chem 3	70	97.2	0.54	0.46 - 0.62	0.65
Final - Chem 1	133	97.2	0.69	0.63 - 0.74	0.80
Final - Chem 2	117	96.8	0.63	0.57 - 0.69	0.78
Final - Chem 3	95	96.8	0.58	0.51 - 0.64	0.91
Average(3):	--	97.3	0.59	-----	0.69
Average(4):	--	96.9	0.63	-----	0.84

Table 7.1.11, continued

(1) The total number of exposures is equal to 15 job descriptions x 270 substances = 4050. Agreement was calculated by categorizing the judgements of each rater into two levels: exposed, at one or more levels of confidence, and exposed, or not, at all other levels.

(2) This represents the number of exposures attributed present by both raters.

(3) The average excludes the comparisons of the individual chemist's assessments with that of the Final Coding.

(4) The average includes the comparisons of the individual chemist's assessments with that of the Final Coding.

As in the previous trial, a separate analysis of agreement was performed for each of the 35 substances which were attributed in any of the four codings. Exposure was categorized as exposed, at any level of confidence, and not exposed. Detailed results are reported in Tables A.8.4 and A.8.5. Inter-rater agreement, averaged over the three comparisons of the chemists, for five ranges of average exposure prevalence, is presented in Table 7.1.12. (Note that there were 29 substances attributed by the three chemists.) Agreement for each of the four non-zero prevalence ranges was about equal using the Kappa index (average κ = 0.28). The value of the index of crude agreement was much higher for those substances in the lowest prevalence range as compared to those in the higher prevalence ranges. Clearly, there was some discordance between the chemists in coding these substances, since excellent agreement was observed for only three of the 29 substances (ie, $p_o \geq 90\%$; Table 7.1.13). There was good agreement observed for 13 substances ($70\% \leq p_o < 90\%$), and there was an equal number of substances for which agreement was poor ($p_o < 70\%$).

Table 7.1.12

Average Pairwise Agreement for Exposure Categorized at Any Level of Confidence as a Function of Average Prevalence of Exposure in the Rubber Industry Trial

Average Prevalence(1)		Number of Substances in Prevalence Range		Average Agreement (2)	
Range	Mean	Number	% of	P_o	κ
%	%		Total	%	
0 - 0	0.0	241	83.3	100.0	1.00
> 0 - 10	5.3	5	1.9	90.3	0.22
>10 - 30	20.3	13	4.8	72.2	0.28
>30 - 50	40.9	5	1.9	61.8	0.28
>50	72.2	6	2.2	71.1	0.34
Totals (3):	32.0	29	10.8	73.3	0.28
(4):	3.4	270	100.0	97.1(6)	0.93
(5):	3.4	270	100.0	97.3(6)	0.59

Table 7.1.12, continued

(1) Based on assessments of exposure to 15 job descriptions and averaged over the original codings by the three chemists.

(2) Values of p_0 and κ for each substance were obtained by averaging agreement, for exposure assessed at any level of confidence, among the three chemists only.

(3) Averaged over the 29 substances for which exposure was coded present by any of the three chemists.

(4) Averaged over all 270 substances on the exposure checklist. For the 241 (270-29) substances for which exposure was not coded in any job description, values of p_0 and κ were set to unity.

(5) Averaged over the three pairs of raters using the collapsed table method (Table 7.1.11).

(6) The difference between these two numbers was due to rounding errors only.

Table 7.1.13

Distribution of the Extent of Agreement for Individual Substances which were Attributed Present at Any Level of Confidence in the Rubber Industry Trial

Agreement (1)	Number of Substances		Average Agreement	
	Number	% of Total	p_o %	κ
Excellent(2)	3	1.1	97.0	0.55
Good	13	4.8	80.2	0.27
Poor	13	4.8	62.3	0.23
Totals (3):	29	10.7	73.9	0.28
(4):	270	100.0	97.2(6)	0.92
(5):	270	100.0	97.3(6)	0.59

Table 7.1.13, continued

(1) Agreement for each individual substance, averaged over the three original assessments of exposure, was categorized as follows:

p_e	
≥90%	Excellent
70-90%	Good
<70%	Poor

(2) Excellent agreement was attained for the following substances (see Table A.8.5):

Rubber Abrasive dust

Floor dust

(3) Averaged over the 29 substances for which exposure was coded present in at least one job description by one of the three chemists.

(4) Averaged over all 270 substances on the exposure checklist. For the 241 (270-29) substances for which exposure was not coded in any job description, values of p_e and κ were set to unity.

(5) Averaged over the three pairs of raters using the collapsed table method (Table 7.1.11).

(6) The difference between these two numbers was due to rounding errors only.

(7.1.2.3) The Final Coding

The Final Coding, was derived after a meeting of the three chemists. It only approximates a true consensus rating since only one chemist re-assigned exposure based on the decisions reached during the meeting. A significantly larger number of exposures were attributed in this coding as compared to the original three. Six substances (inorganic pigments, iron dust, rubber dust ammonia, tin fumes, and combustion products of natural gas) were added to the list of 29 that were attributed originally. These substances, however, accounted for only 0.4% of the total prevalence of exposure (ie, 5.4%). Therefore, the original 29 substances were assigned in many more job descriptions than had been attributed in the original three codings. The collective knowledge of the three chemists may have been responsible for the larger proportion of exposures attributed. Perhaps, the effect was also due to a subconscious tendency by the raters to smooth-out disagreements by adjusting their ratings during the meeting. Furthermore, the chemist who recoded the files may have misinterpreted the decisions that had been made during the meeting.

(7.2) Trials that Compared the Assessments of Exposure of Individual Chemists and External Raters

The purpose of these trials (ie, the Paint Manufacturing and the Welding Trade trials) was to measure the degree of agreement in attributing exposure between individual chemists from the Cancer Study and external judges.

(7.2.1) The Paint Manufacturing Trial

(7.2.1.1) Materials and Methods

Five job descriptions were selected from Cancer Study subjects who had been employed in the paint manufacturing industry (Table A.9.1). Four jobs involved mixing or packaging raw materials while the fifth entailed operating a furnace which converted lead metal to lead oxide.

Three raters, two from the Cancer Study (Chemists 1 and 2) and an external rater (Judge 1), independently assessed exposure from a checklist containing 173 substances. (One substance was added to the exposure coding form containing 172 substances.) A description of the professional experience of the raters was presented in Tables 5.2 and 5.3.

(7.2.1.2) Results

The proportion of exposures attributed at each level of confidence is presented in Table 7.2.1. Forty-two of the 173 substances on the exposure checklist were coded present in at least one job description by at least

one rater. The average proportion of exposures imputed at any level of confidence was 6.4%, or about eleven exposures per job description. The Cancer Study chemists coded a greater number of exposures at the high confidence level than the external judge. The differences, however, were only significant ($p<0.05$; Table A.9.2) between the external judge and Chemist 1. In addition, Chemist 1 assigned about 4.5 times as many exposures at the medium level of confidence as did the other raters. For the most part, Chemist 2 and the external judge attributed exposure only when they were highly confident that the exposure had occurred.

Substances having an average prevalence of exposure greater than 40%, at any level of confidence, are presented in Table 7.2.2. (Estimates of prevalence for each of the 42 substances are given in Table A.9.4.) From this subset, all ten substances attributed by the external judge were ascribed by Chemist 1 and two substances (xylene and toluene) were not coded by Chemist 2. Chemist 2's estimate of prevalence for "solvents", however, agreed well with the other two raters.

Table 7.2.1

Proportion of Exposures Ascribed at Each Level of Confidence in the Paint Manufacturing Trial (1)

Rater	Proportion of Exposures Attributed Present at Confidence Level: Low %	Medium %	High %	Any %
Judge 1	0.0	0.7	3.7	4.4
Chemist 1	0.8	3.1	5.2	9.1
Chemist 2	0.8	0.7	4.3	5.8
Average:	0.5	1.5	4.4	6.4

(1) The denominator used to calculate the percentages is equal to the total number of exposures considered by a rater. This is equal to 5 job descriptions X 173 substances (865 exposures).

Table 7.2.2

Prevalence of Exposure for Selected Substances Ascribed at Any Level of Confidence in the Paint Manufacturing Trial (1)

Substance	Prevalence of Exposure as Attributed by: Judge	Chemist 1	Chemist 2	Average of three raters
	%	%	%	%
Paints and varnishes	100	100	80	93
Solvents	100	80	80	87
Aromatic Hydrocarbons	80	80	80	80
Mineral spirits	40	80	80	67
Metal Oxide Dust	40	80	60	60
Lead Compounds	40	100	20	60
Aliphatic Hydrocarbons	0	80	80	53
Linseed Oil	20	80	60	53
Xylene	80	60	0	47
Toluene	80	60	0	47
Organic Dyes and Pigments	20	60	40	40

Table 7.2.2, continued

(1) Percentages are based on assessments of exposure to 5 job descriptions. Substances were selected for inclusion in this table if the average prevalence of the three raters' codings was greater than 40% (ie, exposure was coded present in at least 2 job descriptions). See Table A.9.4 for the complete list.

For each pattern of concordance presented in Table 7.2.3, the number of observed exposures was significantly larger than would be expected by chance alone ($p<0.001$). Agreement among all raters was excellent at each level of confidence depicted in Table 7.2.4 (MMAI ≥ 0.95; $\kappa \geq 0.57$), and was greatest ($\kappa = 0.76$) at the high level of confidence. The results of a pairwise analysis (Table 7.2.5) indicated that there was very good agreement at all levels of confidence (average $p_o = 96.8\%$; average $\kappa = 0.59$). Agreement between the external judge and the two chemists was highest for exposure assessed at the high level of confidence. For the four exposure variables and the synthetic index (Table A.9.3), moderately good agreement was observed (average $\kappa = 0.43$) for the two comparisons with the external judge. Fair agreement (average $\kappa = 0.36$) was observed, however, for the scale defining concentration. A higher degree of agreement was observed for the comparison of the coding of the two chemists (average $\kappa = 0.51$).

Table 7.2.3

Pattern of Concordance by Level of Confidence Among Three Raters Ascribing Exposure in the Paint Manufacturing Trial (1)

Number of Raters Attributing Exposure:		Number of Exposures Observed and Expected at Confidence Level:					
		High		Medium or High		Any	
Present	Absent	Observed	Expected	Observed	Expected	Observed	Expected
3	0	18	0.1	23	0.2	23	0.1
2	1	18	4.8	22	8.2	28	5.6
1	2	24	104.4	40	136.2	42	125.6
0	3	805	755.7	780	720.4	772	733.7
	Total (2):	865	865	865	865	865	865
	Chi-square (3):	41.4		56.4		33.7	

(1) The expected number of exposures represents the distribution of agreement calculated under the assumption of statistical independence. Agreement was calculated by categorizing the judgements of each rater into two categories: exposed, at one or more levels of confidence, and exposed, or not, at all other levels.
(2) The total number of exposures, 865, is equal to 5 job descriptions x 173 substances.
(3) This is a test, on 1 df, of the null hypothesis that the observed numbers in the concordant cells (3-0 and 0-3) are equal to that expected by chance alone.

Table 7.2.4

Overall Agreement Among all Raters for Exposure Categorized at Three Levels of Confidence in the Paint Manufacturing Trial (1)

Index	Agreement Indices for Exposure Categorized at Confidence Level:		
	High	Medium or High	Any
MMAI	0.97	0.95	0.95
Kappa	0.76	0.57	0.66
95% C.I.	(.72, .80)	(.53,.61)	(.62, .69)

(1) A total of 865 exposures = 5 job descriptions x 173 substances were used in the calculations. A dichotomous exposure scale was used for each comparison: exposed at one or more levels of confidence and exposed, or not, at all other levels.

Table 7.2.5

Pairwise Agreement Using the Collapsed Table Method for Exposure Categorized at Two Levels of Confidence in the Paint Manufacturing Trial (1)

Rater Pair	n_{zz} (2)	p_o %	κ	95% C. I. for κ	κ'
High Confidence Level					
Judge 1 - Chem 1	25	97.4	0.69	0.57 - 0.82	0.78
Judge 1 - Chem 2	19	96.7	0.56	0.41 - 0.71	0.94
Chem 1 - Chem 2	42	97.8	0.62	0.52 - 0.73	0.82
Average(3):	--	97.1	0.63	-----	0.86
Average(4):	--	97.3	0.62	-----	0.85
Any Confidence Level					
Judge 1 - Chem 1	32	93.8	0.52	0.41 - 0.63	0.83
Judge 1 - Chem 2	23	97.1	0.50	0.37 - 0.63	0.58
Chem 1 - Chem 2	42	97.8	0.62	0.52 - 0.73	0.82
Average(3):	--	95.5	0.51	-----	0.71
Average(4):	--	96.2	0.55	-----	0.74

Table 7.2.5, continued

(1) A total of 865 exposures = 5 job descriptions x 173 substances were used in the calculations. A dichotomous exposure scale was used for each comparison: exposed at one or more levels of confidence and exposed, or not, at the other levels.

(2) This represents the number of exposures coded present by both raters.

(3) The average excludes the comparison of the two chemists' codings.

(4) The average includes the comparison of the two chemists' codings.

A pairwise analysis was performed (Table A.9.4) for each of the forty-two substances attributed by any of the raters. As in the previous trials, a dichotomous exposure scale was used. The average inter-rater agreement, between the two chemists and the external judge, was tabulated over four ranges of average exposure prevalence (Table 7.2.6). Both summary measures of agreement decrease with increasing prevalence. Excellent agreement was observed for 18 of the 42 substances (Table 7.2.7). There were 16 substances for which agreement was good and eight for which agreement was poor.

Table 7.2.6

Average Pairwise Agreement for Exposure Categorized at Any Level of Confidence as a Function of Average Prevalence of Exposure in the Paint Manufacturing Trial

Average Prevalence(1)		Number of Substances in Prevalence Range		Average Agreement (2)	
Range %	Mean %	Number	% of Total	P_o %	κ
0 - 0	0.0	131	75.7	100.0	1.00
> 0 - 20	12.5	26	15.0	86.2	0.44
>20 - 40	33.3	7	4.1	67.1	0.29
>40	62.2	9	5.2	60.0	0.19
Totals (2):	26.6	42	24.3	77.4	0.36
(3):	6.4	173	100.0	94.5(5)	0.85
(4):	6.4	173	100.0	95.5(5)	0.51

Table 7.2.6, continued

(1) Based on the assessments of exposure to five job descriptions and averaged over the three codings.

(2) Values of p_0 and ϵ for each substance were obtained by averaging agreement between the external judge and the two chemists.

(3) Averaged over all 173 substances on the exposure checklist. For the 131 (173-42) substances for which exposure was not coded, values of p_0 and ϵ were set to unity.

(4) Averaged over the two comparisons between the external judge and the chemists using the collapsed table method (Table 7.2.5).

(5) The difference in these two numbers was due solely to rounding errors.

Table 7.2.7

Distribution of the Extent of Agreement for Individual Substances which were Attributed Present at Any Level of Confidence in the Paint Manufacturing Trial

Agreement (1)	Number of Substances		Average Agreement	
	Number	% of Total	p_o %	κ
Excellent (2)	18	10.4	92.2	0.63
Good	16	9.3	76.9	0.23
Poor	8	4.6	45.0	0.03
Totals (3) :	42	23.3	77.4	0.36
(4) :	173	100.0	94.5(6)	0.85
(5) :	173	100.0	95.5(6)	0.51

Table 7.2.7, continued

(1) Agreement for each individual substance, averaged over the comparisons between the external judge and the two chemists, was categorized as follows:

P_e	
≥90%	Excellent
70-90%	Good
<70%	Poor

(2) Excellent agreement was attained for the following substances (see Table A.9.4):

Amphibole Asbestos	Silica Dust	Calcium Oxide
Metal Oxide Fumes	Paints, varnishes	Cobalt oxide
Copper oxide	Aliphatic alcohols	Heating oil
Methylene Chloride	Benzidine	Naphtylamine
o-toluidine	Paper dust	Coal dust

Pesticides

Combustion products of coal/coke

Aliphatic saturated halogens

Table 7.2.7, continued

(3) Averaged over the 42 substances for which exposure was coded present in at least one job description by one of the three raters.

(4) Averaged over all 173 substances on the exposure checklist. For the 131 (173-42) substances for which exposure was not coded in any job description, values of p_0 and κ were set to unity.

(5) Averaged over the two comparisons with the external rater using the collapsed table method (Table 7.2.5).

(6) The difference in these two numbers was due solely to rounding errors.

(7.2.2) The Welding Trade Trial

(7.2.2.1) Materials and Methods

Eighteen job descriptions from fourteen subjects who had been employed as welders or solderers were selected for this trial. Fourteen of these job descriptions were from welding occupations, and one each from industrial machinery, engine and related equipment, motor vehicle mechanics and repair, and teaching in a vocational institute. Employment occurred in industries as diverse as appliance manufacturing and pulp and paper. The specific occupations, industries and time periods of employment for each job description are listed in Table A.10.1.

Two chemists from the Cancer Study (Chemists 1 and 3) and a panel of three external judges (Judges 2) ascertained exposure from a checklist of 270 substances. The professional experience of the raters were described in Tables 5.2 and 5.3.

Exposure was attributed using a coding form containing 270 substances. In normal production coding, certain substances were assigned exposure levels automatically by the computer entry system when the presence of certain other substances were indicated. During the trial, it was assumed by the chemists that exposure for these additional substances would be automatically coded. Since this did not occur, the data for eleven substances was lost. Thus, only 259 substances were used in the analysis.

(7.2.2.2) **Results**

Exposure to 88 different substances was assigned in at least one job description by any of the raters. On average, there were 21 exposures per job description attributed by the three raters (average prevalence 8.2%). Chemist 1 assigned the largest number of exposures (9.7%) and Chemist 3 attributed the greatest number of high confidence level exposures (6.5%; Table 7.2.8). The chemists assigned, at any confidence level, one-quarter to one-half more exposures than did the panel of external raters. Most of the differences in the estimates of prevalence of exposure between the various pairs of raters were significant ($p<0.05$; Table A.10.2).

Substances that were assigned, at any level of confidence, an average prevalence of exposure greater than 30% are presented in Table 7.2.9. There was fairly good agreement among all raters in estimating prevalence for each of these substances. Four substances attributed by the chemists (metal dust, silicon carbide, mineral spirits, and solvents) were not coded present by the panel of external raters. (Estimates of prevalence for all 88 substances will be found in Table A.10.4.)

Table 7.2.8

Proportion of Exposures Ascribed at Each Level of Confidence in the Welding Trade Trial (1)

Rater	Proportion of Exposures Attributed Present at Confidence Level: Low %	Medium %	High %	Any %
Judges 2	0.9	1.0	4.7	6.6
Chemist 1	1.7	3.4	4.6	9.7
Chemist 3	0.0	1.9	6.5	8.4
Average:	0.9	2.1	5.3	8.2

(1) The denominator used to calculate the percentages is equal to the total number of exposures considered by a rater. This is equal to 18 job descriptions X 259 substances (4662 exposures).

Table 7.2.9

Prevalence of Exposure for Selected Substances Attributed at Any Level of Confidence in the Welding Trade Trial (1)

Substance	Prevalence of Exposure as Attributed by:			
	Judges 2	Chemist 1	Chemist 3	Average of three raters
	%	%	%	%
Metal Oxide Fumes	89	94	100	94
Nitrogen Oxides	83	94	100	93
Carbon Monoxide	89	94	83	89
Ozone	78	94	94	89
Arc Welding Fumes	83	94	83	89
Iron Fumes	83	89	83	85
Abrasive Dust	67	83	89	80
Iron Dust	72	83	72	76
Mn Fumes	83	61	89	76
Hydrofluoric acid	39	78	100	72
Hydrochloric acid	39	83	94	72
Acetylene	78	56	61	65
Gas Welding Fumes	72	50	72	65

Table 7.2.9, continued

Substance	Prevalence of Exposure as Attributed by: Judges	Chemist 1	Chemist 3	Average of three raters
	%	%	%	%
Mild Steel Dust	72	78	33	61
Metal Dust	0	94	83	59
Pyrolysis	6	94	78	59
Mineral Spirits	0	72	33	57
Pb Fumes	44	72	50	56
n Fumes	56	50	33	46
Silicon Carbide	0	78	56	45
Iron Oxides	78	0	56	44
Alumina	6	78	39	41
Solvents	0	72	33	35

(1) Percentages are based on assessments of exposure to 18 job descriptions. Substances were selected for inclusion in this table if the average prevalence of the three assessments of exposure was greater than 30% (ie, exposure was coded present in at least 5 job descriptions). See Table A.10.4 for the complete list.

The analyses of agreement, using the collapsed table method, is presented in Tables 7.2.10 to 7.2.14. For each comparison based on categorizing exposure at one or more levels of confidence, the observed number of exposures in the concordant cells was significantly larger than would be expected by chance (Table 7.2.10). Very good agreement among the three raters was also indicated by the high values of the Mean Majority Agreement Index (MMAI $\geq$ 0.93) and Kappa ($\kappa \geq 0.62$; Table 7.2.11). Agreement for each pair of raters is presented in Table 7.2.12. Excellent agreement between the panel of external judges and the two chemists was observed for exposure assessed at the high, medium or high, or at any level of confidence (average p_o = 96.0% and κ = 0.61). The degree of concordance was higher for the inter-chemist comparison. For the comparisons between the panel of judges and the two chemists, agreement was marginally good for the scales defining concentration and frequency (average κ = 0.35; Table A.10.3) and about equal for the remaining variables (average κ = 0.45). Agreement was better for the inter-chemist comparison (average over all variables κ = 0.56).

Table 7.2.10

Pattern of Concordance by Level of Confidence Among Three Raters Ascribing Exposure in the Welding Trade Trial (1)

Number of Raters Attributing Exposure:		Number of Exposures Observed and Expected at Confidence Level:					
		High		Medium or High		Any	
Present	Absent	Observed	Expected	Observed	Expected	Observed	Expected
3	0	127	0.7	180	1.8	196	2.5
2	1	104	36.3	145	69.3	162	85.4
1	2	147	662.0	198	884.5	237	968.6
0	3	4284	3963.0	4139	3706.4	4067	3605.5
Total (2):		4662	4662	4662	4662	4662	4662
Chi-square (3):		336.5		491.8		526.0	

(1) The expected number of exposures represents the distribution of agreement calculated under the assumption of statistical independence. Agreement was calculated by categorizing the judgements of each rater into two categories: exposed, at one or more levels of confidence, and exposed, or not, at all other levels.
(2) The total number of exposures, 4662, is equal to 18 job descriptions x 259 substances.
(3) This is a test, on 1 df, of the null hypothesis that the observed numbers in the concordant cells (3-0 and 0-3) are equal to that expected by chance alone.

Table 7.2.11

Overall Agreement Among all Raters for Exposure Categorized at Three Levels of Confidence in the Welding Trade Trial (1)

Index	Agreement Indices for Exposure Categorized at Confidence Level: High	Medium or High	Any
MMAI	0.96	0.94	0.93
Kappa	0.64	0.64	0.62
95% C.I.	(.62, .66)	(.62, .66)	(.60, .64)

(1) A total of 4662 exposures = 18 job descriptions x 259 exposures were used in the calculations. Agreement was calculated by categorizing the judgements of each rater into two levels: exposed, at one or more levels of confidence, and exposed, or not, at all other levels.

Table 7.2.12

Pairwise Agreement Using the Collapsed Table Method for Exposure Categorized at Two Levels of Confidence in the Welding Trade Trial(1)

Rater Pair	n_{ii} (2)	p_o %	κ	95% C. I. for κ	κ'
High Confidence Level					
Judges 2-Chem 1	139	96.7	0.63	0.57 - 0.68	0.64
Judges 2-Chem 3	170	96.1	0.63	0.58 - 0.68	0.76
Chem 1-Chem 3	176	96.5	0.66	0.62 - 0.71	0.82
Average(3):	---	96.4	0.63	-----	0.70
Average(4):	---	96.4	0.64	-----	0.74
Any Confidence Level					
Judges 2-Chem 1	223	93.3	0.55	0.51 - 0.60	0.70
Judges 2-Chem 3	224	97.6	0.61	0.57 - 0.66	0.70
Chem 1-Chem 3	303	97.9	0.69	0.65 - 0.73	0.75
Average(3):	---	95.5	0.58	-----	0.70
Average(4):	---	96.3	0.62	-----	0.72

Table 7.2.12, continued

(1) The total number of exposures, 4662, is equal to 18 job descriptions x 259 substances. Agreement was calculated by categorizing the judgements of each rater into 2 levels: exposed, at one or more levels of confidence, and exposed, or not, at all other levels.

(2) This represents the number of exposures attributed present by both raters.

(3) The average excludes the comparison of the two chemist's coding.

(4) The average includes the comparison of the two chemist's coding.

The detailed results of the analysis of agreement for the 88 substances attributed by any of the three raters is reported in Table A.10.4. Average inter-rater agreement, between the chemists and the panel of judges, is tabulated over five ranges of average exposure prevalence (Table 7.2.13). The index of crude agreement varied considerably across prevalence ranges. The maximum value was observed in the lowest prevalence range (p_o=93.1%) and the minimum value was in the 20-50% range (p_o=53.2%). There did not appear to be any such relationship for the Kappa index which, as noted previously, is not a good indicator of agreement when small samples are employed. Basing the extent of agreement on this former index, it was found that there was excellent agreement for 40 of the 88 substances in which exposure was attributed (Table 7.2.14). There were 22 substances for which agreement was good and 26 for which agreement was poor.

Table 7.2.13

Average Pairwise Agreement for Exposure Categorized at Any Level of Confidence as a Function of Average Prevalence of Exposure in the Welding Trade Trial

Average Prevalence(1)		Number of Substances in Prevalence Range		Average Agreement (2)	
Range %	Mean %	Number	% of Total	P_o %	κ
0 - 0	0.0	171	58.0	100.0	1.00
> 0 - 10	3.8	38	3.3	93.1	0.26
>10 - 20	14.0	13	4.4	78.1	0.34
>20 - 50	32.8	19	6.4	53.2	0.15
>50	74.3	18	6.1	70.4	0.37
Totals (3):	26.0	88	29.8	77.6	0.27
(4):	8.8	259	100.0	92.4(6)	0.75
(5):	8.8	259	100.0	95.5(6)	0.58

Table 7.2.13, continued

(1) Based on the assessments of exposure to 18 job descriptions and averaged over the three codings.

(2) Values of p_0 and κ for each substance were obtained by averaging agreement between the panel of external judges and the two chemists.

(3) Averaged over the 88 substances for which exposure was coded present in at least one job description.

(4) Averaged over all 259 of the 270 substances on the exposure checklist. For the 171 (259-88) substances for which exposure was not coded in any job descriiption, values of p_0 and κ were set to unity.

(5) Calculated using the collapsed table method (Table 7.2.12).

(6) The difference between these two values was due solely to rounding errors.

Table 7.2.14

Distribution of the Extent of Agreement for Individual Substances which were Attributed Present at Any Level of Confidence in the Welding Trade Trial

Agreement (1)	Number of Substances		Average Agreement	
	Number	% of Total	P_o %	κ
Excellent(2)	40	15.4	95.1	0.38
Good	22	8.5	81.5	0.30
Poor	26	10.0	47.7	0.07
Totals (3):	88	34.0	77.6	0.27
(4):	251	100.0	92.4(6)	0.75
(5):	251	100.0	95.5(6)	0.58

Table 7.2.14, continued

(1) Average agreement for each individual substance, between the panel of external judges and the two chemists, was categorized as follows:

P_s	
≥90%	Excellent
70-90%	Good
<70%	Poor

(2) Excellent agreement was attained for the following substances (see Table A.10.4):

Insulation material dust	Construction site dust	Mine dust
Cement dust	Glass fibres	Brass dust
Solder alloy dust	Concrete dust	Aluminum dust
Calcium oxide	Calcium carbide	Chrome dust
Nickel dust	Copper dust	inc dust
Silver dust	Cadmium dust	Tin dust
Coke dust	Cobalt dust	Natural gas
Arc welding fumes	Aluminum fumes	Iron fumes
Silver fumes	Aliphatic alcohols	Bronze dust
Ionizing radiation	Paints and varnishes	Cyanides
Nickel compounds	Caustic solution	Acrolein
Toluene	Turpentine	Cutting fluids
Perchlorethylene	Inorganic acid solution	
Combustion products of natural gas		
Combustion products of coal/coke		

Table 7.2.14, continued

(3) Averaged over the 88 substances for which exposure was coded present in at least one job description by one of the three raters.

(4) Averaged over all 259 substances under consideration. For the 171 (259-88) substances for which exposure was not coded in any job description, values of p_0 and ε were set to unity.

(5) Averaged over the comparisons of the chemists and the external judges using the collapsed table method (Table 7.2.12).

(6) The difference in these two numbers was due solely to rounding errors.

(7.3) Trials that Compared the Consensus Coding of the Chemists with that of External Judges

The purpose of these trials (the Metal Industry and the Chemical Manufacturing trials) was to measure the agreement in attributing exposure between external judges and Cancer Study chemists who used the consensus approach.

(7.3.1) The Metal Industry Trial

(7.3.1.1) Materials and Methods

Seven job descriptions were obtained by interviewing seven workers employed in two factories in the Montreal area that were engaged in the fabrication of metal products. The two plants were chosen by our collaborators at the Département de santé communataire à l'hôpital Sacre-Coeur (referred to as *DSC*) because empirical exposure data to occupational agents had been collected. The data was obtained through the DSC's environmental monitoring program of selected industries. The program was based on a walk-through of the plants by trained industrial hygienists. During these visits, the hygienists measured concentrations of some airborne materials in the vicinity of each employee. In addition, materials and machines used by the employees were recorded and a chemical breakdown of the materials were obtained from the manufacturer. Variation, over time, in the quality of the air and in the materials used could not be assessed since only one visit was made to each plant.

Workers were eligible for selection if: 1) they were employed at either of the two plants at the time of the trial, 2) industrial hygiene data had been collected for the worker, and 3) according to this data, the worker had been exposed to a "typical" set of exposures found in the plant's environment. In addition, subjects were chosen from different occupations which were fairly representative of the operations at the two plants (see Table A.11.1). The interview was performed at each plant by a bilingual interviewer from the Cancer Study. Only the part of the questionnaire concerned with each subject's current employment was administered.

It was intended that these files would be placed in the production stream of the Cancer Study. In order to prevent the chemists from realizing that a special trial was underway, it was necessary to create a lifetime work history for each subject. Previously coded job descriptions from the files of the Cancer Study were selected for the purpose of completing the work history of each subject. The criterion for selection of these other job descriptions was only that they should be reasonably compatible with each other. For example, job descriptions for company executives in unrelated fields were not admissibile. It became necessary, because of time constraints, to give the chemists all seven files at one time. Even though it was impossible to conceal from the chemists that these files were not part of the routine coding, it is believed that the coding was not affected.

Two chemists (Chemists 4 and 5) coded exposure using the consensus method. Exposure was assessed from an exposure coding form consisting

of three hundred substances. The chemists' professional experience was described in Table 5.2.

Two industrial hygienists from the DSC (Judges 3) coded exposure using information derived from the occupational monitoring program. The coding was done at two sessions with one of the principal investigators of the Cancer Study (Chemist 1) and the author present. These persons helped the hygienists quantify exposure to each substance. To ensure that no bias was introduced by the presence of the investigators, the following procedure was followed for each job description: First, without intervention, the presence or absence of each successive substance listed on the exposure coding form was indicated by the hygienists. Following the identification of an exposure, the hygienists assigned codes to each of the four exposure variables. At that point, and only then, did the researchers intervene if it was thought that the raters were not certain how to code the exposure. As a result of this process, it is believed, but by no means assured, that the criteria used for coding the exposure variables was roughly consistent with that used in the Cancer Study.

The hygienists performed two assessments for each subject. The first assessment entailed using the exposure data that had been collected during the occupational surveys. This coding was then updated after reviewing the job description, which had been previously obtained by interview. This latter coding of the external raters, which used the job description and the industrial hygiene data, was compared to the consensus coding of the chemists.

(7.3.1.2) **Results**

Sixty-five of the 300 substances on the exposure coding form were attributed by the raters in at least one job description. The proportion of substances attributed at each level of confidence is presented in Table 7.3.1. The proportion of exposures attributed by the chemists, at any level of confidence, was 4.7%. This is roughly equivalent to attributing 14 exposures per job description. The external judges coded a much larger number of exposures (5.8%). The difference in assigning exposure at the high level of confidence was not significant ($p>0.05$), but a significant difference was observed for exposure categorized at the medium or high and at any level of confidence (Table A.11.2).

A list of substances in which exposure was attributed in more than two job descriptions is given in Table 7.3.2. (See Table A.11.3 for the complete list.) Of the 18 substances listed in the table, four were not attributed by the chemists: silicon carbide, toluene, manganese fumes, and mineral oil.

Table 7.3.1

Proportion of Exposures Ascribed at each Level of Confidence in the Metal Industry Trial (1)

Rater (2)	Proportion of Exposures Attributed Present at Confidence Level: Low %	Medium %	High %	Any %
Consensus	0.2	0.3	4.2	4.7
Judges 3	0.1	1.0	4.7	5.8
Average:	0.2	0.7	4.5	5.3

(1) The denominator used to calculate the percentages is equal to the total number of exposures considered by a rater. This is equal to 7 job descriptions X 300 substances (2100 exposures).

Table 7.3.2

Prevalence of Exposure for Selected Substances Ascribed at Any Level of Confidence in the Metal Industry Trial (1)

Substance	Prevalence of Exposure as Attributed by: Consensus	External Judges	Average of two ratings
	%	%	%
Metallic dust	85.7	71.4	78.6
Stainless steel dust	71.4	71.4	71.4
Alumina	71.4	57.1	64.3
Arc welding fumes	71.4	57.1	64.3
Mild steel dust	42.9	71.4	57.2
Abrasive dust	57.1	57.1	57.1
Iron fumes	57.1	57.1	57.1
Carbon monoxide	57.1	57.1	57.1
Nitrogen oxides	57.1	57.1	57.1
Metal oxide fumes	57.1	57.1	57.1

Table 7.3.2, continued

Substance	Prevalence of Exposure as Attributed by: Consensus %	External Judges %	Average of two ratings %
Ozone	28.6	57.1	42.9
Solvents	28.6	57.1	42.9
Nickel fumes	28.6	57.1	42.9
Ultraviolet radn.	28.6	42.9	35.8
Silicon carbide	0.0	57.1	28.6
Toluene	0.0	57.1	28.6
Manganese fumes	0.0	57.1	28.6
Mineral oil	0.0	57.1	28.6

(1) Percentages were based on assessments of exposure to 7 job descriptions. Substances were selected for inclusion in this table if the average prevalence of the codings by consensus and by the panel of judges was greater than 30% (ie, exposure was coded present in at least 2 job descriptions). (See Table A.11.3 for the complete list.) The judgements of each rater were categorized into exposed, at any level of confidence, and not exposed.

Agreement was assessed for exposure categorized at the high, medium or high, and at any level of confidence. For each pattern of concordance corresponding to these classifications, the observed number of exposures in the concordant cells was greater than expected by chance alone (Table 7.3.3). Very good agreement was also indicated using the summary indices (average κ = 0.58 and average p_o = 96.1%; Table 7.3.4). Agreement was fair for each exposure variable (average κ = 0.46; Table 7.3.5); it was lowest for the scale defining concentration (κ = 0.38).

Table 7.3.3

Pattern of Concordance by Level of Confidence Among Two Raters Ascribing Exposure in the Metal Industry Trial (1)

Number of Raters Attributing Exp:		Number of Exposures Observed and Expected at Confidence Level:					
		High		Medium or High		Any	
Present	Absent	O	E	O	E	O	E
2	0	43	3.1	56	4.4	56	4.6
1	1	76	155.8	79	182.3	84	184.9
0	2	1981	1941.1	1965	1913.3	1960	1908.6
Total (2):		2100	2100	2100	2100	2100	2100
Chi-square (3):		44.2		64.1		60.6	

Table 7.3.3, continued

(1) The expected numbers represent the distribution of agreement calculated under the assumption of statistical independence. Agreement was calculated by categorizing the judgements of each rater into two categories: exposed, at one or more levels of confidence, and exposed, or not, at all other levels.

(2) The total number of exposures, 2100, is equal to 7 job descriptions x 300 substances.

(3) This is a test, on 1 df, of the null hypothesis that the observed numbers in the concordant cells (2-0 and 0-2) are equal to that expected by chance alone.

Table 7.3.4

Agreement using the Collapsed Table Method for Exposure Categorized at Three Levels of Confidence in the Metal Industry Trial (1)

	Agreement Indices for Exposure Attributed Present at Confidence Level:		
	High	Medium or High	Any
$n_{22}(2)$	53	66	68
p_0(%)	96.2	96.1	95.9
κ	0.55	0.59	0.59
95% C.I. for κ	.46 - .64	.51 - .67	.51 - .67
κ'	0.58	0.59	0.65

Table 7.3.4, continued

(1) The total number of exposures, 2100, is equal to 7 job descriptions X 300 substances. Agreement was calculated by categorizing the judgements of each rater into 2 levels: exposed, at one or more levels of confidence, and exposed, or not, at all other levels.

(2) This represents the number of exposures attributed present by both raters.

Table 7.3.5

Inter-Rater Agreement Using the Collapsed Table Method for the Four Exposure Variables and the Synthetic Index in the Metal Industry Trial (1)

	Contact	Frequency	Concentration	Level of Confidence	Synthetic Index(2)
p_o(%)	95.6	94.6	93.6	95.2	94.3
κ	0.51	0.47	0.38	0.52	0.44
95% CI for κ	.43 - .60	.40 - .54	.31 - .44	.45 - .60	.37 - .51
κ'	0.59	0.59	0.50	0.59	0.56

Table 7.3.5, continued

(1) The total number of exposures, 2100, is equal to 7 job descriptions x 300 substances. The original scales of measurement (four categories) were used for each of the four exposure variables.

(2) The synthetic index is defined as frequency x concentration x level of confidence and is categorized into three levels: no, "medium", and "high" exposure.

A separate analysis of agreement was performed for each of the 65 substances attributed in either of the two codings. Detailed results are reported in Table A.11.3. In Table 7.3.6, inter-rater agreement is presented for four ranges of average exposure prevalence. The index of crude agreement for each of the three non-zero prevalence ranges is about equal. The value of Kappa for the two upper prevalence ranges was fairly high ($\kappa \geq 0.49$), but was zero for the lowest range. Of these 65 substances, excellent agreement was observed for 13 (Table 7.3.7). Agreement was good for 45 substances and was poor seven substances.

Table 7.3.6

Average Pairwise Agreement for Exposure Categorized at Any Level of Confidence as a Function of Average Prevalence of Exposure in the Metal Industry Trial

Average Prevalence(1)		Number of Substances in Prevalence Range		Average Agreement	
Range %	Mean %	Number	% of Total	P_0 %	κ
0 - 0	0.0	235	78.3	100.0	1.0
> 0 - 10	6.1	20	6.7	85.7	0.0
>10 - 30	19.1	29	9.7	79.8	0.49
>30	54.6	16	5.3	77.7	0.60
Totals (2):	22.7	65	21.7	81.1	0.34
(3):	4.9	300	100.0	95.9	0.86
(4):	4.9	300	100.0	95.9	0.59

Table 7.3.6, continued

(1) Based on the assessment of exposure to seven job descriptions and averaged over the two ratings.

(2) Averaged over the 65 substances for which exposure was attributed in at least one job description.

(3) Averaged over all 300 substances on the exposure checklist. For the 235 (300-65) substances for which exposure was not coded in any job description, values of p_0 and ϵ were set to unity.

(4) Obtained from Table 7.3.4.

Table 7.3.7

Distribution of the Extent of Agreement for Individual Substances which were Attributed Present at Any Level of Confidence in the Metal Industry Trial

Agreement (1)	Number of Substances		Average Agreement	
	Number	% of Total	p_o %	κ
Excellent (2)	13	4.3	100.0	1.0
Good	45	15.0	80.9	0.21
Poor	7	2.3	47.0	-0.01
Totals (3):	65	21.7	81.1	0.34
(4):	300	100.0	95.9	0.86
(5):	300	100.0	95.9	0.59

Table 7.3.7, continued

(1) Agreement for each individual substance was categorized as follows:

p_e	
≥90%	Excellent
70-90%	Good
<70%	Poor

(2) Excellent agreement was attained for the following substances (see Table A.11.3):

Xylene	Toluene Di-isocyantate
Aliphatic esters	Pyrolysis and combustion fumes
Soldering fumes	Aluminium fumes
Silver fumes	Lead fumes
Carbon black	Paints and lacquers
Carbon monoxide	Nitrogen oxides
Inorganic pigments	

(3) Averaged over the 65 substances for which exposure was coded present in at least one job description.

(4) Averaged over all 300 substances on the exposure checklist. For the 235 (300-65) substances for which exposure was not coded in any job description, values of p_e and κ were set to unity.

(5) Averaged over all ratings using the collapsed table method (Table 7.3.4).

(7.3.2) The Chemical Manufacturing Trial

(7.3.2.1) Materials and Methods

Job descriptions for this trial were produced from job function sheets supplied by the industrial hygiene department of a large Canadian chemical manufacturer. Enumerated on each of these forms was the job title, the department in the plant in which the worker was employed, the precise functions to be performed by the employee, and the machines, tools and materials that would be used on a regular basis. Five of these job functions were chosen from different areas of one of the manufacturer's plants. These were then transformed into job descriptions, with the help of an industrial hygienist from the company (Judge 4) and an interviewer from the Cancer Study. Materials, machines, and protective equipment were included in the job description. In addition, the general working environment in the area in which the job was supposed to have taken place was also mentioned. The job descriptions lacked details that would normally be brought out during an interview (eg, frequency of chemical spills, etc). Nevertheless, it was felt that these job descriptions could provide a sufficient amount of information that would enable the raters to accurately code exposure.

As in the previous trial, it was intended that the Cancer Study chemists would review these job descriptions as part of the production coding process. In order to prevent them from realizing that a special trial was underway, a lifetime job history was created for each hypothetical subject by selecting previously coded job descriptions

from the files of the Cancer Study. The criteria for selecting these job descriptions have been previously described in the Materials and Methods section of the Metal Industry Trial.

Two chemists (Chemists 4 and 5) coded exposure using the consensus method. Exposure was assessed from an exposure coding form consisting of 300 substances. The professional experience of the chemists was illustrated in Table 5.2. The industrial hygienist from the company coded exposure from information on the job function sheets and other industrial hygiene data on record, and from his general knowledge of the industrial environment. The coding was done at two sessions with one of the principal investigators of the Cancer Study (Chemist 1) and the author present at the first session only. The procedures were identical to that described in the Materials and Methods section of the Metal Industry Trial. Two job descriptions were coded in the first session and the remaining three were coded by the external judge alone. The occupations and time periods of employment are presented in Table A.12.1.

(7.3.2.2) **Results**

Thirty-four of the 300 substances on the exposure checklist were attributed in at least one job description by at least one rater. The proportion of substances attributed at each level of confidence is presented in Table 7.3.8. About six exposures per job description were attributed by the chemists (prevalence of exposure = 2.0%). Judge 4 assigned approximately nine exposures per job description (3.0%). The differences in the estimates of the prevalence of exposure at all levels of confidence were significant ($p<0.05$; Table A.12.2).

A list of substances in which exposure was attributed in at least one job description is presented in Table 7.3.9. (See Table A.12.3 for the complete list.) Of this subset, acetic acid and cellulose acetate were highly prevalent (average exposure prevalence of 80% and 60%, respectively).

For each pattern of concordance listed in Table 7.3.10 the number of observed exposures was greater than that expected by chance alone. Good agreement was observed at each of the exposure levels (average κ = 0.47 and p_0 = 97.7%; Table 7.3.11). Agreement was fair and relatively constant for the four exposure variables and the synthetic index (average κ = 0.39; Table 7.3.12).

Table 7.3.8

Proportion of Exposures Ascribed at Each Level of Confidence in the Chemical Manufacturing Trial (1)

Rater	Proportion of Exposures Attributed Present at Confidence Level: Low %	Medium %	High %	Any %
Consensus	0.2	0.3	1.5	2.0
Judge 4	0.3	0.2	2.5	3.0
Average:	0.3	0.3	2.0	2.5

(1) The denominator used to calculate the percentages is equal to the total number of exposures considered by a rater. This is equal to 5 job descriptions X 300 substances (1500 exposures).

Table 7.3.9

Prevalence of Exposure for Selected Substances Ascribed at Any Level of Confidence in the Chemical Manufacturing Trial (1)

Substance	Prevalence of Exposure as Attributed by: Consensus %	Judge 4 %	Average of two ratings %
Acetic acid	60.0	100.0	80.0
Cellulose acetate	60.0	60.0	60.0
Acetone	40.0	40.0	40.0
Alkali, caustic solution	40.0	40.0	40.0
Aliphatic alcohols	40.0	40.0	40.0
Mineral spirits	40.0	20.0	30.0
Aliphatic ketones	20.0	40.0	30.0

Table 7.3.9, continued

(1) Percentages are based on assessments of exposure to 5 job descriptions. Substances were selected for inclusion in this table if the average prevalence was greater than 20% (ie, exposure was coded present in at least 1 job description). See Table A.12.3 for the complete list.

Table 7.3.10

Pattern of Concordance by Level of Confidence Among Two Raters Ascribing Exposure in the Chemical Manufacturing Trial (1)

Number of Raters Attributing Exp:		Number of Exposures Observed and Expected at Confidence Level:					
		High		Medium or High		Any	
Present	Absent	O	E	O	E	O	E
2	0	13	0.6	16	0.7	20	0.9
1	1	32	56.9	35	65.6	35	73.2
0	2	1455	1442.5	1449	1433.7	1445	1425.9
Total (2):		1500	1500	1500	1500	1500	1500
Chi-square (3):		11.3		14.9		21.0	

Table 7.3.10, continued

(1) The expected numbers represent the distribution of agreement calculated under the assumption of statistical independence. Agreement was calculated by categorizing the judgements of each rater into two categories: exposed, at one or more levels of confidence, and exposed, or not, at all other levels.

(2) The total number of exposures, 1500, is equal to 5 job descriptions x 300 substances.

(3) This is a test, on 1 df, of the null hypothesis that the observed numbers in the concordant cells (2-0 and 0-2) are equal to that expected by chance alone.

Table 7.3.11

Pairwise Agreement Using the Collapsed Table Method for Exposure Categorized at Three Levels of Confidence in the Chemical Manufacturing Trial (1)

	Agreement Indices for Exposure Categorized at Confidence Level:		
	High	Medium or High	Any
$n_{11}(2)$	13	16	20
$p_o(\%)$	97.8	97.7	97.7
κ	0.43	0.47	0.52
95% C.I. for κ	.27 - .59	.32 - .62	.38 - .66
κ'	0.58	0.59	0.65

Table 7.3.11, continued

(1) The total number of exposures, 1500, is equal to 5 job descriptions X 300 substances. Agreement was calculated by categorizing the judgements of each rater into 2 levels: exposed, at one or more levels of confidence, and exposed, or not, at all other levels.

(2) This represents the number of exposures attributed present by both raters.

Table 7.3.12

Inter-Rater Agreement Using the Collapsed Table Method for the Four Exposure Variables and the Synthetic Index in the Chemical Manufacturing Trial (1)

	Contact Frequency	Concentration	Level of Confidence	Synthetic Index(2)
p_0 (%)	97.0	97.0	96.7	97.3	97.2
κ	0.39	0.39	0.32	0.44	0.43
95% CI for κ	.27 - .51	.27 - .51	.22 - .42	.30 - .57	.30 - .56
κ'	0.50	0.56	0.49	0.57	0.54

(1) The total number of exposures, 1500, is equal to 5 job descriptions X 300 substances. The original scales of measurement (four categories) were used for each of the four exposure variables.

(2) The synthetic index is defined as concentration x frequency x level of confidence, grouped into 3 categories: no, "medium", and "high" exposure.

A separate analysis of agreement was performed for each of the 34 substances attributed by any of the raters. Detailed results are reported in Table A.12.3. An analysis, stratified by average prevalence, was not performed since the majority of these substances (27 of the 34) were coded with an exposure prevalence less than 20%. Excellent agreement was observed for eleven substances, good agreement for twelve substances, and poor agreement for eleven substances (Table 7.3.13).

Table 7.3.13

Distribution of the Level of Agreement for Individual Substances which were Attributed Present at Any Level of Confidence in the Chemical Manufacturing Trial

Agreement (1)	Number of Substances		Average Agreement	
	Number	% of Total	p_0 %	ε
Excellent (2)	11	3.7	100.0	1.0
Good	12	4.0	80.0	0.05
Poor	11	3.7	58.2	-0.06
Totals (3):	34	11.3	79.4	0.32
(4):	300	100.0	97.7	0.92
(5):	300	100.0	97.7	0.52

Table 7.3.13, continued

(1) Agreement for each individual substance was categorized as follows:

P_e	
≥90%	Excellent
70-90%	Good
<70%	Poor

(2) Excellent agreement was attained for the following substances (see Table A.13.3):

Titanium dioxide	Natural gas
Ozone	Formaldehyde
Caustic acid solution	Benzene
Cellulose acetate	Aliphatic alcohols
Aliphatic aldehydes	Aliphatic esters
Combustion products of natural gas	

(3) Averaged over the 34 substances for which exposure was coded present in at least one job description.

(4) Averaged over all 300 substances on the exposure checklist. For the 266 (300-34) substances in which exposure was not coded in any job description, values of p_e and ε were set to unity.

(5) Obtained using the collapsed table method (Table 7.3.11).

(7.4) The Code/Recode Trial

The purpose of this trial was to evaluate the reliability, or repeatability, of attributing exposure by consensus for a team of Cancer Study chemists.

(7.4.1) Materials and Methods

In the Metal Industry and Chemical Manufacturing trials twelve job descriptions were obtained from sources exterior to the Cancer Study. It was intended that these files would be processed by the team of chemists over a period of four months. In order to prevent the chemists from realizing that a special trial was underway, it was necessary to create a lifetime work history for each subject. For this purpose, previously coded job descriptions were haphazardly selected from the files of the Cancer Study. The criteria for selecting these other job descriptions were only that they should be reasonably compatible with each other. In addition, only job descriptions were selected if no substantial changes had to be made to the time period of employment. As a result of this process, 23 job descriptions were selected. Time periods of employment, and the occupations and industries in which employment occurred is presented in Table A.13.1.

Two chemists (Chemists 4 and 5) participated in both the original coding (referred to as the *First Code*) and the second coding (referred to as the *Recode*). In both cases the coding was performed using the consensus method.

The job descriptions that were utilized in this trial arose from different time periods in the life of the Cancer Study, when coding sheets containing different numbers of substances were used. In none of the job descriptions, however, were substances attributed in the Recode which did not appear on the exposure checklist used in the original coding. Two hundred and seventy substances were used as the basis for the analysis.

(7.4.2) Results

One hundred and eight of the 270 exposures were attributed in at least one of the codings. The proportion of substances attributed at each level of confidence is presented in Table 7.4.1. The proportion of exposures judged present, at any level of confidence, was about 2.3% in both codings, which translates to about six exposures per job description. There were no significant differences at each of the levels of confidence listed in the table (Table A.13.2).

For each pattern of concordance, the number of observed exposures was greater than that expected by chance (Table 7.4.2). Very good agreement was observed at each of the three levels of confidence (average $\kappa = 0.65$, $p_o = 98.5\%$; Table 7.4.3). Agreement for each of the four exposure variables and the synthetic index was also quite good (average $\kappa = 0.60$; Table 7.4.4). Agreement was least for the variable measuring concentration ($\kappa = 0.53$).

A separate analysis of agreement was performed for each of the 108

substances attributed in either coding (Table A.13.3). An analysis, stratified by average prevalence, was not performed since the majority of the substances (105 of the 108) were coded with an exposure prevalence less than 20%. There was excellent agreement for 102 substances (Table 7.4.5); there were six substances in which agreement was good and none in which agreement was poor.

Table 7.4.1

Proportion of Exposures Ascribed at each Level of Confidence in the Code/Recode Trial (1)

Coding	Proportion of Exposures Attributed Present at Confidence Level: Low %	Medium %	High %	Any %
First Code	0.02	0.39	1.92	2.32
Recode	0.06	0.37	1.95	2.38
Average:	0.04	0.38	1.94	2.34

(1) The denominator used to calculate the percentages is equal to the total number of exposures considered by a rater. This is equal to 23 job descriptions X 270 substances (6210 exposures).

Table 7.4.2

Pattern of Concordance by Level of Confidence Among Two Raters Ascribing Exposure in the Code/Recode Trial (1)

Number of Raters Attributing Exp:		Number of Exposures Observed and Expected at Confidence Level:					
		High		Medium or High		Any	
Present	Absent	O	E	O	E	O	E
2	0	74	2.3	97	3.3	99	3.4
1	1	92	234	93	280.4	94	285.2
0	2	6044	5972.3	6020	5926.3	6017	5921.4
Total (2):		6210	6210	6210	6210	6210	6210
Chi-square (3):		90.8		131.2		134.4	

Table 7.4.2, continued

(1) The expected numbers represent the distribution of agreement calculated under the assumption of statistical independence. Agreement was calculated by categorizing the judgements of each rater into two categories: exposed, at one or more levels of confidence, and exposed, or not, at all other levels.

(2) The total number of exposures, 6210, is equal to 23 job descriptions x 270 substances.

(3) This is a test, on 1 df, of the null hypothesis that the observed numbers in the concordant cells (2-0 and 0-2) are equal to that expected by chance alone.

Table 7.4.3

Agreement Using the Collapsed Table Method for Exposure Categorized at Three Levels of Confidence in the Code/Recode Trial (1)

	Agreement Evaluated for Exposure Categorized at Confidence Level:		
	High	Medium or High	Any
$n_{22}(2)$	74	97	99
$p_0(\%)$	98.5	98.5	98.5
κ	0.61	0.67	0.67
95% C.I. for κ	.54 - .68	.61 - .73	.61 - .73
κ'	0.61	0.67	0.68

Table 7.4.3, continued

(1) The total number of exposures, 6210, is equal to 23 job descriptions X 270 substances. The original scales of measurement (four categories) were used for each of the four exposure variables.

(2) This represents the number of exposures attributed present in both assessments.

Table 7.4.4

Inter-Rater Agreement Using the Collapsed Table Method for the Four Exposure Variables and the Synthetic Index in the Code/Recode Trial (1)

	Contact Frequency	Concentration	Level of Confidence	Synthetic Index(2)
p_o(%)	98.3	98.1	97.8	98.3	98.2
κ	0.63	0.58	0.53	0.63	0.61
95% CI for κ	.57 - .69	.52 - .64	.47 - .59	.56 - .69	.55 - .67
κ'	0.66	0.62	0.54	0.64	0.62

(1) The total number of exposures, 6210, is equal to 23 job descriptions X 270 substances.

(2) The synthetic index is defined as concentration x frequency x level of confidence, broken down into three categories: no, "medium", and "high" exposure.

Table 7.4.5

Distribution of the Level of Agreement for Individual Substances which were Attributed Present at Any Level of Confidence in the Code/Recode Trial

Agreement (1)		Number of Substances		Average Agreement	
		Number	% of Total	P_o %	κ
Excellent		102	37.8	96.2	0.55
Good		6	2.2	84.1	0.25
Poor		0	0.0	--	--
Totals	(2):	108	40.0	95.5	0.53
	(3):	270	100.0	98.2(5)	0.81
	(4):	270	100.0	98.5(5)	0.67

Table 7.4.5, continued

(1) Agreement for each individual substance was categorized as follows:

p_e	
≥90%	Excellent
70-90%	Good
<70%	Poor

(2) Averaged over the 108 substances in which exposure was attributed by at least one rater.

(3) Averaged over all 270 substances under consideration. For the 162 (270-108) substances for which exposure was not coded in any job description, values of p_e and κ were set to unity.

(4) Obtained using the collapsed table method (Table 7.4.3).

(5) The difference in these two numbers was due solely to rounding errors.

(7.5) Synthesis of Results

(7.5.1) Agreement in Attributing Exposure at Different Levels of Confidence

In five of the seven trials, no significant differences between raters were observed in estimating prevalence of exposure at the high level of confidence (Table 7.5.1). Significant differences were observed, however, in the Chemical Manufacturing and in the Welding Trade trials. Few exposures were attributed at the low level of confidence. Thus, exposure was attributed only when the raters felt fairly certain that the exposure had occurred. Although not apparent from this summary table, the chemist with the greatest amount of experience (Chemist 1) attributed a higher proportion of exposures in every trial in which he participated.

In Table 7.5.2 the average value of Kappa (and κ'), taken over all pairs of raters in each trial, is summarized for exposure categorized at the high and at any level of confidence. In every trial, good to excellent agreement was observed at each of these levels of confidence. (Similar results were observed for exposure categorized at the medium or high level of confidence.) Averaged over all pairs of raters in all trials, the extent of inter-rater agreement was about equal at all levels of confidence considered in this thesis (average $\kappa = 0.59$ and $\kappa' = 0.70$).

Table 7.5.1

Average Proportion of Exposures Ascribed at Two Levels of Confidence in all Trials of Inter-Rater Agreement (1)

Trial	No. of Raters	Average Proportion of Exposure Attributed at Confidence Level: High Avg. %	High S.D. %	Any Avg. %	Any S.D. %
Comparisons Among Cancer Study Chemists					
Rubber Ind. (2)	3	2.3	0.1	3.4*	0.7
General Comp.	4	2.6	0.3	3.9*	1.0
Paint Mfg. (3)	2	4.8	0.7	7.5*	2.3
Welding Trade (3)	2	5.6*	1.3	9.1*	0.9
Comparisons Between Individual Chemists and External Judges					
Paint Mfg. (4)	3	4.4	0.8	6.4*	2.4
Welding Trade (4)	3	5.3*	1.1	8.2*	1.6

Table 7.5.1, continued

Trial	No. of Raters	Average Proportion of Exposure Attributed at Confidence Level: High Avg. %	High S.D. %	Any Avg. %	Any S.D. %
Comparisons Between Consensus Coding and External Judges					
Metal Ind.	2	4.5	0.4	5.3*	0.8
Chemical Mfg.	2	2.0*	0.7	2.5*	0.7
Reliability of Consensus Coding					
Code/Recode	2	1.9	0.02	2.3	0.04

Table 7.5.1, continued

(1) The exposure prevalence was averaged over all raters, except where noted below. "*" indicates that differences in the proportion of exposures between participating raters was significant at $p<0.05$. Significant differences in the attribution of exposure were evaluated, for pairs of raters, using McNemar's test. Cochran's Q-index was used for situations involving multiple raters. "Avg." represents the average proportion of exposure (in %) for the number of raters listed in column two. "S.D." is the standard deviation.

(2) Excludes the Final Coding.

(3) Excludes the comparison of the two chemists with the external judges.

(4) Excludes the comparison between the chemists.

Table 7.5.2

Average Agreement for Exposure Categorized at Two Levels of Confidence in all Trials of Inter-Rater Agreement (1)

Trial	No. of Raters	Avg. p_o %	Average Agreement for Exposure Categorized at Confidence Level: High κ	High κ'	Any κ	Any κ'
Comparisons Among Cancer Study Chemists						
Rubber Ind. (2)	3	97.9	0.57	0.59	0.59	0.69
General Comp.	4	97.6	0.60	0.64	0.52	0.64
Paint Mfg. (3)	2	97.6	0.63	0.86	0.62	0.82
Welding Trade (3)	2	96.8	0.66	0.82	0.69	0.75
Average:		97.5	0.62	0.73	0.61	0.73
Comparisons Between Individual Chemists and External Judges						
Paint Mfg. (4)	3	96.7	0.63	0.86	0.51	0.71
Welding Trade (4)	3	96.1	0.63	0.70	0.58	0.70
Average:		96.4	0.63	0.78	0.55	0.71

Table 7.5.2, continued

Trial	No. of Raters	Avg. p_o %	Average Agreement for Exposure Categorized at Confidence Level: High κ	High κ'	Any κ	Any κ'
Comparisons Between Consensus Coding and External Judges						
Metal Ind.	2	96.9	0.55	0.59	0.59	0.67
Chemical Mfg.	2	98.4	0.43	0.58	0.52	0.65
Average:		97.7	0.49	0.59	0.56	0.66
Reliability of Consensus Coding						
Code/Recode	2	98.9	0.61	0.61	0.67	0.68
Overall Average:		97.5	0.59	0.69	0.59	0.70

Table 7.5.2, continued

(1) Agreement was calculated by first categorizing the judgements of each rater into two levels: exposed, at one or more levels of confidence, and exposed, or not, at all other levels. Averages of κ and κ' are presented over all pairs of raters participating in each trial, except where noted below.

(2) Excludes the Final Coding.

(3) Excludes the comparison of the external judges with the chemists.

(4) Excludes the comparison between the two chemists.

(7.5.2) Summary of Agreement for the Exposure Variables

For each of the four exposure variables (ie, contact, frequency, concentration, and level of confidence) and the synthetic index, inter-rater agreement was averaged among all pairs of raters who participated in each of the trials (Table 7.5.3). Averaged over all comparisons, agreement was moderately good for all five variables (average κ = 0.48). Without exception, concordance was least for the scales defining intensity of exposure and highest for level of confidence.

Table 7.5.3

Average Agreement, using Kappa, for the Four Exposure Variables and the Synthetic Index in all Trials of Inter-Rater Agreement (1)

Trial	No. of Raters	Avg. p_o (2) %	Average Agreement for: Contact κ	Freq. κ	Conc. κ	Level of Conf. κ	Synth. Index κ
Comparisons Among Cancer Study Chemists							
Rubber Ind. (3)	3	96.6	0.52	0.50	0.41	0.51	0.52
General Comp.	4	95.7	0.45	0.42	0.40	0.47	0.45
Paint Mfg. (4)	2	93.1	0.49	0.48	0.47	0.55	0.57
Welding Trade (4)	2	92.8	0.69	0.53	0.46	0.57	0.54
Average:		94.6	0.54	0.48	0.44	0.53	0.52
Comparisons Between Individual Chemists and External Judges							
Paint Mfg. (5)	3	93.5	0.41	0.43	0.36	0.47	0.51
Welding Trade (5)	3	90.8	0.52	0.34	0.36	0.45	0.39
Average:		92.2	0.47	0.39	0.36	0.46	0.45

Table 7.5.3, continued

Trial	No. of Raters	Avg. p_o (2) %	Average Agreement for: Contact κ	Freq. κ	Conc. κ	Level of Conf. κ	Synth. Index κ
Comparisons Between Consensus and External Judges							
Metal Ind.	2	94.7	0.51	0.47	0.38	0.52	0.44
Chemical Mfg.	2	97.0	0.39	0.39	0.32	0.44	0.43
Average:		95.9	0.45	0.43	0.35	0.48	0.44
Reliability of Consensus Coding							
Code/Recode	2	98.1	0.63	0.58	0.53	0.63	0.61
Overall Average:		94.5	0.51	0.46	0.41	0.51	0.50

Table 7.5.3, continued

(1) Agreement was calculated for the four exposure variables (contact, frequency, concentration, and level of confidence) using the original four-point scales of measurement. For the synthetic index, which was defined as frequency x concentration x level of confidence, agreement was obtained by categorizing the values into three levels: no, "medium", and "high" exposure. Thereafter, an average was taken over all pairs of raters, except where noted below.

(2) Averaged over the 5 variables.

(3) Excludes the Final Coding.

(4) Excludes the comparison of the chemists with the external judges.

(5) Excludes the comparison between the two chemists.

(7.5.3) Summary of Agreement for those Substances Attributed

A summary of the extent of agreement for those substances which were attributed is presented in Table 7.5.4. Substances are omitted from consideration here if the various raters agreed unanimously that there was no exposure. The fourth column of the table represents the index of crude agreement averaged over all pairs of raters and over all attributed substances. The number of substances in which excellent and good agreement was observed is presented in the last four columns. There was excellent agreement observed for 56% of those substances attributed and good agreement was observed for 29% of them. On average, there was good agreement in attributing the presence or absence of exposure (average p_o = 83%).

Table 7.5.4.

Level of Agreement for those Substances Attributed at Any Level of Confidence in all Trials of Inter-Rater Agreement (1)

Trial	Total No. of Substances	No. of Substances Attributed (a)	Avg. p_o %	Number of Substances in which Agreement was (2): Excellent No.	Excellent % of (a)	Good No.	Good % of (a)
Comparisons Among Cancer Study Chemists							
Rubber Ind. (3)	270	29	73.3	3	10.3	13	44.8
General Comp.	172	105	94.2	91	86.7	14	13.3
Paint Mfg. (4)	173	42	78.6	14	33.3	18	42.9
Welding Trade(4)	251	88	87.9	45	51.1	29	33.0
Total:		264	83.5	153	58.0	74	28.0
Comparisons Between Individual Chemists and External Judges							
Paint Mfg. (5)	173	42	77.4	18	42.9	16	38.1
Welding Trade(5)	251	88	77.6	40	45.6	22	25.0
Total:		130	77.5	58	44.6	38	29.2

Table 7.5.4, continued

Trial	Total No. of Substances	No. of Substances Attributed (a)	Avg. p_e %	Number of Substances in which Agreement was (2): Excellent No.	Excellent % of (a)	Good No.	Good % of (a)
Comparisons Between Consensus and External Judges							
Metal Ind.	300	65	82.0	14	21.5	43	66.2
Chemical Mfg.	300	34	79.4	11	32.4	12	35.3
Total:		99	80.7	25	25.3	55	55.6
Reliability of Consensus Coding							
Code/Recode	270	108	95.5	102	94.4	6	5.6
Grand Total:	601		82.9	338	56.2	173	28.8

Table 7.5.4, continued

(1) Agreement was calculated by first categorizing the judgements of each rater into two levels: exposed, at any level of confidence, and not exposed. Averages of p_o are presented over all pairs of raters participating in each trial, except where noted below.

(2) Agreement for each individual substance was categorized as follows:

P_o	
≥90%	Excellent
70 - 90%	Good
<70%	Poor

(3) Excludes the Final Coding.

(4) Excludes the comparison of the chemists with the external judges.

(5) Excludes the comparison between the two chemists.

(8.0) Discussion

(8.1) Validity and Reliability of the Chemists' Coding

The consensus method of coding exposure for metal industry and chemical manufacturing occupations was shown to be valid. A higher degree of concordance was observed, however, in the Metal Industry Trial. This is partly explained by the chemists' greater familiarity with metal industry occupations. The chemists had previously consulted with industrial hygienists and engineers, they had access to published reports of occupational surveys in this industry (DSC, 1984), and they had previously participated in related trials (eg, the Welding Trade Trial). By contrast, the chemists had no previous experience coding exposure for occupations *similar* to those used in the Chemical Manufacturing Trial. They were not able to visit the site or consult company personnel. (The plant from which these job descriptions were generated is unique in Canada.) The chemists had to rely solely on published material as the source of exposure information. Moreover, since the job descriptions for this trial had not been obtained by interviewing subjects, it was felt that the working environment was not as accurately described as in the Metal Industry Trial. Hence, the extent of agreement observed in the chemical trial probably represents a lower limit to the validity of the consensus approach.

The validity of the coding of paint manufacturing and welding occupations by individual chemists can be inferred from the high degree of agreement observed between the chemists and external judges. Since very good agreement was also observed between the individual chemists,

it is likely that a consensus coding by these chemists, if undertaken, would be at least as valid.

Agreement between internal and erternal raters was higher in the Paint Manufacturing and Welding Trade trials than in the Metal Industry Trial. The Cancer Study chemists' greater experience coding paint and welding occupations may be partly responsible for this difference. On the other hand, the external judges in the Metal Industry Trial probably had the most accurate exposure information. In fact, these raters attributed a higher proportion of exposures than in the consensus coding. Thus, it is possible that the lower agreement observed in the metal trial reflected some difficulty on the part of the Cancer Study chemists in identifying exposures. It is unlikely, however, that the metal industry occupations were inherently more complex than those in the other two trials. This should be evident when it is recalled that two of the subjects worked as welders, one as a painter and the remainder worked as polishers and metal formers.

It is likely that the external raters provided the best assessments of exposure that could be obtained under the imposed constraints of time and cost. The external raters in the Metal Industry Trial probably had the best exposure information since environmental measurements for some substances from an occupational survey were used in conjunction with information gleaned from interviews. Even with these accurate measurements of exposure, it was clear that the job description provided additional information not available from the survey. For example, if different materials were used at different times, it is

likely that only those materials used at the time of the survey would be noted. The interviews were able to indicate substances not otherwise catalogued. Overall, it can be concluded that since good agreement was observed in all comparisons with external raters, that the *participating* Cancer Study chemists generally coded exposure as accurately as any other eligible rater.

The results of the Code/Recode trial indicate that the consensus method was quite reliable. The nature of this trial was such that each rater coded the same job descriptions at two different times. However, the author was assured by the participating chemists that they were unaware that they had previously coded these job descriptions. Thus, the high level of agreement observed could not have been due to this effect.

(8.2) Inter-rater Agreement among Cancer Study Chemists

The results of the inter-chemist comparisons, using the Kappa statistic, indicates a high level of agreement between internal raters coding a wide range of occupations. Significant differences in estimates of prevalence were observed, however, between some chemists. (The implication of simultaneously observing significant differences in prevalence between raters and high values of Kappa are discussed in section (8.4).) In particular, it was found that the chemist with the most experience (Chemist 1) attributed a higher proportion of exposures than any other rater. In practice, differences between chemists could be resolved through other related coding activities, such as regular staff meetings.

The coding experience of the chemists in the Cancer Study was different, some worked mainly in primary coding while others worked in secondary coding. The high degree of agreement between the chemists, as measured by the Kappa statistic, suggests that the chemists who participated in these trials can probably work in any facet of the routine coding (eg, primary or secondary coding).

(8.3) Implications to the Cancer Study

(8.3.1) Generalizability to all Job Descriptions on File

Certain factors may affect agreement and, consequently, the accuracy of the chemists' coding. It is likely, for example, that the rater's knowledge of industrial environments, the reliability of his coding, the industry and occupation in which the worker was employed, the time period of the occupation, the location of the job (eg, in the Montreal area or not), and the quality of the job description are some factors that may affect agreement. The trials were not designed to isolate which of these or other factors would affect agreement. Many more job descriptions, sampled in a different way, would be required to answer this type of question. It is possible that the chemists' coding may not be as accurate for certain types of job descriptions. This could imply that the results obtained here should not be extrapolated to the set of job descriptions on file in the Cancer Study. Furthermore, the generalizability of these conclusions to all job descriptions in the Cancer Study may be limited because a small number of job descriptions (about 100) were coded in these trials. In spite of these arguments, there is reason to believe that the results of these trials can indeed

be generalized to those job descriptions that were obtained by *interviewing* subjects. First, good agreement was observed in trials that evaluated specific types of occupations (egs, paint, welding, and rubber). Second, there was good agreement for a large proportion of attributed substances. Finally, and most importantly, the results of the Chemical Manufacturing Trial placed a relatively high lower limit on the accuracy of the coding of exposure for unfamiliar occupations.

(8.3.2) Generalizability to all Chemists

In the Cancer Study, the chemists used a consensus method to assess exposure. Other activities, such as consulting with industrial experts, extensive literature reviews, and site visits tend to increase, with time, the chemists' state of knowledge of industrial environments. Information is kept on file and is shared among the chemists in day-to-day activities and through regular meetings. Thus, the manner in which information is shared within the Cancer Study, the high level of inter-rater agreement between chemists who performed different coding tasks in the Cancer Study, and the belief that the external raters provided excellent assessments of exposure implies that two specific conclusions regarding the participating Cancer Study chemists can probably be generalized to all trained chemists: 1) most trained chemists will attribute exposure as well as any other "expert"; and 2) most trained chemists should be able to perform any coding task within the Cancer Study (ie, primary and secondary coding and special projects).

(8.4) Limitations of the Present Study

The conclusions discussed above may be seriously compromised if the collapsed table method is not an appropriate means of summarizing agreement. In particular, the values obtained for the Kappa statistic may not adequately reflect the degree of agreement averaged over all substances. The value of Kappa obtained from the collapsed table was, however, consistently less than that obtained from an arithmetic average of individual values of Kappa over all substances. This indicates that estimates using the collapsed table were probably conservative.

Furthermore, the qualitative interpretation given to the values of Kappa may not satisfactorily indicate real differences in coding exposure between raters. In particular, the use of Landis and Koch's (1977a) ad hoc system of interpreting values of Kappa may not be appropriate. For example, it is not clear whether a value of Kappa of 0.5, which is considered "good" agreement, reflects a larger difference in the raters' sensitivities and specificities than a value of 0.4, which is considered only fair agreement. In addition, the values of Kappa obtained from the inter-chemist comparisons in the Paint Manufacturing and Welding Trade trials indicate further difficulties in interpretation. In the former trial, the chemists' estimates of prevalence were extremely different (9.1% versus 5.8%). In the latter trial, the values were not as divergent (8.4% versus 9.7%). Yet, the two values of Kappa were very close - 0.62 and 0.69, respectively. This seems to indicate that Kappa may not be sensitive to departures in

estimates of prevalence. The result also implies that the qualitative interpretation conveyed to Kappa may require that differences in estimates of prevalence be taken into account.

One limitation of the present study was that few job descriptions were used in any one trial. As a result, estimates of agreement for individual substances were not precise. In fact, the index of crude agreement was used to indicate the extent of agreement for individual substances since Kappa is highly unstable for small sample sizes. The analysis in which the average index of agreement between raters was calculated as a function of the average prevalence of exposure generally showed that there was higher agreement in the low prevalence categories. This was not unexpected since it is well known (cf, Rogot and Goldberg, 1966) that there is less opportunity to err when there are more ratings in one category; ie, for high or low values of prevalence. It was shown, however, that the average obtained over all substances generally reflected the extent of agreement observed for those substances in the "medium" prevalence categories.

(8.5) Comparison of these Results with other Studies

The results of the present study and those of Baumgarten et al., 1983, in which the occupations reported by subjects in the Cancer Study were shown to be valid, indicate that the determination of exposure in the Cancer Study is accurate. Various other methods have been used to measure past exposure in other occupational settings. The results from these studies, although few in number, lend credibility to the notion

that past exposure to specific agents can be assessed fairly accurately.

It is interesting to compare the results of two studies which assessed the accuracy of the coding of different raters. In the study by Macaluso and his co-workers, 1983, assessments of exposure to polycyclic aromatic hydrocarbons (PAH) and asbestos were compared between a team of chemists and a job/exposure matrix. The average agreement observed was approximately equal (average p_0 = 80% and $\kappa >$ 0.5) to that obtained in the present study. In their study it was also found that there was a larger difference in the two estimates of the odds ratio for PAHs than for the two estimates of the odds ratio for asbestos, even though agreement was *higher* for PAHs (Kappa = 0.58 versus 0.52). Certainly, the differences in odds ratios were due to the different distributions of exposure in the case and control series. It has been shown (Kraemer, 1979, Thompson, 1982, and Walter, 1983) that Kappa is related in a complicated way to the sensitivity and specificity of the two raters and to the prevalence of exposure in the population. (Most other summary measures of agreement will also be functions of these parameters.) Consequently, various combinations of values of these parameters may lead to estimates of Kappa that may not reflect trus differences in risk ratios.

It is instructive to investigate why the study by Soskolne, 1982, reported much lower levels of inter-rater agreement than that obtained here. In his study, job titles containing information on occupation at a chemical plant were used by various raters to attribute exposure to a

set of substances. It may difficult to compare his study with the present one because of the use of different scales of measuring exposure. Moreover, there were large differences in the estimates of prevalence of exposure, which may confound the comparison of the estimates of agreement using the Kappa index (Thompson, 1982, and Walter, 1983). Despite these problems, it is believed that the lov levels of agreement observed by Soskolne may be attributed to a lack of detail in the job title and to an inconsistent usage of the coding criteria.

(8.6) **Further Research**

It has been inferred from the results of the seven agreement trials that the chemists' coding was fairly accurate. It may, however, be worthwhile to investigate the appropriateness of extrapolating the results to all job descriptions on file by investigating the factors that may affect agreement. The results of such a project would certainly be of interest to the epidemiological community. This ambitious project may not be feasible, however, because of the potentially high cost involved.

In all of these trials, the agreement in determining whether an exposure had occurred was higher than that of attributing intensity of exposure. The results of the pneumoconiosis trials (eg, Liddell, 1963) suggest that improvements in inter-rater and intra-rater agreement for determining exposure may occur if finer scales of measurement are used. For example, frequency and concentration could be coded on a ten-point

scale by taking into account whether the rater seriously considered coding the exposure into an adjacent category. It does not seem sufficiently important at this time to justify determining whether agreement would improve with the use of different scales.

Although outside the scope of this thesis, the tables provided in Appendices 7 through 13 may be used to determine whether inter-rater agreement for certain substances was poor. The results of such an analysis may then prove useful in correcting coding deficiencies or in planning future trials.

In the author's opinion, the most important issue is whether the estimates of excess risk deduced from different exposure assessments are large or small. In the context of the Cancer Study, it may be worthwhile to select a few exposures and and at least two sites of cancer to test what kind of differences in risk ratios would be obtained by different raters.

Research into the validity of ascertaining exposure will be important in other epidemiological investigations but, because of the cost involved, may not be assessed through extensive experimental trials. As previously discussed, there are many difficulties in interpreting currently used indices of agreement. One solution may be to correlate indices of agreement with other measures that are more readily interpretable and are epidemiologically important. For example, measuring differences in estimates of risk ratios between different raters is more important than estimating some statistic of agreement.

One potential avenue is the use of numerical simulation. This can be employed to avoid the large cost of having more than two sets of raters judge a complete set of files. Thus, exposure assessments of hypothetical raters can be made by specifying certain parameters (eg, sensitivity, specificity, and prevalence). Risk ratios can be calculated and correlated with values of various indices of agreement (eg, Kappa). The results from such simulations may then provide a more intuitive basis for interpreting agreement indices and may then provide the basis for designing and testing new scales of measurement and, perhaps, other methods of coding exposure.

(8.7) Conclusions

The objectives of this thesis were met by measuring the extent of agreement among participating raters in seven trials. In sum, it was discovered that there was good to excellent agreement in attributing exposure among chemists, and between chemists and external judges. From this, it was inferred that the coding of individual chemists and of chemists using a consensus process was accurate.

It has been previously shown that the reported occupations were fairly accurate (Baumgarten et al., 1983). Thus, it can be tentatively concluded that the methods used to determine exposure in the Cancer Study are fairly accurate.

(9.0) References

Abramson, J H, 1979, "Survey Methods in Community Medicine", second edition, (Churchill Livingstone, London)

Armitage, P, L M Blendis and H C Smyllie, 1966, "The Measurement of Observer Disagreement in the Recording of Signs", J Roy Stat Soc Series A 129: 98-109

Axelson, O, E Dahlgren, C-D Jansson, S O Rehnlund, 1978, "Arsenic Exposure and Mortality: A Case-Referent Study from a Swedish Copper Smelter", Brit J Indust Med 35:8-15

Barron, B A, 1977, "The Effect of Misclassification on the Estimation of Relative Risk", Biometrics 33:414-418

Bartko, J J, 1966, "The Intraclass Correlation Coefficient as a Measure of Reliability", Psychol Rep 19:3-11

Baumgarten, M, J Siemiatycki, G W Gibbs, 1983, "Validity of Work Histories Obtained by Interview for Epidemiologic Purposes", Am J Epi, 118:583-591

Beaumont, J and M Weiss, 1981, "Lung Cancer Among Welders", Journal Occup Med 23:839-844

Bennett, B M, 1967, "Tests of Hypotheses Concerning Matched Samples", J

Roy Stat Soc B 29:468-74

Bennett, B M, 1968, "Note on Chi-Square Tests for Matched Samples", J Roy Stat Soc B 30:368-70

Bennett, B M, 1972, "Measures of Clinicians' Disagreements over Signs", Biometrics 28:607-612

Blot, W, L M Brown, et al., 1983, "Lung Cancer Among Long-term Steel Workers", Am J Epi 117:706-716

Cohen, J, 1960, "A Coefficient of Agreement for Nominal Scales", Educ & Psych Msmt 20:37-46

Cohen, J, 1968, "Weighted Kappa: Nominal Scale Agreement with Provision for Scaled Disagreement or Partial Credit", Psychol Bull 70:213-220

Copeland, K T, H Checkoway, A J McMichael, et al., 1977, "Bias due to Misclassification in the Estimation of Relative Risk", Am J Epi 105:488-495

Copland, L, J Burns, and M Jacobsen, "Classification of Chest Radiographs for Epidemiological Purposes by People not Experienced in the Radiology of Pneumoconiosis", Brit J Indust Med, 38:254-261

Delzell, E and R R Monson, 1981, "Mortality among Rubber Workers: VI. Men with Potential Exposure to Acrylonitrile", Journal Occup Med

24:767-769

Dewar, R, and J Siemiatycki, 1984, "A Program for Point and Interval Calculation of Odds Ratios and Attributable Risks from Unmatched Case-control Data", manuscript

Dice, L R, 1945, "Measures of the Amount of Ecologic Association between Species", Ecology 26:297-302

DSC, 1983, "Projet Pilote en Santé et Sécurité du Travail dans le Secteur de Fabrication de Produits en Metal", Departement de Santé Communataire, Hopital Sacre-Coeur (Montreal), 4 volumes

Efron, B, 1981, "Nonparametric Estimates of Standard Error and The Jackknife, the Bootstrap and Other Methods", Biometrika 68:589-99

Efron, B, 1982, "The Jackknife, the Bootstrap and Other Resampling Plans", CBMS-NSF no 38 (SIAM, Philadelphia)

Felson, B, W K C Morgan, L J Bristol, et al., 1973, "Observations on the Results of Multiple Readings of Chest Films in Coal Miners' Pneumoconiosis", Radiology 109:19-23

Fisher, R A, 1958, "Statistical Methods for Research Workers", (Hafner, New York)

Fleiss, J L, 1965, "Estimating the Accuracy of Dichotomous Judgements",

Psychometrika 30:469-79

Fleiss, J L, 1975, "Measuring Agreement Between Two Judges in the Presence or Absence of a Trait", Biometrics 31:651-659

Fleiss, J L, 1981, "Statistical Methods for Rates and Proportions", 2nd edition (J. Wiley, New York)

Fleiss, J L, and J Cohen, 1973, "The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as a Measure of Reliability", Educat Psychol Meas 33:613-619

Fleiss, J L and M Davies, 1982, "Jackknifing Functions of Multinomial Frequencies, with an Application to a Measure of Concordance", Am J Epi 115:891-845

Fleiss, J L, R L Spitzer, J Endicott and J Cohen, 1972, "Quantification of Agreement in Multiple Psychiatric Diagnosis" Arch Gen Psychiat, 26:168-171

Fletcher, C M and P D Oldham, 1964, "Diagnosis in Group Research", in L J Witts, ed, *Medical Surveys and Clinical Trials*, (Oxford University Press, London)

Gérin, M, L Nadon and M Gauthier, 1983a, "L'Exposition au formaldehyde dans le milieu de travail", internal report to the Institut de recherche en santé et en sécurité du travail du Québec

Gérin, M, J Siemiatycki, L Laroche, C Millet, 1983b, "Translating Job Histories into Histories of Occupational Exposure for Epidemiologic Purpose", in E D Acheson, ed, *Job Exposure Matrices*, (Proceedings of a Symposium, 22 April 1982, Southampton; Medical Research Council, UK)

Gérin, M, J Siemiatycki, H Kemper and D Bégin, 1984, "Obtaining Occupational Exposure Histories in Epidemiologic Case-control Studies", manuscript

Goodman, L A and W H Kruskal, 1954, "Measures of Association for Cross-Classifications", Journal Amer Stat Assn 49:732-764

Greenberg, R A and C H Tamburro, 1981, "Exposure Indices for Epidemiologic Surveillance of Carcinogenic Agents in an Industrial Chemical Environment", Journal Occup Med 23:353-358

Greenland, S, 1980, "The Effect of Misclassification in the Presence of Covariates", Am J Epi 112:564-569

Hardell, L and A. Sandstrom, 1979, "Case-control Study: Soft-tissue Sarcomas and Exposure to Phenoxyacteic Acids or Chlorophenols", Br J Cancer 39:711-717

Hoar, S K, 1983, "Epidemiology and Occupation Classification Systems", in E D Acheson, ed, *Job Exposure Matrices* (Proceedings of a Symposium, 22 April 1982, Southampton; Medical Research Council, UK)

Hoar, S K, A S Morrison, P Cole, D T Silverman, 1980, "An Occupational and Exposure Linkage System for the Study of Occupational Carcinogenesis", Journal Occup Med 22:722-726

Hsieh, C-C, A M Walker, S K Hoar, 1983, "Grouping Occupations According to Carcinogenic Potential: Occupation Cluster from an Exposure Linkage System", Am J Epi 117:575-589

Hubert, L and R G Golledge, 1983, "Rater Agreement for Complex Assessments", Br J Math & Stat Psych 36:207-216

Hull, C H and N H Nie, 1981, "SPSS Update 7-9" (McGraw-Hill, Toronto)

Jarvholm, B, L Lillienberg, et al., 1981, "Cancer Morbidity among Men Exposed to Oil Mist in the Metal Industry", Journal Occup Med 23:333-337

Kendall, M G, 1955, "Rank Correlation Methods", (Hafner Publishing Company, New York)

Kendall, M G and A Stuart, 1961, "The Advanced Theory of Statistics Volume 2", (Hafner Publishing Company, New York)

Kleinbaum, D, L Kupper and H Morgenstern, 1982, "Epidemiologic Research" (Lifetime Learning Publications, Belmont, Calif)

Koran, L M, 1975a, "The Reliability of Clinical Methods, Data, and

Judgements, Part 1", New Eng Journal Med 293:642-646

Koran, L M, 1975b, "The Reliability of Clinical Methods, Data, and Judgements, Part 2", New Eng Journal Med 293:695-701

Kraemer, H C, 1979, "Ramifications of a Population Model for Kappa as a Coefficient of Reliability", Psychometrika 44:461-472

Kraemer, H C, 1980, "Extension of the Kappa Coefficient", Biometrics 36:206-16

Kramer, M S and A R Feinstein, 1981, "Clinical Biostatistics LIV. The Biostatistics of Concordance", Clin Pharm and Therap 29:111-123

Krippendorf, K, 1970, "Bivariate Agreement Coefficients for Reliability of Data", in E F Borgatta, ed, *Social Methodolgy* (Jossey-Bass, San Francisco)

Landis, J R and G G Koch, 1975, "A Review of Statistical Methods in the Analysis Arising from Observational Reliability Studies", Parts 1 and 2, Statistica Neerlandica 29:101-123, 151-161

Landis, J R and G G Koch, 1977a, "The Measurement of Observer Agreement for Categorical Data", Biometrics 33:159-74

Landis, J R and G G Koch, 1977b, "A One-way Components of Variance Model for Categorical Data", Biometrics 33:671-679

Landis, J R and G G Koch, 1977c, "An Application of Hierarchial Kappa-type Statistics in the Assessment of Majority Agreement Among Multiple Observers", Biometrics 33:363-374

Larsen, R J and M L Marx, 1981, "An Introduction to Mathematical Statistics and its Applications" (Prentice-Hall, Englewood Cliffs, N J)

Last, J, 1983, "Dictionary of Epidemiology" (Oxford University Press, Toronto)

Liddell, F D K, 1961, "The Effect of Film Quality on Reading Radiographs of Simple Pneumoconiosis in a Trial of X-Ray Sets" Brit J Industr Med 18:165-174

Liddell, F D K, 1963, "An Experiment in Film Reading", Brit J Industr Med 20:300-312

Liddell, F D K, 1972, "Validation of Classifications of Pneumoconiosis" Annals of N Y Acad of Science 200:527-551

Liddell, F D K, 1974, "Assessment of Radiological Progression of Simple Pneumoconiosis in Individual Miners", Brit J Industr Med 31:185-195

Liddell, F D K, 1977, "Radiological Assessment of Small Pneumoconiotic Opacities", Brit J Industr Med 34:85-94

Liddell, F D K, 1980, "Validation of the UICC/Cincinnati Classification

Radiographs in Terms of Prediction of Mortality of Asbestos Workers", in J C Wagner, ed, *Biological Effects of Mineral Fibres, vol 2* (IARC Scientific Publication no. 30, Lyons, France)

Liddell, F D K and D C Lindars, 1969, "An Elaboration of the ILO Classification of Simple Pneumoconiosis", Brit J Indust Med 26:89-100

Lubin, J, L M Pottern et al., 1981, "Respiratory Cancer among Copper Smelter Workers: Recent Mortality Statistics", Journal Occup Med 23:779-784

Macaluso, M, P Vineis, et al., 1983, "Job Exposure Matrices: Experience in Italy", in E D Acheson, ed, *Job Exposure Matrices* (Proceedings of a Symposium, 22 April 1982, Southampton; Medical Research Council, UK)

Mancuso, T A, A Stewart and G Kneale, 1977, "Radiation Exposures of Hanford Workers Dying from Cancer and other Causes", Health Physics 33:369-385

Maxwell, A E and A E G Pilliner, 1968, "Deriving Coefficients of Reliability and Agreement for Ratings", Brit J Math Stat Psychol 21:105-116

Miller, R G, 1974, "The Jackknife - A Review", Biometrika 61:1-15

Musch, D C, J R Landis, et al., 1984, "An Application of Kappa-Type Analyses to Interobserver Variation in Classifying Chest Radiographs

for Pneumoconiosis", Statistics in Medicine 3:73-83

Nie, N H, C H Hull, et al., 1974, "Statistical Package for the Social Sciences", second edition (McGraw Hill, Toronto)

OCC, 1971, "Canadian Classification and Dictionary of Occupations" (Department of Manpower and Immigration, Ottawa)

Pershagen, G and O Axelson, 1982, "A Validation of Questionnaire Information on Occupational Exposures and Smoking", Scand J Work Environ Health 8:24-28

Pershagen, G., 1978, "Lung Cancer Mortality, Occupational Exposure and Smoking Habits in a Region Surrounding a Smelter", in *International Symposium on the Control of Air Pollution in the Working Environment* (International Labour Organization, Geneva), p179-185

Price, P, 1982, "An Annotated Bibliography of Sources of Information on Occupational Exposures", unpublished

Rogot, E and I D Goldberg, 1966, "A Proposed Index of Measuring Agreement in Test-retest Studies", J Chr Dis 19:991-1006

Rossiter, C E, 1972, "Initial Repeatability of the UICC/Cincinnati Classification of the Radiographic Appearances of Pneumoconiosis", Brit J Industr Med 29:407-419

Scott, W A, 1955, "Reliability of Content Analysis: The Case of Nominal Scale Coding", Pub Opinion Quart 19:321-325

Siegal, S, 1956, "Nonparametric Statistics for the Behaviour Sciences" (McGraw-Hill, Toronto)

Siemiatycki, J, 1979, "Protocol: Monitoring the Occupational Environment for Carcinogens: A Pilot Study in Montreal", manuscript

Siemiatycki, J, M E Day, J Fabry, J A Cooper, 1981a, "Discovering Carcinogens in the Occupational Environment: A Novel Approach", JNCI 66:217-225

Siemiatycki, J, M Gérin, J Hubert, 1981b, "Exposure-Based Case-Control Approach to discovering Occupational Carcinogens: Preliminary Findings", in R Peto and M Schneiderman, *Quantification of Occupational Cancer*, Banbury Report, no 9, pp 471-483 (Cold Spring Harbour Laboratory)

Siemiatycki, J, M Gérin, L Richardson, J Hubert, and H Kemper, 1982, "Preliminary Report of an Exposure-Based, Case-control Monitoring System for Discovery Occupational Carcinogens", Terat, Carc & Mutag 2:169-177

Siemiatycki, J, M Gérin, R Lakhani, R Dewar, J Pellerin and L Richardson, 1983a, "Nickel and Cancer Associations from a Multicancer Occupation Exposure Case-Referent Study: Preliminary Findings",

manuscript

Siemiatycki, J, M Gérin, L Nadon, L Richardson, M Goldberg, 1983b, "Exposition au Formaldehyde et Risques de Cancer: Rapport Preliminaire", internal report to the Institut de recherch en santé et en sécurité du travail du Québec

SIC, 1970, "Standard Industrial Classification Manual" (Catalogue no. 12-501E, Statistics Canada, Ottawa)

Soskolne, C, 1982, "Upper Respiratory Cancer Among Refinery and Chemical Plant Workers: A Case-control Study in Baton Rouge, Louisiana", PhD Dissertation, University of Pennsylvania

Spitzer, R L, J Cohen, et al., 1967, "Quantification of Agreement in Psychiatric Diagnosis", Arch Gen Psychiat 17:83-87

Spitzer, R L, J L Fleiss, 1967, "A Re-analysis of the Reliability of Psychiatric Diagnosis", Brit J Psychiat 125:341-347

Thomas, D C, W O Spitzer, J K MacFarlane, 1981, "Inter-Observer Error Among Surgeons and Nurses in Presymptomatic Detection of Breast Cancer", J Chr Dis 34:617-626

Thomas, D C, J Siemiatycki, R Dewar, J Robins, M Goldberg, and B G Armstrong, 1984a, "The Problem of Multiple Inference in Studies Designed to Generate Hypotheses", manuscript

Thomas, D C, J Siemiatycki, M Goldberg, and R Dewar, 1984b, "Methods for Relating Several Exposure Factors to Several Diseases in Case-Heterogeneity Studies", manuscript

Thompson, W D, 1982, "Design Issues in the Assessment and Control of Misclassification Errors", presented at the 16th Annual Meeting of the Society for Epidemiologic Research (Cincinatti)

Wolfson, C, 1978, "Measures of Agreement for Qualitative Data", unpublished MSc thesis, Department of Mathematics, McGill University, Montreal

Walter, S D, 1983, "The Maximum Value of Kappa", presented at the 17th Annual Meeting of the Society for Epidemiologic Research (Winnipeg)

Appendix 1

Glossary of Terms and Abbreviations

ANOVA:

Analysis of Variance.

Cancer Study:

An abbreviation for the study "Monitoring the Occupational Environmental for Carcinogens".

Chance Agreement:

The agreement that would be expected by chance if the raters were judging the trait independently.

Chance-corrected Indices:

Summary measures of agreement which are evaluated by taking into account the agreement that is expected by chance if the raters assess the trait independently.

Chemist:

An internal rater employed in the Cancer Study who assessed occupational exposure from detailed job descriptions obtained usually from interviewing subjects.

Chemist Team:

The team of Cancer Study chemists who assessed occupational exposure, by a consensus method, after revieving job descriptions.

Collapsed Table Method:

A statistical method used to assess inter-rater agreement for

attributing exposure to a set of predefined substances. For two raters, one contingency table is obtained by crosstabulating the assessments of exposure for each substance under consideration. These tables are then combined by summing the corresponding cells of each table. Statistics can then be calculated from this "collapsed table" to provide a summary measure of inter-rater agreement.

Concentration:

The variable used to indicate, on a relative scale, the average, weekly level dose to each substance listed on the exposure coding form. Concentration is coded on a four-point scale: no, "low", "medium", and "high" exposure.

Concordant Cells (of Agreement):

The agreement between two observers judging a trait on a scale having K categories can be represented by a KxK contingency table. The values in each of the K cells on the diagonal represent the number of observations for which there is perfect agreement. These are known as the concordant cells of agreement.

Consensus Coding:

The process whereby the team of Cancer Study chemists reviews job descriptions and then attributes exposure to substances listed on the exposure coding form. At least two chemists score each job description; the first chemist performs the preliminary coding and the second verifies and modifies the work of the first. Other chemists may also be called-on to offer their expertise.

Contact:

The variable used to indicate whether exposure to a substance was absent or was either cutaneous, respiratory, or both.

d:

The variable representing the number of raters who participated in an agreement trial.

df:

Degrees of freedom.

DSC:

Departement de Sante Communautaire or Community Health Department.

Exposure Checklist:

See Exposure Coding Form

Exposure Coding Form:

The list of chemicals, compounds and physical agents that were under consideration in the Cancer Study.

Exposure Profile:

The type and extent of exposure that a subject may have incurred in any one job. It is defined by the set of values attributed to each of the four exposure variables (contact, frequency, concentration, and level of confidence) for each substance on the exposure coding sheet.

Exposure Variables:

The variables that indicate the level of exposure ("frequency", "concentration"), the type of exposure ("contact"), and the trustworthiness of the ascertainment of exposure ("level of confidence") as determined by the rater who interpreted the job description.

External Judge:

Any rater, other than a chemist, who participated in the agreement trials reported in this thesis. These raters possessed expert knowledge of the uses of materials and processes used in certain industries and occupations.

Final Coding:

In the Rubber Industry Trial, a fourth assessment of exposure was derived after a meeting of the three participating chemists. The purpose of the meeting was to form a consensus with regard to assigning exposure to rubber industry occupations. This Final Coding, which was evaluated by one of the chemists, was then used in the analysis of inter-rater agreement.

Frequency:

The variable used to indicate, in each occupation, the average amount of time of exposure to each substance on the chemical coding sheet. Frequency is coded on a four-point ordinal scale; viz: no exposure, less than 5%, 5-30%, and greater than 30% of the time.

Independence:

This term is used in two different ways: a) it refers to the probability that one event will occur regardless of the outcome of another event (Larsen & Marx, 1981). Therefore, the ratings of two observers will be statistically independent if they use entirely different criteria to judge a trait. If this is true, then any observed agreement would be explainable by chance alone. b) Independence may also refer to the situation in which the raters are unaware of each other's assignments. This situation will occur as a result of study design and is sometimes referred to as "design independence".

Index of Crude Agreement:

This index, which is also known as the overall proportion of agreement, is defined as the sum of the concordant cells in a KxK contingency table. It is denoted by the symbol p_o.

Job Description:

For each occupation of a subject enrolled in the Cancer Study, relevant information is collected for the purpose of ascertaining exposure to certain occupational agents. A description of the tasks, machines and materials used, and the general working environment are obtained usually by interviewing subjects. The job description represents, for the trials reported in this thesis, the fundamental unit of observation.

Job Exposure Matrix:

A device that is used to infer the extent of exposure in occupational environments from a set of predefined substances. The set of job titles and substances under consideration are specified in advance. Each job title on file will be accorded exposure levels to a subset of these substances using information obtained from the literature and other sources. The exposure levels may depend on the time of the occupation or on other factors. An exposure profile for an individual can then be obtained by specifying the job title, the time period of employment and other relevant variables. The linkage is usually performed automatically by computer.

Joint Distribution (of Rater Assessment):

The distribution of the probability of the occurrence of an event that is simultaneously dependent on two or more variables, each having their own distribution (Larsen and Marx, 1981). For example, two raters who are independently scoring a categorical variable having K categories will generally produce different frequency distributions (i.e., different marginal distributions). The joint distribution of rater ascertainment will result when the two scores are cross-tabulated. The new random variable that is created by this process will have KxK values. The distribution of this new variable is simply the joint distribution of rater assessment.

κ':

The ratio of Kappa to κ_{max}.

Kappa (κ):

A statistic that is used to measure agreement for raters who judge a trait on a categorical scale. It is corrected for chance agreement by subtracting the agreement that would be expected if the raters were judging the trait independently. It is then standardized by dividing by the maximum excess agreement.

Kappa Maximum:

The maximum value of Kappa that would result given the observed marginal proportions. The symbol (κ_{max}) is used to refer to this quantity.

Level of Confidence:

The variable used to indicate the degree of trustworthiness of exposure. Four values are used by the raters: no exposure, "possible" exposure, "probable" exposure and "almost certain" exposure.

Marginal Proportions:

The frequency of occurrence of the categories of any variable used to describe a trait in a population of subjects. Suppose, for example, that a variable, which is composed of four categories, is measured by two raters. The marginal proportions are simply the two frequency distributions for each of the two raters. If a 4x4 table is used to represent the cross-classification of the ratings of the 2 observers, then the frequency distribution obtained above will be identical to that displayed in the margins of the table.

MMAI:

Mean Majority Agreement Index.

MPDI:

Mean Pair Disagreement Index.

n:

The number of job descriptions used in any agreement trial.

n_{11}:

The number of exposures jointly attributed present by two raters.

Overall Proportion of Agreement:

A synonym for the Index of Crude Agreement.

P_0:

This symbol represents the Index of Crude Agreement.

Pattern of Concordance:

The distribution of the number of exposures classified according to the number of raters who attribute the exposure present and absent. If there are three raters attributing exposure then there will be 4 categories of interest. For example, there will be one category where two of the three raters attribute exposure while the third rater codes it absent.

Prevalence of Exposure:

The frequency of exposure obtained from assessing exposure to a set of

job descriptions. It refers to the frequency of exposure to any of the substances on the exposure coding form. It also refers to the total frequency of exposure with respect to all substances and job descriptions. In the former instance the denominator is the number of job descriptions whereas in the latter case the denominator is the number of substances multiplied by the number of job descriptions.

Primary Coder:

The process of consensus coding, as performed by the team of chemists, consists of at least two chemists sequentially reviewing each job description. The chemist who first reviews the job description and assigns exposure is known as the primary coder. The chemist, or chemists, who subsequently review the file are known as *Secondary Coders*.

Q:

Cochran's Q index. It is used for testing the equality of the marginal proportions in situations where there are more than two raters.

QI: This refers to the questionnaire/interview which is used to elicit complete occupational and personal history for subjects enrolled in the Cancer Study. It is composed of two parts: the SAQ1 which is used to elicit a lifetime work history and the interview which is used to obtain information on personal variables and details for each occupation.

s:

The number of substances on the chemical coding form.

s':

The number of substances on the exposure checklist that were assigned, in any agreement trial, a prevalence of exposure greater than zero, as estimated by any of the raters.

Sampling Unit:

The constituent members of a population that are being sampled. The job description is the sampling unit for the trials reported in this thesis.

SAQ:

An acronym for self-administered questionnaire. SAQ1 refers to the questionnaire used to elicit lifetime work histories from subjects.

SDAI:

Standard Deviation Agreement Index.

Secondary Coder:

The process of consensus coding, as performed by the team of chemists, consists of at least two chemists sequentially reviewing each job description. The chemist who first reviews each job description and assigns exposure is known as the Primary Coder. The chemist, or chemists, who subsequently review the file are known as Secondary Coders.

Synthetic Index:

A composite index of exposure defined by multiplying the values of the exposure variables frequency, concentration and level of confidence.

Uncorrected Indices:

Any measure of agreement that does not account for agreement that would be expected by chance.

Unit of Observation:

See Sampling Unit.

Work History:

The set of job descriptions that comprise the lifetime, occupational history of a subject.

95% CI:

Ninety-five percent confidence interval for Kappa derived using an asymptotic estimate of variance.

Appendix 2

Some Questionnaires Used in the Cancer Study to Elicit Occupational Exposure from Subjects

Work history I.D. no.: _____

Job. No. _____ Name: _____ Surname: _____

W 1. You worked at _____

from 19 ___ to 19 ___ as a _____

Was this full-time or part-time? _____

Was there shiftwork? _____ If yes, what hours? _____

Company

W 2. Main activity of company or organization? _____

_____

W 2. a) PROBE for other relevant activities or products _____

_____

W 2 b) IF INDUSTRIAL, PROBE for production process used _____

_____

Job

W 3. Now I would like to ask some questions about your job and about the type of environment where you worked. How would you describe the place where you usually worked?

factory or plant	construction site	office
laboratory	warehouse	restaurant or hotel
vehicle	outdoors	others? _____

W 4. In what department of the company or organization did you work?

_____

_____

W 5. JOB DESCRIPTION

I would like you to describe in detail your specific tasks. Try to describe what you did and how you did it. We are particularly interested in any materials that you manipulated or machines that you used.

_____

_____

_____

_____

_____

_____

_____

(OPERATION AND MAINTENANCE OF MACHINES, VEHICLES; LOADING, UNLOADING CONTAINERS; CLEANING; SUBSTANCES OR CHEMICALS USED - PROBE FOR FUNCTIONS AND NAMES)

W 6. Did you ever have to replace someone else? _____

IF YES, how often and what were your tasks? _____

_____

W 7. What other work was being done around you by other workers?

_____

_____

W 8. Can you describe the (room, office, workshop) where you worked?

_____

_____

(SIZE, NUMBER OF PEOPLE, NUMBER OF MACHINES, TEMPERATURE).

W 9. Was there dust or smoke or fumes or gases in the environment where you worked?

_____

IF YES, PROBE FOR DESCRIPTION, SOURCE, NAMES (CIGARETTE SMOKE)

_____

_____

_____

_____

W 10. Did you work with: oils or solvents or acids or detergents? _____

IF YES, PROBE FOR FUNCTIONS, NAMES. _____

_____

_____

_____

W 11. Did your job involve exposure to radiation? (X-rays, micro waves...) _____

IF YES, PROBE FOR FUNCTION. _____

_____

W 12. Did you have to wear any protective equipment while at work? _____

IF YES, WHAT AND WHY? _____

_____

W 13. Did this job have a bad effect on your physical health? _____

IF YES, WHY? _____

_____

W 14. Was there anything you did not like to do? _____

_____

_____

W 15 How stressful was the job? Did this job make you feel tense or anxious most of the time? _____

IF YES, WHY? _____

_____

PAINTERS

A) In what type of industry did you work as a painter?

1. construction - private houses
   - large buildings, bridges, other metal structures
2. assembly or repair of cars, trucks, vehicles
3. ship building or repair
4. household appliances: fridges, stoves...
5. furniture, office equipment...
6. other: _____

B) To what type of surface did you apply the paint?

1. wood
2. metal
3. exterior masonry
4. walls, ceilings
5. other: _____ _____

C) What type of paint did you use?

1. exterior: oil
   latex
   metal paint
2. interior: oil
   latex
   enamel (washrooms)
3. special paints: varnishes
   primers, undercoats...what type: _____
   wood stains
4. What colours were used:
   always white
   other colours: specify _____

D) What did you apply the paint with?

1. brush
2. roller
3. gun

E) Did you have to clean or strip surfaces before applying the paint?

1. sanding
2. degreasing (solvent)
3. stripping (to remove an old layer of paint or varnish):
   caustic solution
   special solvent

F) What products did you use to clean your brushes, rollers, guns?

1. solvent like: varsol, naphta
2. turpentine
3. other: _____

G) What products did you use to clean your hands, face...?

1. solvent like: varsol, naphta
2. turpentine
3. other: _____

H) Did you wear protective equipment during your work?

1. gloves
2. mask
3. other: _____

MACHINIST

A. Type of machine tools

What kind of machine tool did you use?

Lathe, drill, saw, milling machine, boring machine, planin machine, broaching machine, truing machine, grinding wheel sharpener, screw or thread-cutter.

What kind of knife were you using?

B. Cutting oils

Did you ever use cutting oils?

What type: Straight cutting oils?
Soluble cutting oils (Emulsion type)?
Synthetic cutting oils?

Trade names:
Colour and appearance:
How much of your working time did you use them?

C. Lubrification oils and greases

Did you ever use lubrifiers?

Trade names.
Did you do the maintenance on the machines?

D. Metals

What kind of metals were you using?

Stainless steel
Mild steel
Copper
Brass
Bronze
Cast Iron
Special alloys (please specify)
Aluminum
Others

Please specify alloy numbers if possible:
What pieces were you machining?

MACHINIST (page 2)

E. Solvents

What were you using to clean your hands?

- Mineral spirits (varsol)
- Naphta
- Kerosene
- Chlorothene

How many times a day?
What were you using to degrease metal parts?
- Trichloroethylene
- Carbon tetrachloride
- Mineral spirits
- Trade names, ex. chlorothene, royalene

Describe the procedure.
Was the liquid heated?
Did you ever use acids, alkalis (caustic), or any other cleaner?

F. Welding

Did you do any welding?

What kind? Arc, gas, etc.
If welding was done near you, how far away?
(If more than 30% of his working time, fill out welder questionnaire)

G. Grinding

Did you do any grinding?

- What kind of wheel? Wheel number if possible?
- Did you use oils?
- Did you do any wheel straightening?
- What kind of abrasives?
- If grinding was done near you, how far away?

H. Heating

What kind of heating was used in the plant?

- Coke
- Coal
- Salamander

Appendix 3

Exposure Coding Form

Exposure	Code	Exposition	Cont.	Conc.	Freq.	Fiab.	Remarques
(1) SOLIDS (1A) INORGANIC SOLIDS		(1) SOLIDES (1A) SOLIDES INORGANIQUES					
abrasives dust	110001	poussières d'abrasifs					
insulation material dust (inorg.)	110002	poussières d'isolants (inorg.)					
construction site dust	110003	poussières de chantiers de construction					
excavation dust	110004	poussières d'excavation					
metallic dust	110005	poussières métalliques					
inorganic pigments	110024	pigments inorganiques					
extenders	110026	matières de charge					
cement (Portland)	110010	ciment (Portland)					
concrete	110016	béton					
brick	110014	briques					
refractory brick	110017	briques réfractaires					
clay	110015	argile					
ashes	110029	cendres					
silica	110009	silice					
alumina	111301	alumine					
silicon carbide	111401	carbure de silicium					
asbestos (chrysotile)	110007	amiante (chrysotile)					
asbestos (amphiboles)	110008	amiante (amphiboles)					
talc	110013	talc					
glass fibers	110012	fibres de verre					
glass dust	110011	poussières de verre					
mineral wool fibers	110025	fibres de laine minérale					
sulfur	111600	soufre					
sodium carbonate	111101	carbonate de sodium					
sodium silicate	111110	silicate de sodium					
Mg carbonate	111202	carbonate de Mg					
calcium oxide (lime)	112001	oxyde de calcium (chaux)					
calcium sulfate (gypsum)	112002	sulfate de calcium (gypse)					
calcium carbide	112003	carbure de calcium					

Exposure	Code	Exposition	Cont.	Conc.	Freq.	Fiab.	Remarques
calcium carbonate	112005	carbonate de calcium					
titanium dioxide	112201	bioxyde de titane					
iron oxides	112601	oxydes de fer					
zinc oxide	113001	oxyde de zinc					
lead oxides	118201	oxydes de plomb					
lead chromate	118205	chromate de plomb					
basic lead carbonate	118204	carbonate basique de plomb					
Cr dust	112400	poussière de Cr					
Fe dust	112600	poussière de Fe					
Ni dust	112800	poussière de Ni					
Cu dust	112900	poussière de Cu					
Zn dust	113000	poussière de Zn					
Cd dust	114800	poussière de Cd					
Sn dust	115000	poussière de Sn					
Pb dust	118200	poussière de Pb					
bronze dust	110018	poussière de bronze					
brass dust	110019	poussière de laiton					
stainless steel dust	110020	poussière d'acier inoxydable					
mild steel dust	110021	poussière d'acier doux					
solder alloy dust	110022	poussière d'alliage à soudure					
aluminum alloy dust	110028	poussière d'alliage d'aluminium					
(1B) ORGANIC SOLIDS		(1B) SOLIDES ORGANIQUES					
fabrics dust	160001	poussières de tissus					
synthetic fibers	150001	fibres synthétiques					
plastic dusts	150002	poussières de plastiques					
rubber dust	170006	poussière de caoutchouc					
organic dyes and pigments	130001	teintures et pigments organiques					
coal dust	170001	poussière de charbon					
carbon black	170002	noir de carbone					
soot	170004	suie					

Exposure	Code	Exposition	Cont.	Conc.	Freq.	Fiab.	Remarques
coke dust	170005	poussière de coke					
graphite dust	170008	poussière de graphite					
charcoal dust	170009	poussière de charbon de bois					
cellulose (paper fiber)	170003	cellulose (fibre de papier)					
cotton dust	140001	poussière de coton					
wool fibers	140002	fibres de laine					
silk fibers	140003	fibres de soie					
flax fibers	140008	fibres de lin					
rayon fibers	150007	fibres de rayonne					
acrylic fibers	150008	fibres acryliques					
polyester fibers	150009	fibres de polyester					
nylon fibers	150010	fibres de nylon					
acetate fibers	150011	fibres d'acétate					
wood dust	140004	poussière de bois					
cork dust	140009	poussière de liège					
grain dust	140005	poussière de grains					
tobacco dust	140017	poussière de tabac					
flour dust	140006	poussière de farine					
starch dust	140012	poussière d'amidon					
sugar dust	140013	poussière de sucre					
fur dust	140007	poussière de fourrure					
hair dust	140010	poussière de cheveux					
felt dust	140015	poussière de feutre					
leather dust	140016	poussière de cuir					
urea	125002	urée					
tannic acid	145001	acide tannique					
rosin	140014	colophane					
cellulose acetate	150012	acétate de cellulose					
cellulose nitrate	150013	nitrate de cellulose					
polyethylene	150014	polyéthylène					
polypropylene	150015	polypropylène					
polystyrene	150016	polystyrène					
polyvinylchloride	150017	chlorure de polyvinyle					

Exposure	Code	Exposition	Cont.	Conc.	Freq.	Fiab.	Remarques
polyvinylacetate	150018	acétate de polyvinyle					
polyamides	150019	polyamides					
poly-acrylates	150020	poly-acrylates					
acrylonitrile-butadiene-styrene	150021	acrylonitrile-butadiène -styrène					
alkyds	150022	alkydes					
epoxies	150023	époxy					
phenol-formaldehyde	150024	phénol-formaldéhyde					
urea-formaldehyde	150025	urée-formaldéhyde					
melamine-formaldehyde	150026	mélamine-formaldéhyde					
polyurethanes	150027	polyuréthanes					
polyesters	150028	polyesters					
styrene-butadiene rubber	150029	caoutchouc SBR					
polychloroprene	150030	polychloroprène					
polybutadiene	150031	polybutadiène					
natural rubber	140018	caoutchouc naturel					
(2) GASES (2A) INORGANIC GASES		(2) GAZ (2A) GAZ INORGANIQUES					
coal gas	270001	gaz de houille					
hydrogen	210100	hydrogène					
carbon monoxide	210601	monoxyde de carbone					
hydrogen cyanide	210602	cyanure d'hydrogène					
ammonia	210701	ammoniac					
nitrogen oxides (NO_x)	210703	oxydes d'azote (NO_x)					
ozone	210801	ozone					
sulfur dioxide	211601	bioxyde de soufre					
hydrogen sulfide	211602	sulfure d'hydrogène					
chlorine	211700	chlore					
hydrogen fluoride	210901	fluorure d'hydrogène					
hydrogen chloride	211701	chlorure d'hydrogène					

Exposure	Code	Exposition	Cont.	Conc.	Freq	Fiab.	Remarques
(2B) ORGANIC GASES		(2B) GAZ ORGANIQUES					
anesthetic gases	260001	gaz anesthésiants					
spray gases	260002	gaz propulseurs					
natural gas	220001	gaz naturel					
methane	220101	méthane					
propane	220102	propane					
formaldehyde	220501	formaldéhyde					
ethylene oxide	220603	oxyde d'éthylène					
methyl chloride	220805	chlorométhane					
ethylene	221101	éthylène					
acetylene	221102	acétylène					
vinyl chloride	221301	chlorure de vinyle					
phosgene	225001	phosgène					
(3) FUMES AND SMOKES		(3) FUMÉES					
cooking	370002	cuisson des aliments					
plastics pyrolysis	370013	pyrolyse de plastiques					
rubber pyrolysis	370014	pyrolyse de caoutchouc					
engine emissions	370004	gaz d'échappement					
engine emissions (diesel only)	370006	gaz d'échappement (diesel seulement)					
engine emissions (jet fuel)	370011	gaz d'échappement (jet fuel)					
engine emissions (propane)	370012	gaz d'échappement (propane)					
combustion products of coal/coke	370005	prod. combustion charbon/coke					
combustion prod. of liquid fuel	370008	prod. combustion combustible liquide					
combustion prod. of wood	370009	prod. combustion du bois					
combustion prod. of natural gaz	370010	prod. combustion gaz naturel					
pyrolysis and combustion (other organic products	370001	pyrolyse et combustion (autres produits organiques)					

Exposure	Code	Exposition	Cont.	Conc.	Freq.	Fiab.	Remarques
gas welding fumes	310001	fumees de soudage au gaz					
arc welding fumes	310002	fumées de soudage à l'arc					
soldering fumes	310003	fumees de brazage tendre					
metal oxide fumes	310004	fumées d'oxydes métalliques					
Mg fumes	311201	fumées de Mg					
Al fumes	311301	fumées de Al					
Cr fumes	312401	fumées de Cr					
Mn fumes	312501	fumées de Mn					
Fe fumes	312601	fumées de Fe					
Ni fumes	312801	fumées de Ni					
Cu fumes	312901	fumées de Cu					
Zn fumes	313001	fumées de Zn					
Ag fumes	314701	fumées de Ag					
Cd fumes	314801	fumées de Cd					
Sn fumes	315001	fumées de Sn					
Au fumes	317901	fumées de Au					
Pb fumes	318201	fumées de Pb					
(4) LIQUIDS AND VAPOURS (4A) INORGANIC LIQUIDS AND VAPOURS		(4) LIQUIDES ET VAPEURS (4A) LIQUIDES ET VAPEURS INORGANIQUES					
plating solution	410005	solution de placage					
inorganic acid solution	410001	acides inorganiques en solution					
alkali, caustic solution	410002	alcalis, caustiques en solution					
javex	410003	eau de javel					
nitric acid	410704	acide nitrique					
hydrogen peroxide	410802	peroxyde d'hydrogène					
phosphoric acid	411501	acide phosphorique					
sulfuric acid	411603	acide sulfurique					
mercury	418000	mercure					

Exposure	Code	Exposition	Cont.	Conc	Freq.	Fiab	Remarques
(4B) ORGANIC LIQUIDS AND VAPOURS		(4B) LIQUIDES ET VAPEURS ORGANIQUES					
wood varnishes and stains	470002	vernis et teintures à bois					
metal coatings	470005	revêtements à métaux					
paints, varnishes, lacquers (others)	470001	peintures, vernis laques (autres)					
inks	470003	encres					
adhesives (synth.)	460002	adhésifs (synth.)					
animal, vegetal glues	440001	colles animales et végétales					
solvents	460003	solvants					
waxes, polishes	460004	cires, polis					
lubricating oils and greases	460012	huiles et graisses lubrifiantes					
cutting fluids (metal)	460013	huiles de coupe (metaux)					
hydraulic fluid	460019	fluide hydraulique					
mineral oil (other)	460022	huile minérale (autre)					
gasoline (leaded)	460005	essence (au plomb)					
avgas	460026	avgas					
kerosene	460006	kérosène					
jet fuel	460025	jet fuel					
diesel oil	460007	diesel					
heating oil	460008	huile à chauffage, mazout					
petroleum ether	460028	éther de pétrole					
mineral spirits	460009	essence minérale					
mineral spirits with BTX	460027	essence minérale avec BTX					
crude oil	460011	pétrole brut					
asphalt	460014	asphalte					
coal tar and pitch	460015	goudron et poix					
creosote	460017	créosote					
turpentine	440002	térébenthine					
linseed oil	440003	huile de lin					

Exposure	Code	Exposition	Cont.	Conc.	Freq.	Fiab.	Remarques
paraffin	420001	paraffine					
silicone (oil, grease)	420002	silicone (huile, graisse)					
PCB oil	460029	huile BPC					
methanol	420201	méthanol					
ethanol	420202	éthanol					
ethyleneglycol	420203	éthylèneglycol					
isopropanol	420204	isopropanol					
formic acid	420403	acide formique					
acetic acid	420401	acide acétique					
acetone	421501	acétone					
diethylether	420602	diethyléther					
carbon tetrachloride	420801	tétrachlorure de carbone					
chloroform	420802	chloroforme					
methylene chloride	420803	chlorure de méthylène					
1.1.1-trichlorethane	420804	1.1.1-trichloréthane					
carbon disulfide	421001	disulfure de carbone					
trichloroethylene	421302	trichloréthylène					
perchloroethylene	421303	perchloréthylène					
methyl methacrylate	421404	méthacrylate de methyle					
benzene	430101	benzène					
toluene	430102	toluène					
xylene	430103	xylène					
styrene	430104	styrène					
phenol	430201	phénol					
trinitrotoluene	430701	trinitrotoluene					
camphor	440501	camphre					
(5) CHEMICAL FAMILIES		(5) FAMILLES CHIMIQUES					
Be compounds	510499	composés du Be					
Al compounds	511399	composés de Al					
Cr compounds	512499	composés du Cr					
Mn compounds	512599	composés du Mn					
Fe compounds	512699	composés du Fe					

Exposure	Code	Exposition	Cont	Conc.	Freq.	Fiab.	Remarques
Co compounds	512799	composes du Co					
Ni compounds	512899	composes du Ni					
Cu compounds	512999	composes du Cu					
Zn compounds	513099	composés du Zn					
As compounds	513399	composes de As					
Se compounds	513499	composés du Se					
Ag compounds	514799	composés de Ag					
Cd compounds	514899	composés du Cd					
Sn compounds	515099	composés du Sn					
Te compounds	515299	composés du Te					
Sb compounds	515199	composés du Sb					
Au compounds	517999	composés de Au					
Hg compounds	518099	composés du Hg					
Pb compounds	518299	composes du Pb					
cyanides	510001	cyanures					
fluorides	510002	fluorures					
chromates	510003	chromates					
nitrates	510005	nitrates					
hypochlorites	510004	hypochlorites					
alkanes (C_1-C_4)	520198	alcanes (C_1-C_4)					
alkanes (C_4-C_{17})	520199	alcanes (C_4-C_{17})					
alkanes (C_{17}+)	520197	alcanes (C_{17}+)					
unsaturated aliphatic hydrocarbons	521199	hydrocarbures aliphatiques insaturés					
aliphatic alcohols	520299	alcools aliphatiques					
aliphatic aldehydes	520599	aldéhydes aliphatiques					
aliphatic ketones	521599	cetones aliphatiques					
aliphatic esters	521499	esters aliphatiques					
chlorinated alkanes	520899	alcanes chlorés					
chlorinated alkenes	521399	alcènes chlorés					
fluorocarbons	521699	fluorocarbones					
glycol ethers	521999	éthers-oxydes					
aromatic hydrocarbons (mononuclear)	530199	hydrocarbures aromatiques mononucléaires					
aromatic hydrocarbons (polynuclear)	530198	hydrocarbures aromatiques polynucléaires					

Exposure	Code	Exposition	Cont.	Conc.	Freq.	Fiab.	Remarques
aromatic alcohols	530299	alcools aromatiques					
aromatic amines	530399	amines aromatiques					
phthalates	531799	phthalates					
isocyanates	531899	isocyanates					
(6) OTHERS		(6) AUTRES					
bleach	990022	décolorant					
cleaning agents (exc. solvents)	990005	produits de nettoyages (exc. solvants)					
cosmetics	990007	cosmétiques					
pharmaceuticals	990008	produits pharmaceutiques					
photographic products	990009	produits photographiques					
laboratory products	990012	produits de laboratoire					
fertilizers	990013	engrais					
pesticides	990014	pesticides					
biocides	990021	biocides					
(7) RADIATIONS		(7) RADIATIONS					
ionizing radiations	890001	radiations ionisantes					
radio frequency microwaves	890002	radio fréquences micro-ondes					
ultraviolet	890003	ultraviolet					

Appendix 4

Instructions to Chemists in the Use of The Exposure Coding Form

Written by M. Gérin for Use in the Cancer Study

UTILISATION DE LA FEUILLE D'EXPOSITION

1) ID: identifie le numéro du dossier (un dossier pour chaque patient).

2) job no: se rapporte au numéro chronologique de l'emploi considéré en commençant par le premier emploi occupé. On remplit une fiche d'exposition pour chaque emploi de chaque personne.

3) dates: année du début et de la fin de l'emploi en question. En règle générale, on ne considère pas un emploi de durée inférieure à 6 mois.

4) codes: a) code industriel à 3 chiffres tel que tiré de la "classification des activités économiques", 1970, Bureau fédéral de la statistique du gouvernement fédéral canadien. Ce code représente l'activité principale de la compagnie employeur (p. ex: industrie textile, construction, commerce, gouvernement).

b) code professionnel à 7 chiffres tel que tiré de "la classification canadienne descriptive des professions - 1971" du ministère de la main-d'oeuvre et de l'immigration du gouvernement fédéral canadien. Ce code représente la profession de l'employé durant cette période (p. ex: soudeur, nettoyeur, employé de bureau).

Remarques:
- lorsqu'il y a ambiguité ou indécision dans le choix d'un code, il faut choisir le code qui représente le mieux l'environnement de travail de la personne;
- un propriétaire d'entreprise qui serait soumis à des expositions devrait être classé plutôt comme le contremaître équivalent;
- un soldat qui exerce des fonctions inhabituelles qui se rapprochent plus de celles d'une autre profession devrait recevoir le code équivalent à cette profession (p. ex. mécanicien).

5) contact: code indiquant la voie d'exposition la plus probable.

1: voie respiratoire uniquement: pour poussières, gaz, fumées et vapeurs;

2: voie cutanée uniquement: indique un contact d'une matière liquide ou solide avec la peau (p. ex: solutions caustiques);

3: voies respiratoires et cutanées: lorsque l'exposition se fait simultanément par les deux voies (p. ex. pour la plupart des liquides organiques, s'il y a contact cutané, il y a aussi présence de vapeur et contact respiratoire).

Remarques:
- un contact sera généralement 1 ou 3, rarement 2. Le mode de contact est surtout utile pour distinguer l'exposition à un même produit sous forme de vapeur ou de liquide (série des "matières" inorganiques et organiques);
- pour les expositions physiques, on utilise le chiffre 9.

6) Concentration: représente le niveau relatif de concentration moyenne durant le temps d'exposition. Cet indice prend les valeurs de 1 à 3:

1: niveau faible

2: niveau moyen

3: niveau élevé

Il n'y pas de référence absolue pour classer une exposition à un niveau donné, mais il faut considérer une échelle différente pour chaque produit, échelle qui peut être reliée aux divers métiers conduisant à cette exposition (p. ex: poussière bois: pour un polisseur de meubles sans protection, concentration 3; pour un peintre dans le même atelier, concentration 1) (autre exemple: benzène: pour un travailleur manipulant du benzène pur: concentration 3; si le benzène est une impureté dans un solvant industriel: concentration 1).

7) fréquence: correspond à la proportion du temps de travail pendant laquelle l'exposition a lieu à un niveau proche de son niveau moyen.

1: moins de 5% du temps (moins de ½ h. par jour)

2: de 5 à 30% du temps (entre ½ h. et 3 h. par jour)

3: de 30 à 100% du temps (plus de 3 h. par jour).

On pense ici à une journée de travail de 8-9 heures environ.

Remarque: si dans la période de temps considérée l'exposition n'a pas lieu tous les jours, les pourcentages précédents s'appliquent également au niveau d'une semaine, d'un mois ou même d'une ou plusieurs années (p. ex. une semaine par an = fréquence 1).

8) fiabilité: correspond à l'estimation que se fait le codeur de la probabilité que l'exposition ait eu lieu effectivement.

3: correspond à une quasi certitude (ex: gazoline pour un pompiste)

2: indique une bonne probabilité (ex: émissions de soudage au gaz pour un mécanicien auto)

1: indique une possibilité mais faible ou une grande incertitude (ex: amiante et débardeur).

Plusieurs éléments peuvent concourrir à l'incertitude concernant la réalité d'un contact ou d'une exposition. Tout d'abord l'information tirée de l'interview peut être insuffisante pour se faire une idée précise de l'environnement de l'individu. Il peut également y avoir une probabilité intrinsèque qu'un produit donné ait été utilisé dans un emploi ou une industrie déterminée. A ceci s'ajoute le niveau d'incertitude du codeur lui-même qui juge de sa capacité à décerner cette exposition.

9) liste des expositions

La feuille d'exposition liste une série de substances chimiques et mélanges dans divers états physiques, et comprend également quelques procédés industriels et quelques expositions physiques.

a) code: à chacun des produits est associé un numéro à 6 chiffres ou code; ce code reflète la nature physique (poussière, gaz...) et la nature chimique de l'exposition considérée (inorganique, organique, fonctions chimiques...). Notons qu'il n'est pas nécessaire de connaître la base de ce code pour attribuer les expositions.

b) classes d'expositions: à chaque exposition correspond une définition précise. Les expositions sont regroupées en classes reconnaissables par le 1er chiffre du code:

- les poussières: 1
- les gaz: 2
- les fumées: 3

- les liquides et vapeurs (sous forme de vapeur, liquides ou solutions): 4
- les groupes chimiques: 5
- les expositions physiques: 8
- les autres (temporaires, procédés...): 9

Dans chaque classe où applicable, on divise les produits en inorganiques et organiques. On liste en premier les expositions correspondant à des mélanges (qu'ils soient de composition variable ou fixe) puis les expositions à des composés chimiques simples.

Les sections suivantes présentent une description générale de chacune des classes.

c) les poussières: les poussières résultent généralement de la désintégration mécanique d'un matériau solide en particules fines. On considère ici celles qui sont suffisamment fines pour être en suspension dans l'air et pénétrer dans le corps par le nez ou la bouche. Certaines fibres sont également incluses. On retrouve sur la liste des poussières inorganiques de type général (abrasifs, isolants) et spécifique (carbure de silicium, gypse) ainsi que des poussières et fibres organiques.

d) les gaz: les gaz comprennent deux groupes: les inorganiques et les organiques.

e) les fumées: les fumées sont composées de particules solides provenant de la condensation de substances formées lors du chauffage de la combustion ou décomposition à la chaleur de solides, liquides ou gaz (ex: fumées d'oxydes métalliques); dans un sens plus large, on inclut aussi les gaz et vapeurs en émanant (ex: gaz (fumées) d'échappement, fumées de soudage).

f) les liquides et vapeurs : sous ce terme on regroupe les produits et mélanges chimiques qui ne sont dans leur état habituel en milieu de travail ni des poussières, ni des gaz, ni des fumées. Il s'agit alors de substances normalement liquides ou dissoutes ou dispersées dans un solvant. Elles se manifestent sous forme liquide (contact cutané) ou sous forme de vapeurs (contact respiratoire) ou les deux.

On retrouvera sous forme liquide dissoute des composés inorganiques à faible volatilité, par exemple: acide phosphorique en solution dans l'eau, la plupart des sels

inorganiques solubles (ne sont pas listés), les bases en solution. Par contre les acides halogénés peuvent se retrouver également sous forme de vapeurs.

La liste des expositions organiques comprend des mélanges liquides complexes (ex: peintures, encres) pour lesquels l'exposition sera soit respiratoire soit cutanée et respiratoire à la fois. Il en est de même de la plupart des composés identifiés qui pourront se retrouver à l'état pur ou en mélange, dans une phase liquide ou vapeur (ex: benzène, trichloréthylène).

g) la liste des groupes chimiques: on a regroupé ici des classes d'exposition par leur nature chimique: le métal pour les composés métalliques, la fonction chimique pour les composés organiques. Ces catégories sont utilisées soit systématiquement chaque fois qu'un composé en faisant partie est codé ailleurs dans la liste (ex: hydrocarbures aromatiques et benzène) ou indépendamment si on ne désire pas spécifier le composé particulier (composés du plomb) ou lorsqu'on a affaire à un mélange de produits de même nature chimique sans en reconnaître les constituants précisément.

h) autres expositions: cette classe regroupe des produits ou mélanges difficilement classifiables ailleurs ou en attente de classification ou redéfinition.

i) expositions physiques: pour les deux expositions listées, on utilise le code 9 comme code de contact.

10) <u>méthode de codage</u>: Nous recommandons la façon suivante pour procéder au codage:

a) lecture de l'histoire occupationnelle complète du cas envisagé, c'est-à-dire les différentes feuilles d'interview et le questionnaire autoadministré; ceci permet de se faire une idée globale de la carrière de la personne et du type de spécialisation qu'elle a pu atteindre.

A ce stade, des codes industriels et occupationnels ont été attribués et les emplois regroupés si nécessaire.

b) lecture approfondie de chaque feuille d'interview correspondant à un emploi. Chacune des

questions posées devrait permettre de préciser l'environnement de travail de la personne. Il s'agit alors d'intégrer cette information (souvent incomplète ou hétéroclite) et de la compléter si nécessaire par la consultation: - d'ouvrages ou articles se rapportant à ce métier ou cette industrie, - de personnes ressources dans ces mêmes domaines, - de dossiers similaires antérieurement codés, et d'établir une liste d'expositions possibles.

c) on inscrit alors ces expositions sur la feuille d'expositions chimiques.

- si l'exposition fait partie de la liste, on entoure le code correspondant et on attribue des valeurs dans les colonnes contact, concentration, fréquence et fiabilité.

- si l'exposition n'en fait pas partie, on la rajoute dans un espace vide, si possible dans son groupe (poussières, gaz...). Un numéro de code lui sera attribué avant l'entrée dans l'ordinateur. On lui attribue les valeurs nécessaires dans les colonnes contact, concentration, fréquence et fiabilité.

d) on procède alors à une lecture-revision systématique de la feuille d'expositions groupe par groupe et produit par produit en incluant celles des expositions sur la liste qui auraient pu être oubliées lors de la première évaluation.

Appendix 5

Justification for Using the Collapsed Table Method as a Means of Assessing Inter-Rater Agreement

(A.5.1) Introduction

In this thesis, agreement has been measured for raters attributing exposure to physical and chemical agents by reviewing detailed job descriptions obtained usually by interview. A summary measure of agreement has been obtained by pooling each rater's assessment of exposure for each of the s substances on the exposure checklist and for each of the n job descriptions. The purpose is to obtain an "average" measure of inter-rater agreement over all substances and job descriptions. The process consists of adding together the cells of the s contingency tables corresponding to each substance. Thus, one "collapsed" table is formed having n x s observations. (This is referred to as the *collapsed table method*.)

At first sight, it might appear that the exposure has replaced the job description as the sampling unit. Heuristic arguments are presented in this Appendix that demonstrate that this is not the case. In fact, it will be shown that this process is identical to taking a simple arithmetic average of the index of crude agreement, p_0, over the s contingency tables. By way of an example, it will also be demonstrated that the averages of the expected value of this index, p_E, and the Kappa index, κ, are *approximately* equal to that obtained from the collapsed table. No formal proof of the latter two assertions will be given.

(A.5.2) An Example of the Validity of the Collapsed Table Method as a Means of Measuring Average Inter-rater Agreement

Consider the following two tables which represent agreement between two raters (R-1 and R-2) who assessed exposure on a dichotomous scale to two substances (s=2) for each of 30 job descriptions (n=30).

SUBSTANCE 1

R-1 \ R-2	-	+	
-	6	12	18
+	2	10	12
	8	22	30

$p_o = 0.533$

$p_e = 0.453$

$\kappa = 0.146$

SUBSTANCE 2

R-1 \ R-2	-	+	
-	5	3	8
+	2	20	22
	7	23	30

$p_o = 0.833$

$p_e = 0.624$

$\kappa = 0.556$

Suppose now that these two tables are combined, as described above, for the purpose of obtaining summary measures of agreement; viz,

COLLAPSED TABLE

R-1 \ R-2	-	+	
-	11	15	26
+	4	30	34
	15	45	60

$p_o = 0.683$

$p_e = 0.533$

$\kappa = 0.321$

The difference between the arithmetic average of each of these quantities, over the two strata, and the values derived from the collapsed table are small (<10%; see Table A.5.1). Thus, it appears that the collapsed table method provides a means of summarizing inter-rater agreement over the complete set of exposures.

Table A.5.1

Comparison of Agreement Indices

Agreement Statistic	Value of agreement statistic for: Substance Number 1	Substance Number 2	Average (1)	Collapsed Table	% Difference (2)
p_o	0.533	0.833	0.683	0.683	0.0
p_e	0.453	0.624	0.539	0.533	1.1
κ	0.146	0.556	0.351	0.321	9.4

(1) Arithmetic average of the values obtained from the contingency tables for Substances 1 and 2.

(2) Calculated as (Average-Collapsed)/Collapsed x 100%.

(A.5.3) More General Considerations

For the index of crude agreement it can also be shown that when the number of job descriptions (ie, observations) in each table is equal, the average of p_o over the original tables will always be equal to that obtained from the collapsed table. When the condition of equal sample sizes does not hold, the index of crude agreement obtained from the collapsed table is identically equal to a weighted average over the original tables. The weights are equal to the inverse of the total number of observations in each table.

It appears, however, that the value of Kappa obtained from the collapsed table can not be written as a weighted average over the constituent contingency tables. Various weighting schemes have been tried (see Table A.5.2), but all yield values different from that obtained from the collapsed table.

It would also be of interest to compare the value obtained from the collapsed table to that obtained when an average is formed by weighting each individual value of Kappa by the reciprocal of its estimated variance (Fleiss, 1981). This procedure would be applied only if it was believed that the underlying inter-rater agreement for each of the substances was equal. In general, this would not be the case, although it might be a reasonable assumption for certain clusters of substances. In any event, this procedure was not followed in the thesis since, as was noted in Chapter 3, little confidence can be placed in asymptotic (and, perhaps, jackknifed) variances when table sizes are small ($n<50$),

as is encountered in this thesis.

Lastly, there may be problems with the estimates of the variance of Kappa obtained from the collapsed table. The variance is calculated from the collapsed table using asymptotic statistical theory which assumes that each entry in the table is an independent observation. Since the values in the collapsed table are not independent it is not clear whether the variance obtained from the collapsed table, and the resultant confidence intervals, are interpretable in the usual statistical way. Nevertheless, confidence intervals are calculated for all comparisons.

Table A.5.2

Weighted Averages of Kappa Derived from the Previous Example (1)

Weight Factor	Value of κ derived from use of the Weight Factor	% Difference with respect to collapsed value (2)
1	0.351	9.4
$1 - p_2$	0.313	-2.5
$\frac{1 - p_2}{(p_{12} + p_{21})}$	0.451	41.9

Table A-5-2, continued

(1) A weighted average for K is defined as:

$$\frac{\sum x_i * W_i}{\sum W_i}$$

where the W_i are weights and the x_i are the individual values of Kappa obtained for each stratum.

(2) The value obtained from the collapsed table is 0.321. The percent difference is calculated from the formula: (Weighted Average-Collapsed)/Collapsed x 100%.

Appendix 6

A FORTRAN Program for the Calculation of Eggs Statistics

```
      PROGRAM KAPPAS(BLDOUT,OUTPUT,LABEL,TAPE1=BLDOUT,
     *               TAPE6=OUTPUT,TAPE2=LABEL,OUTMAP,TAPE8=OUTMAP)
C
C
C
C
C     ******************************************************************
C
C
C          PROGRAM NAME: KAPPAS                 OBJECT DECK: KAPPA
C                  DATE: 13SEP83                        NAME: M. GOLDFARB
C
C
C
C
C               P R O G R A M    D E S C R I P T I O N
C               = = = = = = =    = = = = = = = = = = =
C
C
C     THIS PROGRAM IS DRIVEN BY THE OUTPUT FROM SPSS CROSSTABS
C     OPTION=10
C
C     IT IS ASSUMED THAT THE FORM FOR EACH CROSSTABS TASK
C     WILL BE AS FOLLOWS:
C
C         TABLES = RATER1 (LIST) BY RATER2 (LIST) BY STRATUM (1 VAR)
C
C                  RATER1 SHOULD NOT CORRESPOND WITH ANY OF THE VARS
C                  IN RATER2 LIST (EG TABLES=  R2 BY R2 R3 R4 BY STRATA
C                  IS NOT MEANINGFUL)
C                  THEREFORE,
C
C           TABLES= R1 R2 BY R3 R4 R5 BY STRATA
C
C                  IS CORRECT.
C
C     THE LABELS FILE REQUIRES NAMELIST AND THE NAMES FOR RATER1, RATER2
C     AND STRATA
C     THEREFORE, THE FORMAT FOR THE ABOVE EXAMPLE IS:
C
C          (COLUMN 2 FOR NAMELIST)
C               $
C               $PARMS
C               TITLE=TEXT,MODE=X,DIM=X,ETC...
C               $END
C              RATER1 LABEL
C              RATER2 LABEL
C         --- BLANK LINE      ---
C              RATER3 LABEL
C              RATER4    =
C              RATER5    =
C         --- BLANK LINE      ---
C              STRATUM 1 LABEL
C              STRATUM 2 LABEL
C                .       .
C                .       .
C              STRATUM N LABEL
C         --- BLANK LINE      ---
```

```
C
C       OTHER FORMS FOR TABLES= CARD ARE NOT ALLOWED. FOR EXAMPLE,
C
C                       TABLES= B1 BY B2 BY B3 / B5 BY B7
C
C       WILL PRODUCE UNPREDICTABLE RESULTS
C
C       IN ADDITION, ALL VALUES FOR RATER1, RATER2 AND STRATA
C       VARIABLES MUST BE ORDERED, STARTING FROM 1, BY
C       CONSECUTIVELY INCREASING INTEGERS (IE, 1, 2, 3, .... )
C
C       A TASK IS DEFINED AS A TABLES= CARD
C       MULTIPLE SETS OF LABELS ARE REQUIRED (AS ABOVE)
C       THE NAMELIST MUST APPEAR AT LEAST ONCE. BY SETTING
C       ONENAM=T THEN NO FURTHER NAMELISTS WILL BE READ.
C       SAMLAB=T FIRST SET OF LABELS TO BE USED THROUGHOUT
C       IN ADDITION, BY SETTING SAMLAB=T ONLY ONE SET OF
C       LABELS ARE REQUIRED ALTHOUGH MULTIPLE NAMELISTS
C       ARE PERMITTED.
C
C         NAMELIST PARMS:
C
C                       TITLE = TITLE OF JOB (80CHARS)
C                       MODE  = CALCULATION MODE (DEFAULT = 0 )
C                       WEIGHT = WEIGHT USED IN TABLE
C                                SYMMETRIC MATRIX OF DIMENSION #DIM#
C                                DIAGONAL ELEMENTS ARE UNITY (IE,
C                                   WEIGHT(J,J)=1. - THIS IS DEFAULT)
C                                OFF DIAGONAL ELEMENTS 0<=WEIGHT<1.
C                       DIM = DIMENSION OF WEIGHT ARRAY (MUST BE
C                                SAME SIZE AS CROSSTAB TABLES)
C                                DEFAULT = 0
C                     ONENAM = LOGICAL VARIABLE INDICATING ONLY ONE
C                                NAMELIST READ TO BE DONE FOR ALL
C                                FURTHER CROSSTAB TASKS
C                                DEFAULT IS .TRUE.
C
C                       SAMLAB = IF TRUE THEN ONE SET OF LABELS USED
C                                FOR ALL FURTHER TASKS
C                                DEFAULT = FALSE
C                       SAMSTR = IF TRUE THEN ONE SET OF STRATA LABELS USED
C                                FOR ALL FURTHER TASKS
C                                DEFAULT = FALSE
C
C                       JACKNF = IF TRUE THEN JACKKNIFE ESTIMATES OF
C                                VARIANCE WILL BE CALCULATED
C                                DEFAULT = TRUE
C
C                       MODE REFERS TO TYPE OF STATS REQUIRED
C
C                       MODE = 0 - INDIVIDUAL KAPPAS BASED ON 2 RATERS
C                                  K<=30 CATEGORIES
C                                - THIS IS THE DEFAULT
C
C
C                       MODE = 1  - SAME AS MODE = 0 BUT SUMMARY KAPPA OVER
C                                   ALL STRATA
C
C                       MODE = 2  - SAME AS MODE = 0 BUT SUMMARY KAPPA OVER
C                                   ALL TABLES
```

```
C
C       NOTES
C
C           LABELS HAVE A MAXIMUM OF 20 CHARACTERS
C           IF LIMITS EXCEEDED OR NO BLANK LINE APPEARS AT END
C               OF LABELS THEN NO LABELS WILL BE PRODUCED.
C           A MAX OF 200 LABELS IN ALL 3 CATEGORIES IS ALLOWED
C           TOTAL MAX NO. OF TABLES IS 500 (MODE = 1)
C           NOTE THAT THE MAXIMUM NO. OF TABLES MAY BE RESTRICTED
C            BY SPSS TO LEVELS BELOW THE MAX DIMENSIONS
C           MAX SIZE OF ARRAY IS 30 X 30
C           ONLY 1 CONTROL VARIABLE IS ALLOWED
C
C           THIS PROGRAM CANNOT BE OVERLAID
C
C       FILES REQUIRED
C
C       FILE NAME     UNIT    CONTENTS
C        BLDOUT         1     OUTPUT FROM SPSS CROSSTABS, OPTION 10
C        LABEL          2     NAMELIST PARMS (SEE *GETLAB*)
C        OUTPUT         6     KAPPA STATS OUTPUT
C        OUTMAP         9     KAPPA STATS OUTPUT ONTO A FILE FOR FURTHER
C                             PROCESSING
C
C       --REQUIREMENTS OF INPUT FILES
C
C       BLDOUT          OUTPUT BY SPSS CROSSTABS OPTION 10.
C                       SINCE CDC VERSION OF SPSS CREATES A FILE
C                       CALLED *BLDOUT1* FOR THIS PURPOSE, THIS
C                       FILE MUST BE RENAMED BEFORE BEING USED.
C
C       LABEL           FOR EACH CROSSTABS PROCEDURE IN THE SPSS RUN,
C                       THE FOLLOWING LINES MUST BE ENTERED AS ABOVE.
C
C      ***********************************************************
C
C                   U P D A T E   L O G
C                   = = = = = =   = = =
C
C         DESCRIPTION                                       DATE
C         ***********                                       ****
C
C          BUG FIXED WHEN THERE ARE NO STRATA FOR                19JAN84
C              SUMMARY CALLS (KAPOVR) - VAR=NOSUMM
C
C         INDEX OF CRUDE AGREEMENT ADDED TO PRINTOUT
C
C
C
C
C      ***********************************************************
C
```

```
      IMPLICIT INTEGER (A - Z)

      CHARACTER*80 STRATA
      CHARACTER*10 TITLE*20,DATE,TIME,M(2)
      CHARACTER*20 RATE1(200), RATE2(200), STRATA(200)
      CHARACTER*20 BLNK20
      REAL WEIGHT(30,30), TABLE(30,30), DF
      REAL KAPPA(200), CRSE(200), NULSE(200), KAPCNT, KAPSUM,
     +     PVALUE, DF, KAPMAX(200), PL(200), VARJCK(200), SVJACK(500),
     +     PO(200)
      REAL SVKAP(500), SVLRSE(500), CI, CIT, KAPAVG, KAPSMJ
      REAL C1, C2, C3, C4, KAPCNJ, PVALUJ, DFJ
      LOGICAL UNDEF(200), SVUND(500), NOVAR(200), SVNOVR(500),
     +        BLNKST, END1, NOSUMR

      DIMENSION IDP(2), SIZTBL(200)
      REAL ROWV, COLV, STRTV
      DIMENSION NUMBER(200)
      LOGICAL END, NOLAB, SAMLAB, NAML, WT, OVERAL, SAMSTR,
     +        JACKNF

      COMMON /T/ M,ITITLE
      COMMON / MAX / MAXTAB, MAXSTR, MAXARY


      COMMON  / LABS / TITL, WT, NRATE1, NRATE2, NSTR, NAML,
     +                 JACKNF

C
C.... DIM = SIZE OF WEIGHT ARRAY (ENTERED FROM NAMELIST)
C.... NRATE1 = NO OF RATER1 LABELS
C.... NRATE2 = NO OF RATER2 LABELS
C.... NSTR   = NO OF STRATUM LABELS
C.... NAML   = IF TRUE THEN ONLY ONE NAMELIST READ IN ALL LABEL SETS
C.... SAMLAB = IF TRUE SAME SET OF LABELS AS FIRST TO BE USED THROUGHOUT
C.... WT = IF TRUE WEIGHTS ENTERED
C.... NOSTRT = NO. OF TABLES IN STRATA (RESET TO ZERO FOR EACH NEW
C....          *SPSS TABLE*)
C.... TOTABL = TOTAL NO. OF TABLES  OVER ALL CROSSTABS TASKS
C.... MAXTAB = MAX NO. OF TABLES  OVER ALL CROSSTABS TASK
C.... MAXSTR = MAX NO. OF STRATA
C.... NCRSTB = NO. OF CROSSTABS TASKS CARDS GENERATED
C.... MAXARY = MAX SIZE OF TABLE (MAXARY X MAXARY)
C.... UNDEF(K) = IF TRUE, THEN KAPPA(K) IS UNDEFINED
C.... NOVAR(K) = IF TRUE, THEN VARIANCE OF KAPPA IS UNDEFINED
C.... BLNKST  = IF SET TO .TRUE. THEN XTAB TASK HAS NO STRATA
C.... NOSUMR  = IF SET TO .TRUE. THEN THE SUMMARY OVER STRATA IS UNDEFINED
C
C

      DATA RATE1,RATE2,STRATA/600*'                    '/
      DATA BLNKST / .FALSE. /
      DATA UNDEF, SVUND, NOVAR, SVNOVR / 1400 * .FALSE. /
      DATA BLNK20/'                    '/
      DATA TABLE, WEIGHT / 900 * 0., 900 * 0./
      DATA SVKAP, SVLRSE /500*0., 500*0. /
      MAXTAB = 500
      MAXSTR = 200
      MAXARY = 30
      NCRSTB = 0
      SAMLAB = .FALSE.
      SAMSTR = .FALSE.
      OVERAL = .FALSE.
      NOSUMR = .FALSE.
```

```
C... INITIALIZE WEIGHT

      DO 50 K = 1, 50
         WEIGHT(K,K) = 1.
 50   CONTINUE

      NOSIRT = 0
      DIM = 0
      TOTABL = 0
      END  =  .FALSE.
      MODE = 0
      NOLAB = .FALSE.

C.... GET DATE AND TIME FOR THIS RUN

      H(1) = DATE()
      H(2) = TIME()

C.... GET TABLES, IF THERE ARE ANY

      CALL GETTAB( RATE1, RATE2, STRATA,
     +             WEIGHT, NOLAB, SAMLAB, SAMSTR, MODE )
      CALL WRITABL( RATE1, RATE2, STRATA,
     +             WEIGHT, NOLAB, MODE )

C
C.... FORMAT FOR SPSS OUTPUT (OUTT = 1)
C
C.... INP(1) = TABLE NO. (REFERS TO TAKING 1 VARIABLE AT A TIME
C....          FROM EACH VAR LISTS CONNECTED WITH A BY)
C.... INP(2) = CELL FREQUENCY (UNWEIGHTED)
C.... INP(3) = VALUE OF ROW VARIABLE
C.... INP(4) = VALUE OF COL VARIABLE
C.... INP(5) = VALUE OF CONTROL VARIABLE
C
C

      LINO = 60
      READ(1,1000,END=900) INP, ROWV, COLV, STRALP
C1000 FORMAT(4X,F4.0,4F8.0)
 1000 FORMAT(4X,I4,I8,2F8.0,A8)

      IF ( STRALP .EQ. BLNK20(1:8) ) THEN
         STRTV = 1
         BLNKST = .TRUE.
      ELSE
         READ ( STRALP, *(F8.0)* ) STRTV
      END IF

      ROW = IROW = INT(ROWV)
      COL = ICOL = INT(COLV)
```

```
C
C.... INITIALIZE FOR NEW CROSSTABS TABLE AND
C
  101 CONTINUE
      CALL ZERO ( KAPPA, MAXSTR )
      CALL ZERO ( ERRSE, MAXSTR )
      CALL ZERO ( NULSE, MAXSTR )
      CALL ZERO ( KAPMAX, MAXSTR )
      CALL ZERO ( VARJCK, MAXSTR )
      CALL ZERO ( PC, MAXSTR )
      CALL ZERO ( PO, MAXSTR )
      DO 106 K = 1, MAXSTR
        UNDEF(K) = .FALSE.
        NOVAR(K) = .FALSE.
        NUMDEF(K) = 0
  106 CONTINUE
      NCRSTB = NCRSTB + 1

C
C.... INITIALIZE FOR NEW TABLE
C

  102 CONTINUE
      NUSTR1 = 0
      IPO = 00
      LTABLE = INP(1)
      ROW = INT ( ROWV )
      COL = INT ( COLV )
      END1 = .FALSE.

C
C.... INITIALIZE FOR NEW STRATUM
C

  105 CONTINUE
      CALL ZERO ( TABLE, MAXARY * MAXARY )
      ISTRAT = INT ( STRTV )
      TABLE( ROW, COL ) = INP(2)

C
C.... GET NEXT CELL FROM SPSS CROSSTABS OUTPUT
C

  110 CONTINUE
      READ(1,1000,IOSTAT=IOS,END=250) INP, ROWV, COLV, STRALP

  118 CONTINUE
      IF ( STRALP .EQ. BLNK20(1:8) ) THEN
         STRTV = 1
         BLNKST = .TRUE.
      ELSE
         READ ( STRALP, '(F8.0)' ) STRTV
      END IF

      ROW = INT ( ROWV )
      COL = INT ( COLV )
      NTABLE = INP(1)
      NSTRAT = INT ( STRTV )

C.... CHECK FOR END OF TABLE CONDITIONS
C....       - STRATA NO < LAST STRATA NO
C....       - COL < LAST COL
```

```
C.... DEBUG
C     WRITE ( 6, * ) ROW, LROW, COL, LCOL, END1, NSTRAT, LSTRAT

      IF ( .NOT. BLNKST .AND. (NSTRAT .NE. LSTRAT)) GO TO 300
      IF ( BLNKST .AND. (NTABLE .NE. LTABLE) ) END1 = .TRUE.
C     IF ( BLNKST .AND. ( ROW .LT. LROW )) END1 = .TRUE.
      IF ( BLNKST .AND. ( COL .LT. LCOL )) END1 = .TRUE.
      IF ( END1 ) GO TO 300

C.... PUT CELL INTO TABLE

  200 CONTINUE
      TABLE ( ROW, COL ) = INP(2)
      LROW = ROW
      LCOL = COL
      GO TO 110

  250 END = .TRUE.

C
C**** START PROCESSING CURRENT STRATUM TABLE
C

  300 CONTINUE
      TOTABL = TOTABL + 1
      NOSTRT = NOSTRT + 1

C.... DEBUG
C     WRITE (6,*) IOS, END1, END, TOTABL, ((TABLE(J,K),J=1,4),K=1,4),
C    +            ROW, LROW, COL, LCOL, BLNKST, NTABLE, LTABLE,
C    +            NSTRAT, LSTRAT


      IF ( NOSTRT .GT. MAXSTR ) THEN
         WRITE ( 6, 1100 ) NOSTRT, MAXSTR
 1100    FORMAT (10 **** NO OF STRATA TABLES (',I5,') ',
     +            '(GREATER THAN MAXIMUM (',I5,')'/)
         GO TO 900
      END IF

C
C.... CALCULATE KAPPA FOR TABLE AND SID ERRORS
C

      SIZTBL(NOSTRT) = MAX0 ( LROW, LCOL )
      CALL KAPPAC ( MODE, KAPPA(NOSTRT), KAPMAX(NOSTRT),
     +      PO(NOSTRT), ERRSE(NOSTRT), NULSE(NOSTRT),
     +  NUMDER(NOSTRT), UNDEF(NOSTRT), SIZTBL(NOSTRT), TABLE, MLIGHT,
     +     NOVAR(NOSTRT), VARJCK(NOSTRT) )

C.... SAVE KAPPAS AND SID ERRORS FOR MODE=2 PROCESSING

      IF ( MODE .EQ. 2 .AND. TOTABL .LE. MAXTAB) THEN
         SVKAP(TOTABL) = KAPPA(NOSTRT)
         SVLPSE(TOTABL) = ERRSE(NOSTRT)
         SVUND (TOTABL) = UNDEF(NOSTRT)
         SVJACK(TOTABL) = VARJCK(NOSTRT)
         SVNOVR(TOTABL) = NOVAR(NOSTRT)
      END IF
```

```
C
C**
C**** IF NEW TABLE DETECTED, WRITE SUMMARY TABLE AND RESULTS
C**
C

  360  CONTINUE
       IF ( LEND .OR. EOF1 ) GO TO 360

       IF ( NSTRAT .LT. LSTRAT ) GO TO 360
       IF ( NTABLE .EQ. LTABLE ) GO TO 105

  360  CONTINUE
       IF ( LINE .GE. 60 ) GO TO 305
       CALL HEADG ( .TRUE., NOLAB, NR1, NR2, RATE1, RATE2, LINE)

  305  CONTINUE

C
C.... CALCULATE SUMMARY KAPPA AND CHI SQUARE FOR HOMOGENEITY
C

       IF ( MODE .EQ. 1 .OR. MODE .EQ. 2 .AND. NOSTRT .GT. 1) THEN
          CALL KAPUVR ( KAPPA, LRGSE, UNDEF, NOVAR, KAPAVG, MODE,
     +                  NOSTRT, KAPSUM, CI, KAPCHI, PVALUE, DF, *355)
          C1 = KAPSUM + 1.96 * SQRT(1./CI)
          C2 = KAPSUM - 1.96 * SQRT(1./CI)

C.... JACKKNIFE

       IF ( JACKNF ) THEN
          CALL KAPUVR ( KAPPA, VARJCK, UNDEF, NOVAR, KAPAVG, MODE,
     +                  NOSTRT, KAPSMJ, CIJ, KAPCHJ, PVALUJ, DFJ, *355)
          C3 = KAPSMJ + 1.96 * SQRT(1./CIJ)
          C4 = KAPSMJ - 1.96 * SQRT(1./CIJ)
       END IF

       END IF

       GO TO 370

C.... SUMMARY CANNOT BE DONE

  355  CONTINUE
       NOSUMR = .TRUE.

  370  CONTINUE


C.... GET POINTERS TO LABELS

       NR1 = 1 + (LTABLE - 1) / NRATE2
       NR2 = LTABLE - NRATE2 * ( NR1 - 1 )

       CALL WAY( NR1, NR2, LSTRAT, LINE, RATE1, RATE2, STRATA,
     +   KAPPA, NOLAB, LRGSE, NOLSE, KAPSUM, KAPCHI,
     +   C1, C2, C3, C4, PVALUE, DF, NUMBER, UNDEF, NOVAR,
     +   OVERAL, KAPMAX, PC, PO, KAPAVG,
     +   KAPSMJ, VARJCK, PVALUJ, DFJ, KAPCHJ, SIZTHL, NOSUMR )
       NOSUMR = .FALSE.
```

```
      IF ( END ) GO TO 400
      IF ( NSTRAT .LT. ISTRAT .AND. NTABLE .LE. LTABLE) GO TO 375
      ICOL = LROW = 1
      IF ( NTABLE .GT. LTABLE ) GO TO 102
      IF ( END1) GO TO 375

C
C**** NEW CROSSTABS TASK: GET TABLES, ETC.
C
C**** INITIALIZE
C

  375 CONTINUE
      IF ( SAMLAB ) GO TO 390
      DO 380 K = 1, MAXSTR
         RATE1(K)=RATE2(K)=BLNK20
  380    CONTINUE

      IF ( SAMSTR ) GO TO 390
      DO 385 K = 1, MAXSTR
         STRATA(K)=BLNK20
  385    CONTINUE

  390 CONTINUE
      CALL GETLAB(RATE1, RATE2, STRATA, WEIGHT, NOLAB,
     +            SAMLAB, SAMSTR, MODE)
      CALL WRTLAB( RATE1, RATE2, STRATA, WEIGHT,
     +             NOLAB, MODE )
      LINE = 60
      ICOL = 1
      IROW = 1
      END1 = .FALSE.
      GO TO 101

C
C**** END OF PROCESSING
C

C**** IF MODE=2 CALCULATE OVERALL KAPPA

  400 CONTINUE
      IF ( MODE .EQ. 2 .AND. TOTABL .LE. MAXTAB) THEN
         OVERAL = .TRUE.
         CALL KAPOVR(SVKAP, SVLRSE, SVUND, SVNOVR, KAPAVG, 1,
     +               TOTABL, KAPSUM, C1, KAPCHI, PVALUE, DF, *900)
         C1 = KAPSUM + 1.96 * SQRT(1./C1)
         C2 = KAPSUM - 1.96 * SQRT(1./C1)

C**** JACKKNIFE

         IF ( JACKNF ) THEN
            CALL KAPOVR(SVKAP, SVJACK, SVUND, SVNOVR, KAPAVG, 1,
     +                  TOTABL, KAPSMJ, C1J, KAPCHJ, PVALUJ, DFJ, *900)
            C3 = KAPSMJ + 1.96 * SQRT(1./C1J)
            C4 = KAPSMJ - 1.96 * SQRT(1./C1J)
         END IF

         CALL NOSUM( TOTABL, LINE, KAPSUM, KAPCHI, C1, C2, C3, C4,
     +               PVALUE, DF, OVERAL, KAPAVG, KAPSMJ,
     +               PVALUJ, DFJ, KAPCHJ, NOSUMR )
         OVERAL = .FALSE.
         NOSUMR = .FALSE.
      END IF
```

```
C
C**** END
C
  900   CONTINUE
        STOP
        END
        SUBROUTINE GETLAB( RATE1, RATE2, STRATA, WEIGHT, NOLAB,
     +                     SAMLAB, SAMSTR, MODE )
C
C
C       READ LABEL FILE AND SET LABEL FLAG TO TRUE IF
C       VALID LABELS HAVE NOT BEEN READ
C
        IMPLICIT INTEGER ( A - Z )
        CHARACTER*20 RATE1(200), RATE2(200), STRATA(200)
        CHARACTER*10 TITLE*60,H(2)

        LOGICAL NAMEL, NOLAB, ONENAM, WT, SAMLAB, SAMSTR, FIRST
        LOGICAL JACKNF
        REAL WEIGHT ( 30, 30 )

        COMMON / LABS / DIM, WT, NRATE1, NRATE2, NSTRAT, NAMEL,
     +                  JACKNF
        COMMON /T/ H,TITLE
        COMMON / MAX / MAXTAB, MAXSTR, MAXARY

C
C.... MODE REFERS TO TYPE OF STATS REQUIRED
C.... MODE = 0    - INDIVIDUAL KAPPAS BASED ON 2 RATERS FOR K<=30
C....               CATEGORIES
C....             - THIS IS THE DEFAULT
C
C.... MODE = 1    - SAME AS MODE = 0 BUT SUMMARY KAPPA OVER
C....               ALL STRATA WILL BE CALCULATED
C
C.... MODE = 2    - SAME AS MODE = 0 BUT SUMMARY KAPPA OVER
C....               ALL TABLES WILL BE CALCULATED
C....  DIM = DIMENSION OF WEIGHT ARRAY
C.... ONENAM = ONLY ONE NAMELIST TO BE READ (DEFAULT = TRUE)
C.... SAMLAB = SAME LABELS AS IN FIRST SET TO BE USED IN ALL TABLES
C.... SAMSTR = SAME STRATA LABELS ONLY TO BE USED THEREAFTER
C.... JACKNF = JACKKNIFED VARIANCES CALCULATED
C

        NAMELIST / PARMS / TITLE, MODE, WEIGHT, DIM, ONENAM, SAMLAB,
     +                     SAMSTR, JACKNF

        NOLAB = .FALSE.
        DATA FIRST, NAMEL, ONENAM /  3 * .TRUE. /

        IF ( NAMEL ) READ(2,PARMS,END=100)
        IF ( ONENAM ) NAMEL = .FALSE.
        IF ( .NOT. ONENAM ) NAMEL = .TRUE.
        IF ( DIM .NE. 0 ) WT = .TRUE.
        IF ( DIM .EQ. 0 ) THEN
           WT = .FALSE.
           CALL ZERO ( WEIGHT, MAXARY * MAXARY )
           DO 50 K = 1, MAXARY
              WEIGHT(K,K) = 1.
   50      CONTINUE
        END IF
```

```
C
C.... READ LABELS FOR RATE1 (MAX = 200 )
C

      IF ( SAMLAB .AND. .NOT. FIRST ) GO TO 8000
      NRATE1 = 1
      NRATE2 = 1
  100 CONTINUE
      READ(2,1001,END=900) RATE1(NRATE1)
 1001 FORMAT(A20)

C.... END OF VALID RATE1 LABELS IS FLAGGED WITH A BLANK LINE

      IF( RATE1(NRATE1) .EQ. ' ' ) GO TO 200

      IF( NRATE1 .GT. MAXSTR ) THEN
         WRITE(6,2000)
 2000    FORMAT(' *** TOO MANY LABELS: ALL LABELS IGNORED ***')
         NOLAB = .TRUE.
         GO TO 8000
      ELSE
         NRATE1 = NRATE1 + 1
         GO TO 100
      END IF

  200 CONTINUE
      NRATE1 = NRATE1 - 1

C
C.... READ RATE2 LABELS
C

  300 CONTINUE
      READ(2,1001,END=900) RATE2(NRATE2)

C.... END OF VALID RATE2 LABELS IS FLAGGED WITH A BLANK LINE

      IF( RATE2(NRATE2) .EQ. ' ' ) GO TO 350

      IF( NRATE2 .GT. MAXSTR ) THEN
         WRITE(6,2000)
         NOLAB = .TRUE.
         GO TO 8000
      ELSE
         NRATE2 = NRATE2 + 1
         GO TO 300
      END IF
```

```
  550  CONTINUE
       NRATE2 = NRATE2 - 1

C
C.... READ STRATA LABELS
C

       IF ( SAMSTR .AND. .NOT. FIRST ) GO TO 8000
       NSTRAT = 1
  400  CONTINUE
       READ ( 2, 1001, END = 500 ) STRATA(NSTRAT)
       IF ( STRATA(NSTRAT) .EQ. ' ' ) GO TO 500
       NSTRAT = NSTRAT + 1
       IF ( NSTRAT .GT. MAXSTR ) GO TO 500
       GO TO 400

C.... END OF STRATA LABELS

  500  CONTINUE
       NSTRAT = NSTRAT - 1
       GO TO 8000

C
C.... END OF LABELS FOR RATER1 AND RATER2 ONLY
C

  900  CONTINUE
       IF ( SAMLAB ) GO TO 8000
       WRITE(6,2002)
 2002  FORMAT(T0,*** UNEXPECTED END OF LABEL FILE: LABELS IGNORED ***/1)
       NOLAB = .TRUE.
       LINE = LINE + 3

C
C**** RETURN
C

 8000  CONTINUE
       FIRST = .FALSE.
       RETURN
       END
       SUBROUTINE VARLAB( RATE1, RATE2, STRATA, WEIGHT, NOLAB, MODE)
C
C      WRITE OUT LABELS READ IN ORDER TO CHECK WITH TABLE INDEX
C      CREATED BY SPSS CROSSTABS OPTION 10
C
       IMPLICIT INTEGER ( A - Z )
       CHARACTER*20 RATE1(200), RATE2(200), STRATA(200)
       REAL WEIGHT (50,50)
       LOGICAL NAMEL, WT, NOLAB, JACKNF

       COMMON / LABS / DIM, WT, NRATE1, NRATE2, NSTRAT, NAMEL,
      +                JACKNF

       IF ( NAMEL ) THEN
          WRITE ( 6, 1000 ) MODE, NRATE1, NRATE2, NSTRAT
 1000  FORMAT ( T0 PARAMETERS SET IN THIS RUN:T/
      +         T  ********** *** ** **** ****T//
      +         T  MODE .....T, I2 //
      +         T  NO OF RATER1 LABELS .....T, I3 //
      +         T  NO OF RATER2 LABELS .....T, I3 //
      +         T  NO OF STRATA LABELS .....T, I3 //)
          LINE = LINE + 13
       END IF
```

```
C.... WRITE OUT WEIGHT TABLE

      IF ( LABEL .AND. WT) THEN
         WRITE ( 6, 1200 )
 1200    FORMAT ( '0 WEIGHT TABLE ENTERED AS FOLLOWS:'/
     +           ' ' ****** ***** ******* ** *******'///)
         LINE = LINE + 2

         DO 50 J = 1, DIM
            WRITE ( 6, 1250 ) (WEIGHT(J,I), I = 1, DIM )
 50         CONTINUE
 1250    FORMAT ( '0',3(10(F5.3,3X/) ) )
      END IF
      CALL HEAD6( .TRUE., NOLAB, NR1, NR2, RATE1, RATE2, LIN6)

      IF ( .NOT. NOLAB ) THEN
         WRITE(6,2000)
 2000    FORMAT('0  CHECK THE FOLLOWING INDEX AGAINST THE INDEX '/
     >   1X,'  PROVIDED BY SPSS CROSSTABS OPTION 10.  IF THE'/
     >   1X,'  CORRESPONDENCE IS NOT EXACT, THE LABELS ARE PROBABLY'/
     >   1X,'  INCORRECT IN THE FOLLOWING LISTING')

         IF ( MODE .LE. 2 ) THEN
            WRITE(6,2001)
 2001       FORMAT('0',10X,'TABLE  RATER1',18X,'RATER2',/
     +           1X,'0',12,'0',/)

            LINE = 12
            I = 0
            DO 100 IS=1,NRATE1
               DO 100 IX=1,NRATE2
                  I = I+1
                  WRITE(6,2002) I, RATE1(IS), RATE2(IX)
                  LINE = LIN6+1
                  IF( LINE .GE. 60 ) THEN
                     CALL HEAD6( .TRUE., NOLAB, NR1, NR2,
     +                    RATE1, RATE2, LINE)
                  END IF
 100           CONTINUE
      END IF
 2002    FORMAT(1A,'1',I4,2X,A20,4X,A20)

C.... WRITE OUT STRATA

         IF ( NSTRAT .EQ. 0 ) GO TO 8000
         CALL HEAD6( .TRUE., NOLAB, NR1, NR2,
     +                RATE1, RATE2, LINE)
         WRITE ( 6, 2003 )
 2003    FORMAT ('0',' STRATUM LABELS ARE AS FOLLOWS:'/
     +           ' ' ******* ****** *** ** ********'///
     +           '0', 'STRATUM  NAME OF STRATUM'/ 13,'0','///)
```

```
            I = 0
            DO 200 K = 1, NSTRAT
               I = I + 1
               WRITE ( 6, 2004 ) I, STRATA(K)
               LINE = LINE + 1
               IF ( LINE .GE. 60 ) THEN
                  CALL HEADER( .TRUE., NOLAB, NR1, NR2,
     +                         RATE1, RATE2, LINE)
               END IF
  200       CONTINUE

 2004       FORMAT ( I12, I4, 3X, A20 )
            END IF

 8000       CONTINUE
            RETURN
            END
            SUBROUTINE HEADER ( ONLYT, NOLAB, NR1, NR2, RATE1, RATE2, LINE)
C
C           WRITE HEADER LINE WITH TIME AND DATE STAMP
C
            IMPLICIT INTEGER ( A - Z )
            CHARACTER*10 TIME,DATE*80,H(2)
            CHARACTER*20 RATE1(NR1), RATE2(NR2)
            LOGICAL NOLAB, ONLYT, NAMEL, JACKNF

            COMMON  / LABS / DIM, MT, NRATE1, NRATE2, NSTR, NAML,
           +                 JACKNF
            COMMON /T/ H,TITLE
            WRITE(6,2000) TITLE,H
 2000       FORMAT(T11,I10,A60,5X,A10,2X,A10//)
            LINE = 2

            IF ( ONLYT ) RETURN

            IF ( .NOT. NOLAB )
           +   WRITE(6,9400) RATE1(NR1), RATE2(NR2)
 9400       FORMAT ( T01, ' TABLES LABELED AS: RATE#1 = ', A20,
           +         1X, ' RATE#2 = ', A20 // )

            IF (  NOLAB )
           +   WRITE ( 6, 9450 )
 9450       FORMAT ( T01, 'TABLES ARE UNLABELED'/
           +         1X,  'XXXXXX XXX XXXXXXXXX'// )

            IF ( JACKNF ) THEN
               WRITE(6,9500)
 9500          FORMAT(T01,1X,'NO.', T20, 'STRATUM', T55,
           +        'KAPPA', T45, 'KAPPA', T55, 'EXPECTED',
           +        T65, 'STANDARD ERROR', T80, '95 % C.I. ',
           +        T95, 'TOTAL COUNTS'/
           +        T45, 'MAX', T55, 'PROB',
           +        T65, 'LARGE SAMPLE', T80, 'LARGE SAMPLE',
           +        T95, 'IN TABLE'/ T55, 'CRUDE ADJ',
           +        T65, 'JACKKNIFE', T80, 'JACKKNIFE', T95, 'TABLE SIZE'/
           +        T65, 'NULL ASSOCN', T80, 'NULL ASSOCN'// )
               LINE = 10
            END IF
```

```
      IF ( .NOT. JACKNF ) THEN
         WRITE(6,9550)
 9550    FORMAT('1',1X,'NULL',T20, 'STRATUM',T35,
     +        'KAPPA',T45, 'KAPPA',T55, 'EXPECTED',
     +        T65, 'STANDARD ERROR',T80, '95 % C.I. ',
     +        T95, 'TOTAL COUNTS'/
     +        T45, 'MAX', T55, 'PROB',
     +        T65, 'LARGE SAMPLE', T80, 'LARGE SAMPLE',
     +        T95, 'IN TABLE'/ T55, 'CROSS TABS',
     +        T65, 'NULL ASSUMPT', T80, 'NULL ASSUMPT',
     +        T95, 'TABLE SIZE'// )
      LINE = 10
      END IF

      RETURN
      END
      SUBROUTINE PR ( NR1, NR2, LSTRAT, LINE, RATE1, RATE2, STRATA,
     +     KAPPA, NULAB, LRGSE, NULSE, KAPSUM, KAPCHI,
     +     C2, C1, C4, C3, PVAL, DF, NUMBER, UNDEF, NOVAR, OVERAL,
     +  KAPMAX, PL, PU, KAPAVG, KAPSMJ, VARJCK, PVALUJ, DFJ, KAPCHJ,
     +     SIZTBL, NOSUMR )

C
C     ROUTINE TO WRITE KAPPA STATS TO PRINTER
C

      IMPLICIT INTEGER ( A - Z )

      CHARACTER*20 RATE1(NR1), RATE2(NR2), STRATA(LSTRAT)
      REAL KAPPA(LSTRAT), LRGSE(LSTRAT), NULSE(LSTRAT),
     +     KAPSUM, KAPCHI, PVAL, C1, C2, DF, C3, C4,
     +     CIMIN1, CIMIN2, CIMIN3, CIMAX1, CIMAX2, CIMAX3,
     +     KAPMAX(LSTRAT), PC(LSTRAT), VARJCK(LSTRAT), KAPAVG,
     +     KAPSMJ, DFJ, KAPCHJ, PVALUJ, PU(LSTRAT)
      LOGICAL NULAB, NAMEL, AT, UNDEF(LSTRAT), NOVAR(LSTRAT),
     +        OVERAL, JACKNF, NOSUMR
      CHARACTER*20 NOVAL, ONEVAL
      DIMENSION NUMBER(LSTRAT), SIZTBL(LSTRAT)

      COMMON / LABS/ DIA, AT, NRATE1, NRATE2, NSTRAT, NAMEL,
     +               JACKNF
      DATA NOVAL / 'UNDEFINED          '/
      DATA ONEVAL /'PERFECT AGREEMENT  '/

C
C.... WRITE HEADINGS
C

      IF ( LINE .GE. 60 )
     +   CALL HEADR ( .FALSE., NULAB, NR1, NR2, RATE1, RATE2, LINE)

C
C.... LABELS PROCESSING
C

      IF ( .NOT. NULAB ) THEN
         DO 1000 K = 1, LSTRAT

            CIMIN1 = KAPPA(K) - 1.96 * LRGSE(K)
            CIMIN2 = KAPPA(K) - 1.96 * VARJCK(K)
            CIMIN3 = KAPPA(K) - 1.96 * NULSE(K)
            CIMAX1 = KAPPA(K) + 1.96 * LRGSE(K)
            CIMAX2 = KAPPA(K) + 1.96 * VARJCK(K)
            CIMAX3 = KAPPA(K) + 1.96 * NULSE(K)
```

```
                IF ( .NOT. UNDEF(K) .AND. JACKNF )
     +          WRITE ( 6, 9000 ) K, STRATA(K),
     +          KAPPA(K), KAPMAX(K), PC(K),
     +          LRGSE(K), CIMIN1, CIMAX1, NUMBER(K), PO(K),
     +          VARJCK(K), CIMIN2, CIMAX2, SIZTBL(K),
     +          NULSE(K), CIMIN3, CIMAX3

                IF ( .NOT. UNDEF(K) )
     +          WRITE ( 9, 9315 ) K, RATE1(NR1), RATE2(NR2),
     +          KAPPA(K), KAPMAX(K), PC(K),
     +          LRGSE(K), CIMIN1, CIMAX1, NUMBER(K), PO(K),
     +          VARJCK(K), CIMIN2, CIMAX2, SIZTBL(K),
     +          NULSE(K), CIMIN3, CIMAX3
                IF ( .NOT. UNDEF(K) .AND. .NOT. JACKNF )
     +          WRITE ( 6, 9020 ) K, STRATA(K),
     +          KAPPA(K), KAPMAX(K), PC(K),
     +          LRGSE(K), CIMIN1, CIMAX1, NUMBER(K), PO(K),
     +          NULSE(K), CIMIN3, CIMAX3, SIZTBL(K)
                IF ( UNDEF(K) .AND. KAPPA(K) .EQ. 0.)
     +          WRITE ( 6, 9010 ) K, STRATA(K),
     +                  NOVAL, PC(K), PO(K)
                IF ( UNDEF(K) .AND. KAPPA(K) .EQ. 1.)
     +          WRITE ( 6, 9010 ) K, STRATA(K),
     +                  ONEVAL, PC(K), PO(K)
                LIN6 = LIN6 + 2
                IF ( LIN6 .GE. 60 ) THEN
                   CALL HEAD6 ( .FALSE., NOLAB, NR1, NR2,
     +                          RATE1, RATE2, LIN6 )
                END IF
   1000     CONTINUE

            ELSE

C
C...  DO LARGLS
C

            DO 1200 K = 1, LSTRAT

                CIMIN1 = KAPPA(K) - 1.96 * LRGSE(K)
                CIMIN2 = KAPPA(K) - 1.96 * VARJCK(K)
                CIMIN3 = KAPPA(K) - 1.96 * NULSE(K)
                CIMAX1 = KAPPA(K) + 1.96 * LRGSE(K)
                CIMAX2 = KAPPA(K) + 1.96 * VARJCK(K)
                CIMAX3 = KAPPA(K) + 1.96 * NULSE(K)

                IF ( .NOT. UNDEF(K) .AND. JACKNF )
     +          WRITE ( 6, 9100 ) K, KAPPA(K), KAPMAX(K), PC(K),
     +          LRGSE(K), CIMIN1, CIMAX1, NUMBER(K), PO(K),
     +          VARJCK(K), CIMIN2, CIMAX2, SIZTBL(K),
     +          NULSE(K), CIMIN3, CIMAX3

                IF ( .NOT. UNDEF(K) )
     +          WRITE ( 9, 9315 ) K, RATE1(NR1), RATE2(NR2),
     +          KAPPA(K), KAPMAX(K), PC(K),
     +          LRGSE(K), CIMIN1, CIMAX1, NUMBER(K), PO(K),
     +          VARJCK(K), CIMIN2, CIMAX2, SIZTBL(K),
     +          NULSE(K), CIMIN3, CIMAX3
```

```
                IF ( .NOT. UNDEF(K) .AND. .NOT. JACKNF )
     +           WRITE ( 6, 9120 ) K, KAPPA(K), KAPMAX(K), PC(K),
     +          IRGSF(K), CIMIN1, CIMAX1, NUMOFM(K), PO(K),
     +          POLSE(K), CIMIN3, CIMAX3, SIZIBL(K)
                IF ( UNDEF(K) .AND. KAPPA(K) .EQ. 0.)
     +          WRITE ( 6, 9110 ) K, NOVAL, PC(K), PO(K)
                IF ( UNDEF(K) .AND. KAPPA(K) .EQ. 1.)
     +          WRITE ( 6, 9110 ) K, ONEVAL, PC(K), PO(K)
                LINO = LINO + 2
                IF ( LINO .GE. 60 ) THEN
                   CALL HEADO ( .FALSE., NOLAB, NR1, NR2,
     +                          RATE1, RATE2, LINO )
                END IF
   1200      CONTINUE
             END IF

C
C     RETURN IF NO OF STRATA = 1
C

      IF ( LSTRAT .LE. 1 ) RETURN

C
C     FORMATS
C

 9000 FORMAT(I0I,1X,I3, T10, A20, T35, F7.5, I45, F7.5, I55, F7.5,
     +       I65, F8.5, I80, F6.3,
     +       1X, F6.3, I95, I5/ I55, F7.5,
     +       I65, F8.5, I80, F6.3, 1X, F6.3, I95, I5/
     +       I65, F8.5, I80, F6.3, 1X, F6.3 )

 9020 FORMAT(I0I,1X,I3, T10, A20, T35, F7.5, I45, F7.5, I55, F7.5,
     +       I65, F8.5, I80, F6.3,
     +       1X, F6.3, I95, I5/ I55, F7.5,
     +       I65, F8.5, I80, F6.3, 1X, F6.3, I95, I5 )
 9010 FORMAT(I0I, 1X, I3, T10, A20, T35, A20, I55, F7.5/ I55, F7.5)

 9100 FORMAT(I0I,1X, I3, T35, F7.5, I45, F7.5, I55, F7.5,
     +       I65, F8.5, I80, F6.3, 1X,
     +       F6.3, I95, I5 / I55, F7.5,
     +       I65, F8.5, I80, F6.3, 1X, F6.3, I95, I5 /
     +       I65, F8.5, I80, F6.3, 1X, F6.3 )

 9120 FORMAT(I0I,1X, I3, I35, F7.5, I45, F7.5, I55, F7.5,
     +       I65, F8.5, I80, F6.3, 1X,
     +       F6.3, I95, I5 / I55, F7.5,
     +       I65, F8.5, I80, F6.3, 1X, F6.3, I95, I5 )
 9110 FORMAT ( I0I, 1X, I3, I35, A20 , I55, F7.5/ I55, F7.5)

 9315 FORMAT(1X,I3, 1X, A20, 1X, A20, 1X, F7.5, 1X, F7.5, 1X, F7.5,
     +       1X, F8.5, 1X, F6.3,
     +       1X, F6.3, 1X, I5/1X, F7.5,
     +       1X, F8.5, 1X, F6.3, 1X, F6.3, 1X, I5,
     +       1X, F8.5, 1X, F6.3, 1X, F6.3 )

      IF ( NOSUMM ) RETURN
```

```
C
C
C     SUMMARY STATS ROUTINE
C
C
      ENTRY WGSUM ( ISTRAT, LINE, KAPSUM, KAPCHI, C2, C1,
     +              C4, C3, PVAL, DF, OVERAL, KAPAVG,
     +              KAPSMJ, PVALUJ, DFJ, KAPCHJ, NOSUMR )

      IF ( NOSUMR ) RETURN
      IF ( OVERAL )
     +   WRITE ( 6, 9300 )
 9300    FORMAT ( 1H1, 115, 'O V E R A L L   K A P P A   S U M M E U',
     +                  'T   O V E R   A L L   T A B L E S'/// )

      WRITE (6, 9200) ISTRAT, KAPAVG, KAPSUM, KAPCHI, PVAL, DF, C1, C2
 9200 FORMAT (////' S U M M A R Y   K A P P A   S T A T S'/
     +         ' = = = = = = = =   = = = = =   = = = = =='//
     +         ' NO. OF STRATA ..................... ', I4/
     +         ' (NOTE THAT NOT ALL STRATA ARE USED IF SOME KAPPAS',
     +         ' WERE UNDEFINED)' //
     +         ' SUMMARY KAPPA (UNWEIGHTED) .... ', F5.3//
     +         ' LARGE SAMPLE VARIANCES          '/
     +         ' ===== ====== ==========          '//
     +         ' SUMMARY KAPPA (WEIGHTED) ...... ', F5.3//
     +         ' CHI SQ FOR HOMOGENEITY ........ ', F10.4/
     +         ' (H0: STRATA ARE HOMOGENEOUS )' //
     +         ' PVALUE FOR CHI SQUARE ......... ', F7.5,' ( DF = ',
     +         F6.0, ' )'//
     +         ' 95% C.I. FOR OVERALL KAPPA .... ',F5.3,' TO ',F5.3//)
      LINE = LINE+17

      IF ( JACKNF ) THEN
      WRITE (6, 9400) KAPSMJ, KAPCHJ, PVALUJ, DFJ, C3, C4,
 9400 FORMAT (' JACKKNIFED VARIANCES'/
     +         ' ========== ==========='//
     +         ' SUMMARY KAPPA (WEIGHTED) ...... ', F5.3//
     +         ' CHI SQ FOR HOMOGENEITY ........ ', F10.4/
     +         ' (H0: STRATA ARE HOMOGENEOUS )' //
     +         ' PVALUE FOR CHI SQUARE ......... ', F7.5,' ( DF = ',
     +         F6.0, ' )'//
     +         ' 95% C.I. FOR OVERALL KAPPA .... ',F5.3,' TO ',F5.3//)
      LINE = LINE+12
      END IF

      RETURN
      END
      SUBROUTINE KAPPAC ( MODE, KAPPA, KAPMAX, PC, PO, LRGSE, NULSE,
     +                    SUMMER, UNDEF, TABSIZ, TABLE, WEIGHT, NOVAR,
     +                    VARJCK )

C
C     REFER TO J. FLEISS, 'STAT METHODS FOR RATES & PROPORTIONS',
C        CHAP 13, (2ND EDITION) FOR FORMULAE
C
C     NOTE THAT NO WEIGHTS REQUIRES A DIAGONAL MATRIX
C        SUCH THAT ALL ELEMENTS (J,J)= 1 AND (J,K) = 0
C     THIS IS THE DEFAULT
C
```

```
      IMPLICIT INTEGER ( A - Z )
      REAL KAPPA, LRGST, NULST, TABLE(30,30),
     +     WEIGHT(30,30)
      REAL PO, PC, PROW(30), PCOL(30), WTROW(30), WTCOL(30)
      REAL A, B, C, D, E, KAPMAX, POM, VARJCK
      LOGICAL UNDEF, NOVAR, NAMEL, WT, JACKNF
      REAL ALMZER

      COMMON / MAX / MAXTAB, MAXSIR, MAXARY
      COMMON / LABS / DIM, WT, NRATE1, NRATE2, NSTRAT, NAMEL,
     +                JACKNF

      DATA ALMZER / 1.E-12 /

C.... INITIALIZE ACCORDING TO CALCULATION MODE

      IF ( MODE .LE. 2 ) THEN
         CALL ZERO ( PROW, MAXARY )
         CALL ZERO ( PCOL, MAXARY )
         CALL ZERO ( WTRO.., MAXARY )
         CALL ZERO ( WTCOL, MAXARY )
         PO = 0
         PC = 0
      END IF

C
C.... COUNT TOTAL NO. OF OBSERVATIONS
C

      NUMBER = 0
      DO 100 K = 1, TABSIZ
         DO 100 J = 1, TABSIZ
            NUMBER = NUMBER + TABLE(K,J)
  100 CONTINUE

C.... CHECK FOR AN EMPTY TABLE

      IF ( NUMBER .EQ. 0 ) THEN
         UNDEF = .TRUE.
         LRGST = 0.
         NULST = 0.
         GO TO 8000
      END IF

C
C.... KAPPA
C

C     WRITE(6,*)  ((TABLE(I,M),L=1,4),M=1,4)
      CALL CALCNK ( TABLE, WEIGHT, TABSIZ, NUMBER, PROW, PCOL, PO,
     +              PC, KAPPA, UNDEF )
      IF ( UNDEF ) THEN
         LRGST = 0.
         NULST = 0.
         VARJCK = 0.
         GO TO 8000
      END IF
```

```
C
C.... KAPPA MAXIMUM
C

      PON = 0.
      DO 650 K = 1, TABSIZ
         A = AMIN1 ( PROW(K), PCOL(K) )
         PON = PON + A
  650 CONTINUE

      KAPMAX = ( PON - PC ) / ( 1. - PC )

C.... DEBUG
C     WRITE ( 6, * ) PO, PC, KAPPA, (PROW(J),J=1,5), (PCOL(K),K=1,5)
C
C.... WEIGHTED AVG OF WEIGHTS
C.... WTROW = W(I.)
C.... WTCOL = W(.J)
C

      DO 700 J = 1 ,TABSIZ
         DO 700 K = 1 ,TABSIZ
         WTROW(J) = WTROW(J) + WEIGHT(J,K)*PCOL(K)
  700 CONTINUE

      DO 750 J = 1 ,TABSIZ
         DO 750 K = 1 ,TABSIZ
         WTCOL(J) = WTCOL(J) + WEIGHT(K,J)*PROW(K)
  750 CONTINUE
C.... DEBUG
C     WRITE ( 6, * ) (WTROW(K),K=1,5), (WTCOL(J),J=1,5)

C
C.... STANDARD ERRORS
C

C.... LARGE SAMPLE STD ERRORS

      C = 0.
      DO 800 J = 1, TABSIZ
         DO 800 K = 1, TABSIZ
         A = WEIGHT(J,K) - (WTROW(J) +WTCOL(K))*(1. - KAPPA)
         A = TABLE(J,K) *A * A
         C = C + A
  800 CONTINUE
      B = KAPPA - PC * (1. - KAPPA )
      C = C - B * B
      IF ( C .LE. ALMZER ) C = 0.

      ERRSE  = SQRT (C / (NUMBER*(1.-PC)**2) )
```

```
C
C... SE FOR NO ASSOCN
C

      D = 0.
      DO 900 J = 1, TABSIZ
         DO 900 K = 1, TABSIZ
         E = (WEIGHT(J,K) - (WTCOL(K) * WTROW(J)))**2
         E = E*PROW(J) * PCOL(K)
         D = D + E
  900 CONTINUE
      D = D - PL * PC
      IF (D .LE. ALMZER ) D = 0.

      NULSE  = SQRT (D / (NUMBER* ( 1. -PL) **2 ) )

C.... DEBUG
C        WRITE ( 6, * ) KAPPA, PD, P, A, B, C, D, E, LRGSE, NULSE
C
C
C.... JACKKNIFE
C

      IF ( JACKNF )
     +    CALL JCKNIF ( NUMBER, KAPPA, TABSIZ, TABLE, WEIGHT,
     +                  NOVAR, VARJCK )

C
C**** RETURN
C

 8000 CONTINUE
      RETURN
      END
      SUBROUTINE ZERO ( TABLE, N )
      REAL TABLE(N)

      DO 100 K = 1, N
         TABLE(K) = 0.
  100 CONTINUE

      RETURN
      END
      SUBROUTINE BLNK ( TABLE, N )
      CHARACTER TABLE(N)

      DO 100 K = 1, N
         TABLE(K) = ' '
  100 CONTINUE

      RETURN
      END
```

```
      SUBROUTINE KAPSUM ( KAPPA, LRGSE, UNDEF, NOVAR, KAPAVG,
     +                    MODE, NSTRAT, KAPSUM, CI, KAPCHI, PVALUE, DF, * )
C
C.... ROUTINE TO CALCULATE SUMMARY KAPPA
C.... SEE FLEISS, P221
C
C
      IMPLICIT INTEGER ( A - Z )

      REAL KAPPA(NSTRAT), LRGSE(NSTRAT), A, CI, KAPSUM, KAPCHI, C, D,
     +     PVALUE, DF, KAPAVG
      LOGICAL UNDEF(NSTRAT), NOVAR(NSTRAT)

      IF ( MODE .LE. 2) THEN
         A= 0.
         KAPAVG = 0.
         CI = 0.
         DO 1000 K = 1, NSTRAT
            KAPAVG = KAPAVG + KAPPA(K)
            IF (UNDEF(K).OR.NOVAR(K) .OR. LRGSE(K) .EQ. 0.) GO TO 1000
            C = LRGSE(K) * LRGSE(K)
            A = A + KAPPA(K) / C
            CI = CI + 1. / C
 1000    CONTINUE

         IF ( A .EQ. 0. .OR. CI .EQ. 0. ) RETURN 1
         KAPSUM = A / CI
         KAPAVG = KAPAVG / NSTRAT

      END IF

C
C.... CALCULATION OF CHI SQUARE FOR HOMOGENEITY
C

      KAPCHI = 0.
      NOVALS = 0

      IF ( MODE .LE. 2) THEN
         DO 2000 K = 1, NSTRAT
            IF (UNDEF(K).OR.NOVAR(K) .OR. LRGSE(K) .EQ. 0.) GO TO 2000
            C = (KAPPA(K) - KAPSUM) ** 2
            D = 1. / (LRGSE(K) * LRGSE(K) )
            KAPCHI = KAPCHI + C*D
            NOVALS = NOVALS + 1
 2000    CONTINUE

C.... PVALUE

      DF = NOVALS - 1
      CALL MDCH ( KAPCHI, DF, PVALUE, IER )
      PVALUE = 1. - PVALUE

      END IF

C
C**** RETURN
C

      RETURN
      END
```

```
      SUBROUTINE CALCHI ( TABLE, WEIGHT, TABSIZ, NUMBER, PROW, PCOL,
     +                    PO, PC, KAPPA, UNDEF )

      IMPLICIT INTEGER ( A - Z )
      REAL TABLE(30,30), PCOL(TABSIZ), PROW(TABSIZ),
     +     WEIGHT(30, 30)
      REAL PO, PC, KAPPA, ALMZER
      LOGICAL UNDEF

      COMMON / MAX / MAXTAB, MAXSIZ, MAXARY

      DATA ALMZER / 1.E-12 /
      UNDEF = .FALSE.

C
C.... CALCULATE PROPORTIONS
C

      DO 340 K = 1, TABSIZ
         DO 340 J = 1, TABSIZ
            TABLE(K,J) = (TABLE(K,J)/NUMBER)
  340 CONTINUE
C.... DEBUG
C     WRITE ( 6, * ) ((TABLE(J,K),K=1,5),J=1,5)


C     ENTRY FOR TACHNITEF ROUTINE ONLY

      ENTRY CALNKJ ( TABLE, WEIGHT, TABSIZ, NUMBER, PROW, PCOL,
     +               PO, PC, KAPPA, UNDEF )

C.... INITIALIZE ACCORDING TO CALCULATION MODE

      CALL ZERO ( PROW, MAXARY )
      CALL ZERO ( PCOL, MAXARY )
      PO = 0
      PC = 0

C
C.... PROW, PCOL
C.... PROW = P(I,.) = SUM(J) (P(IJ))
C.... PCOL = P(.,J) = SUM(I) (P(IJ))
C

      DO 400 J = 1, TABSIZ
         DO 400 K = 1, TABSIZ
         PROW(J) = PROW(J) + TABLE(J,K)
  400 CONTINUE

      DO 450 J = 1, TABSIZ
         DO 450 K = 1, TABSIZ
         PCOL(J) = PCOL(J) + TABLE(K,J)
  450 CONTINUE

C
C.... PO
C

      DO 500 J =1, TABSIZ
         DO 500 K =1, TABSIZ
         PO = PO + TABLE(J,K)*WEIGHT(J,K)
  500 CONTINUE
```

```
C
C.... PL
C
      DO 600 J = 1, TABSIZ
         DO 600 K = 1, TABSIZ
            PL = PL + WEIGHT(J,K)*PROW(J)*PCOL(K)
  600 CONTINUE
C
C.... KAPPA
C
      IF ( ABS ( 1. - PC ) .LE. ALMZER .AND.
     +     ABS ( PO - PC ) .LE. ALMZER ) THEN
         UNDEF = .TRUE.
         KAPPA = 1.
         GOTO 8000
      END IF
      IF ( ABS ( 1. - PL ) .LE. ALMZER ) THEN
         UNDEF = .TRUE.
         KAPPA = 0.
         GOTO 8000
      END IF

      KAPPA = ( PO - PC ) / ( 1. - PC )

 8000 CONTINUE

C.... DEBUG
C     WRITE(6,*) KAPPA, PO, PC , ((TABLE(L,M),L=1,TABSIZ),M=1,TABSIZ)
      RETURN
      END
      SUBROUTINE JCKNIF ( NUMBER, KAPPA, TABSIZ, TABLE, WEIGHT,
     +                    UNDEF, VARJCK )

C
C     REF: MILLER, BIOMETRIKA 61:1-15,1974 FOR A REVIEW
C          FLEISS AND DAVIS AJE 115:841-845,1982 FOR AN EXAMPLE
C          AND EFRON, BIOMETRIKA 68:589-599,1981 FOR A COMPARISON
C              OF METHODS
C
C     ON 4 BYTE WORD MACHINES (EG, IBM) THIS ROUTINE SHOULD
C     BE USED IN DOUBLE PRECISION MODE
C
      IMPLICIT INTEGER ( A - Z )

      REAL TABLE ( 30, 30 ), PSEUDO ( 900 ), A
      REAL WEIGHT( 30, 30 )
      REAL JACKNF, PROW(30), PCOL(30), PO, PC, VARJCK, KAPPA
      REAL WORK(900)
      LOGICAL UNDEF, INCFLG
      UNDEF = .FALSE.
      VARJCK = 0.
      INCFLG = .FALSE.
      IF ( NUMBER .LE. 30 ) INCFLG = .TRUE.
```

```
      DO 8000 J = 1, TABSIZ
         DO 8000 K = 1, TABSIZ

         NPSEUD = (J-1) * TABSIZ + K
         NOCEL(NPSEUD) = TABLE(J,K) * NUMBER
         IF ( TWOCLG ) NOCEL(NPSEUD) = NOCEL(NPSEUD) + 0.5

C....    CONVERT TABLE TO NEW PROPORTIONS

         IF ( NOCEL(NPSEUD) .LE. 0. ) THEN
            PSEUDO(NPSEUD) = KAPPA
            GO TO 5000
         END IF
         CALL CNVDWN ( TABLE, TABSIZ, J, K, NUMBER )
         CALL CALNKJ(TABLE, WEIGHT, TABSIZ, NUMBER-1, PROW, PCOL, PO,
     +                PL, PSEUDO(NPSEUD), UNDEF )
      A = PSEUDO(NPSEUD)

C....    RETURN TABLE TO ITS NATURAL STATE

C.... DEBUG/
C     WRITE ( 6, * )  ((TABLE(L,M),M=1,5),L=1,5)
         CALL CNVUP ( TABLE, TABSIZ, J, K, NUMBER )

C.... DEBUG
C     WRITE ( 6, * )  ((TABLE(L,M),M=1,5),L=1,5)

         IF ( UNDEF )  GO TO 9000

 5000    CONTINUE
         PSEUDO(NPSEUD) = NUMBER*KAPPA - (NUMBER-1.)*PSEUDO(NPSEUD)
C.... DEBUG
C     WRITE(6,*) NPSEUD, NOCEL(NPSEUD), A, PSEUDO(NPSEUD), TABLE(J,K),
C    +           NUMBER, KAPPA
 6000 CONTINUE

C.... CALCULATE JACKKNIFE ESTIMATE OF KAPPA

      JACKNF = 0.
      DO 7000 K = 1, TABSIZ * TABSIZ
         JACKNF = JACKNF + NOCEL(K) * PSEUDO(K)
 7000 CONTINUE

      JACKNF = JACKNF / NUMBER

C.... CALCULATE VARIANCE ESTIMATE OF JACKKNIFED ESTIMATE

      VARJCK = 0.
      DO 8000 K = 1, TABSIZ * TABSIZ
         A = (PSEUDO(K) - JACKNF )
         A = A * A * NOCEL(K)
         VARJCK = VARJCK + A
 8000 CONTINUE

      VARJCK = VARJCK / (NUMBER*(NUMBER-1.) )
      VARJCK = SQRT(VARJCK)

C.... DEBUG
C     WRITE ( 6, * ) JACKNF, VARJCK
 9000 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE CNVDWN ( TABLE, TABSIZ, ROW, COL, NUMBER )

C
C     ROUTINE TO CONVERT PROPORTIONS IN TABLES
C     USED IN JACKKNIFE PROCEDURE
C

      IMPLICIT INTEGER ( A - Z )

      REAL TABLE ( 30, 30 )

C.... CONVERT TO N-1 TABLE

      TABLE(ROW,COL) = (TABLE(ROW,COL)*NUMBER - 1.)/NUMBER

      DO 1000 J = 1, TABSIZ
         DO 1000 K = 1, TABSIZ
         TABLE(J,K) = (TABLE(J,K)*NUMBER)/(NUMBER - 1.)
 1000 CONTINUE

      RETURN

C.... CONVERT TO ORIGINAL TABLE

      ENTRY CNVUP ( TABLE, TABSIZ, ROW, COL, NUMBER )

      DO 2000 J = 1, TABSIZ
         DO 2000 K = 1, TABSIZ
         TABLE(J,K) = (TABLE(J,K)*(NUMBER-1.))/NUMBER
 2000    CONTINUE
      TABLE(ROW,COL) = (TABLE(ROW,COL)*NUMBER + 1.)/NUMBER
      RETURN
      END
      BLOCK DATA
      IMPLICIT INTEGER ( A - Z )
      COMMON / I / H, TITLE
      COMMON / LABS / DIM, MT, NRATE1, NRATE2, NSTRAT, NAMEL,
     +                JACKNF
      LOGICAL MT, NAMEL, JACKNF
      CHARACTER*10 TITLE(8), H(2)
      DATA NRATE1, NRATE2, NSTRAT, DIM / 3*1, 0 /
      DATA MT / .FALSE. /, JACKNF / .TRUE. /
      DATA TITLE, H / 10 * ' ' , ' ' /
      END
```

Appendix 7

Supplementary Tables for the General Comparison Trial

Table A.7.1

Occupations, Industries, and Time Periods of Employment for Job Descriptions used in the General Comparison Trial

Occupation/ Industry Description	Standard Industrial Code(1)	Occupation Code (2)	Time Period of Employment
Leather worker/ Small leather goods	179	8569-158	1932-34
Tar chaser/ Chemical mfg	378	8178-214	1934-38
Enlisted serviceman/ Armed Forces	902	6117-999	1938-47
Spray Painter/ Railway Transport	503	8595-126	1947-77
Drill Press Operator/ Motor Vehicle mfg	323	8315-150	1942-45
Kitchen helper/ Food wholesale	614	6198-134	1945-47
Leather worker/ Small Leather Goods	179	8569-294	1947-51
Office Clerk/ Shipbuilding	327	4197-160	1951-72

Table A.7.1, continued

Occupation/ Industry Description	Standard Industrial Code(1)	Occupation Code (2)	Time Period of Employment
Management, NEC/ University, College	806	1179-299	1972-79
Construction Labourer/ Building Construction	404	8798-114	1934-35
Motor Vehicle Maint./ Motor Vehicle Repair	658	8581-114	1935-39
Airframe Assembler/ Aircraft Mfg	321	8515-118	1939-45
Patternmaker/ Aircraft Mfg	321	8395-299	1945-74
Cook/ Hospital	821	6121-134	1968-72
Labourer, Concrete/ Other construction	409	8718-114	1972-79

Table A.7.1, continued

Occupation/ Industry Description	Standard Industrial Code(1)	Occupation Code (2)	Time Period of Employment
Wood Worker/ Hospital	821	8541-126	1979-80
Labourer/ Scrap Metal Wholesale	627	9918-199	1980-80
Fish Cleaner/ Fish Products	102	8217-130	1926-59
Construction Labourer/ Building Construction	404	8798-114	1959-66
Fish Cleaner/ Fish Products	102	8217-130	1967-73

(1) Refer to SIC, 1970.

(2) Refer to OCC, 1971.

Table A.7.2

Statistical Tests of Equality of the Proportion of Exposure Attributed Present at a Given Level of Confidence between all Raters in the General Comparison Trial (1)

Rater Pair	Chi-Square Value for Exposure Categorized at Confidence Level (2): High	Medium or High	Any
Chem 1 - Chem 2	6.8**	29.8**	29.4**
Chem 1 - Chem 4	1.6	33.6**	45.6**
Chem 1 - Chem 6	3.9*	1.7	3.1
Chem 2 - Chem 4	1.8	0.4	2.7
Chem 2 - Chem 6	0.2	16.4**	13.7**
Chem 4 - Chem 6	0.5	20.6**	25.4**
Overall (3)	7.3	53.4**	62.3**

Table A.7.2, continued

(1) A total of 3040 exposures = 20 job descriptions x 172 substances were used in the calculations.

(2) The tests in the first part of the table are derived using McNemar's Test (1 df). Note that: * indicates $p < .05$; ** indicates $p < .01$.

(3) This test was evaluated using Cochran's Q statistic on 3 df.

Table A.7.3

Inter-Rater Agreement using the Collapsed Table Method for the Four Exposure Variables and the Synthetic Index in the General Comparison Trial (1)

Rater Pair	Contact			Frequency			Concentration			Level of Confidence			Synthetic Index (2)		
	P_o %	κ	κ'	P_o %	κ	κ'	P_o %	κ	κ'	P_o %	κ	κ'	P_o %	κ	κ'
Chem 1 - Chem 2	95.9	0.49	0.65	95.4	0.42	0.55	95.2	0.40	0.52	95.7	0.46	0.65	95.5	0.44	0.52
Chem 1 - Chem 4	95.9	0.46	0.66	95.5	0.41	0.58	95.4	0.39	0.55	95.9	0.46	0.60	95.7	0.43	0.61
Chem 1 - Chem 6	94.4	0.38	0.57	94.3	0.37	0.50	94.1	0.35	0.42	94.9	0.43	0.46	94.7	0.42	0.50
Chem 2 - Chem 4	97.6	0.58	0.62	97.2	0.52	0.61	97.0	0.49	0.56	97.6	0.59	0.67	97.4	0.55	0.66
Chem 2 - Chem 6	95.4	0.38	0.47	95.8	0.44	0.56	95.6	0.41	0.50	95.9	0.45	0.54	95.9	0.44	0.57
Chem 4 - Chem 6	95.7	0.39	0.59	95.6	0.37	0.48	95.4	0.34	0.46	96.1	0.45	0.60	95.8	0.40	0.56
Average (3):	95.8	0.45	0.59	95.6	0.42	0.55	95.5	0.40	0.50	96.0	0.47	0.59	95.8	0.45	0.57
Average (4):	96.4	0.51	0.64	96.0	0.45	0.58	95.9	0.43	0.54	96.4	0.50	0.64	96.2	0.47	0.60

(1) A total of 3440 exposures = 20 job descriptions x 172 substances were used on the calculations.
The original scales of measurement (four categories) were used for each of the four exposure variables.

(2) The synthetic index is defined as frequency x concentration x level of confidence, grouped into 3 categories: no , "medium" and "high" exposure.

(3) This average includes the comparisons with Chemist 6.

(4) This average excludes the comparisons with Chemist 6.

Table A.7.4

Frequency of Exposure for all Attributed Substances in the General Comparison Trial (1)

Substance	Prevalance of Exposure as Attributed by: Chem 1 %	Chem 2 %	Chem 4 %	Chem 6 %	Average %
Abrasive dust	30	20	15	25	22.5
Insulating material dust	5	10	5	5	6.3
Construction dust	15	15	15	20	16.3
Mine dust	0	10	10	0	5.0
Metallic dust	25	35	20	20	25.0
Asbestos (Chrysotile)	15	15	15	5	12.5
Asbestos (Amphibole)	15	15	5	5	10.0
Silica dust	10	10	25	20	16.3
Cement dust	15	15	10	10	12.5
Glass fibres	10	5	5	5	6.3
Brick dust	0	10	0	5	3.8
Silicon carbide	5	5	0	15	6.3
Sulfur	0	0	0	5	1.3
Potassium nitrate	0	0	0	5	1.3
Gypsum	10	5	5	10	7.5
Titanium dioxide	10	5	0	15	7.5
Iron oxides	10	5	15	10	10.0
Floor dust	0	0	0	40	10.0
Lead dust	5	0	0	0	1.3
Sodium chloride	0	5	5	5	3.8
Coal dust	5	5	5	10	6.3
Carbon dust	5	5	0	5	3.8
Rubber dust	0	0	0	5	1.3
Fabric dust	5	5	0	0	2.5
Plastic dust	10	5	5	5	6.3
Wool fibres	0	5	0	0	1.3
Wood dusts	20	20	15	20	18.8
Flour dusts	5	0	5	5	3.8
Fur dust	0	5	10	0	3.8
Leather dust	0	0	0	10	2.5
Carbon dust	5	0	0	0	1.3
Coal gas	0	0	0	5	1.3

Table A-7.4, continued

Substance	Prevalence of Exposure as Attributed by: Chem 1 %	Chem 2 %	Chem 4 %	Chem 6 %	Average %
Carbon monoxide	5	5	0	20	7.5
Nitrogen oxides	0	0	0	15	3.8
Sulfur dioxide	0	0	0	10	2.5
Spray gases	0	5	0	0	1.3
Natural gas	0	5	0	0	1.3
Propane	0	5	5	0	2.5
Formaldehyde	10	0	0	0	2.5
Vinyl chloride	5	0	0	0	1.3
Pyrolysis and combustion fumes	15	10	10	5	10.0
Cooking fumes	5	5	5	10	6.3
Engine emissions	20	20	20	15	18.8
Combustion products of wood	0	0	0	5	1.3
Gas welding fumes	5	5	5	15	7.5
Arc welding fumes	10	0	0	10	5.0
Metal oxide fumes	15	15	10	10	12.5
Lead fumes	5	5	0	0	2.5
Inorganic acid solution	5	0	0	0	1.3
Caustic solution	5	0	0	0	1.3
Paints and varnishes	15	10	10	20	13.8
Wood stains and varnishes	10	10	10	10	10.0
Dyes and pigments	5	10	5	5	6.3
Organic dyes and pigments	10	5	0	5	5.0
Adhesives	15	15	10	5	11.3
Solvents	45	25	20	30	30.0
Waxes and polishes	10	0	0	5	3.8
Gasoline	5	5	10	15	8.8
Diesel oil	0	0	0	5	1.3
Mineral spirits	20	10	15	15	15.0
Aromatic napthas	5	5	0	5	3.8
Lubricating oils and grease	30	15	20	20	21.3
Cutting fluids	5	10	5	0	5.0
Coal tar and pitch	5	5	5	5	5.0
Creosote	5	0	0	0	1.3
Pitch	5	0	0	0	1.3
Glues	0	0	0	5	1.3

Table A.7.4, continued

Substance	Prevalence of Exposure as Attributed by: Chem 1 %	Chem 2 %	Chem 4 %	Chem 6 %	Average %
Turpentine	10	5	0	10	6.3
Linseed oil	10	10	5	10	8.8
Methanol	10	5	0	0	3.8
Ethanol	5	0	0	0	1.3
Ethyleneglycol	5	0	0	0	1.3
Carbon tetrachloride	5	0	0	0	1.3
Methylene chloride	10	0	5	10	6.3
Tetraethyl lead	0	0	0	5	1.3
Trichloroethylene	5	0	0	0	1.3
Benzene	25	0	0	0	6.3
Toluene	15	0	0	0	3.8
Xylene	5	0	0	0	1.3
Phenol	10	0	0	0	2.5
Benzidine	10	0	0	0	2.5
Naphtylamine	15	0	0	0	3.8
O-toluidine	10	0	0	0	2.5
Aluminum compounds	0	0	5	5	2.5
Chrome compounds	5	0	0	0	1.3
Iron compounds	5	10	0	5	5.0
Zinc compounds	5	0	0	0	1.3
Cadmium compounds	5	0	0	0	1.3
Lead compounds	15	5	5	0	6.3
Aliphatic hydrocarbons	20	0	15	25	15.0
Aliphatic alcohols	10	5	0	0	3.8
Aliphatic acids	10	0	0	0	2.5
Aliphatic aldehydes	5	0	0	0	1.3
Aliphatic saturated halogens	10	0	0	5	3.8
Aliphatic unsaturated halogens	5	0	0	0	1.3
Aromatic hydrocarbons	25	0	5	20	12.5
Aromatic alcohols	5	0	0	0	1.3
Rubber	10	5	10	5	7.5
Plastics	5	5	10	5	6.3
Leather	10	10	10	10	10.0
Foodstuff	15	10	15	15	13.8

Table A.7.4, continued

Substance	Prevalence of Exposure as Attributed by: Chem 1 %	Chem 2 %	Chem 4 %	Chem 6 %	Average %
Cleaning agents	10	15	5	10	10.0
Pharmaceuticals	0	5	0	0	1.3
Non-specific dust	0	5	20	5	7.5
Metal oxide dust	10	15	15	25	16.3

(1) The table is based on the assessment of exposure to 20 job descriptions using a dichotomous scale: exposed, at any level of confidence, and not exposed.

Table A.7.5

Inter-Rater Agreement for all Attributed Substances in the General Comparison Trial (1)

Substance	Chem 1–Chem 2			Chem 1–Chem 4			Chem 1–Chem 6			Chem 2–Chem 4			Chem 2–Chem 6			Chem 4–Chem 6			Average	
	P_o %	κ	κ' (2)	P_o %	κ	κ' (2)	P_o %	κ	κ' (2)	P_o %	κ	κ' (2)	P_o %	κ	κ' (2)	P_o %	κ	κ' (2)	P_o %	κ
Abrasive dust	90.0	0.24	1.0	85.0	0.58	1.0	95.0	0.88	1.0	95.0	0.83	1.0	95.0	0.86	1.0	90.0	0.69	1.0	91.0	0.76
Insulating material dust	95.0	0.64	1.0	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.64	1.0	95.0	0.64	1.0	100.0	1.0	1.0	97.5	0.82
Construction dust	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.83	1.0	100.0	1.0	1.0	95.0	0.83	1.0	95.0	0.83	1.0	97.5	0.92
Mine dust	90.0	0.0	u	90.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	90.0	0.0	u	90.0	0.0	u	93.3	0.33
Metallic dust	90.0	0.77	1.0	75.0	0.29	0.34	85.0	0.67	1.0	75.0	0.39	0.62	95.0	0.89	1.0	80.0	0.55	1.0	83.3	0.59
Asbestos (Chrysotile)	80.0	0.22	0.22	80.0	0.22	0.27	90.0	0.46	1.0	90.0	0.61	0.61	90.0	0.46	1.0	90.0	0.46	1.0	86.7	0.41
Asbestos (Amphibole)	80.0	0.22	0.22	80.0	−0.08	0.17	90.0	0.46	1.0	90.0	0.46	1.0	90.0	0.46	1.0	90.0	−0.05	−0.05	86.7	0.25
Silica dust	90.0	0.44	0.44	75.0	0.17	0.34	90.0	0.62	1.0	90.0	0.62	1.0	85.0	0.50	1.0	85.0	0.57	0.66	84.2	0.42
Cement dust	100.0	1.0	1.0	95.0	0.77	1.0	95.0	0.78	1.0	95.0	0.77	1.0	95.0	0.77	1.0	90.0	0.44	0.44	95.0	0.76
Glass fibres	95.0	0.64	1.0	95.0	0.64	1.0	95.0	0.64	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.82
Brick dust	90.0	0.0	u	100.0	1.0	1.0	95.0	0.0	u	90.0	0.0	u	95.0	0.64	1.0	95.0	0.0	u	94.2	0.27
Silicon carbide	100.0	1.0	1.0	95.0	0.0	u	90.0	0.46	1.0	95.0	0.0	u	90.0	0.46	1.0	85.0	0.0	u	92.5	0.32
Sulfur	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.0	u	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	97.5	0.50
Potassium nitrate	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.0	u	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	97.5	0.50
Gypsum	95.0	0.64	1.0	95.0	0.64	1.0	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.64	1.0	95.0	0.64	1.0	96.7	0.76
Titanium dioxide	95.0	0.64	1.0	90.0	0.0	u	95.0	0.78	1.0	95.0	0.0	u	90.0	0.46	1.0	85.0	0.0	u	91.7	0.31
Iron oxides	95.0	0.64	1.0	95.0	0.72	1.0	90.0	0.44	0.44	90.0	0.46	1.0	85.0	−0.07	−0.11	85.0	0.32	0.42	90.0	0.43
Floor dust	100.0	1.0	1.0	100.0	1.0	1.0	60.0	0.0	u	100.0	1.0	1.0	60.0	0.0	u	60.0	0.0	u	80.0	0.50
Lead dust	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	0.0	u	97.5	0.33
Sodium chloride	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.50
Coal dust	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.0	u	100.0	1.0	1.0	95.0	0.64	1.0	95.0	0.64	1.0	97.5	0.71
Carbon dust	100.0	1.0	1.0	95.0	0.0	u	100.0	1.0	1.0	95.0	0.0	0.04	100.0	1.0	1.0	95.0	0.0	u	97.5	0.50
Rubber dust	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.0	u	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	97.5	0.50
Fabric dust	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	96.7	0.33
Plastic dust	95.0	0.64	1.0	95.0	0.64	1.0	95.0	0.0	u	100.0	1.0	1.0	90.0	−0.05	−0.05	90.0	−0.05	−0.05	94.2	0.36
Wool fibres	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	97.5	0.50
Wood dusts	100.0	1.0	1.0	95.0	0.83	1.0	100.0	1.0	1.0	95.0	0.83	1.0	100.0	1.0	1.0	95.0	0.83	1.0	97.5	0.92
Flour dusts	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	97.5	0.50
Fur dust	95.0	0.0	u	90.0	0.0	u	100.0	1.0	1.0	95.0	0.64	1.0	95.0	0.0	u	90.0	0.0	u	94.2	0.27
Leather dust	100.0	1.0	1.0	100.0	1.0	1.0	90.0	0.0	u	100.0	1.0	1.0	90.0	0.0	u	90.0	0.0	u	95.0	0.50
Carbon dust	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.50
Coal gas	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.0	u	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	97.5	0.50

Table A.7.5, continued

Substance	Chem 1–Chem 2			Chem 1–Chem 4			Chem 1–Chem 6			Chem 2–Chem 4			Chem 2–Chem 6			Chem 4–Chem 6			Average	
	p_o %	κ	κ' (2)	p_o %	κ	κ' (2)	p_o %	κ	κ' (2)	p_o %	κ	κ' (2)	p_o %	κ	κ' (2)	p_o %	κ	κ' (2)	p_o %	κ
Carbon monoxide	100.0	1.0	1.0	95.0	0.0	u	85.0	0.35	1.0	95.0	0.0	u	85.0	0.35	1.0	80.0	0.08	u	90.0	0.28
Nitrogen oxides	100.0	1.0	1.0	100.0	1.0	1.0	85.0	0.0	u	100.0	1.0	1.0	85.0	0.0	u	85.0	0.0	u	92.5	0.50
Sulfur dioxide	100.0	1.0	1.0	100.0	1.0	1.0	90.0	0.0	u	100.0	1.0	1.0	90.0	0.0	u	90.0	0.0	u	95.0	0.50
Spray gases	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	97.5	0.50
Natural gas	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	97.5	0.50
Propane	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	96.7	0.33
Formaldehyde	95.0	0.0	u	90.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	96.7	0.50
Vinyl chloride	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.50
Pyrolysis and combustion fumes	85.0	0.32	0.42	95.0	0.77	1.0	90.0	0.46	1.0	90.0	0.44	0.44	85.0	−0.07	−0.11	85.0	−0.07	−0.11	88.3	0.31
Cooking fumes	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.64	1.0	100.0	1.0	1.0	95.0	0.64	1.0	95.0	0.64	1.0	97.5	0.82
Engine emissions	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.83	1.0	100.0	1.0	1.0	95.0	0.83	1.0	95.0	0.83	1.0	97.5	0.92
Combustion products of wood	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.0	u	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	97.5	0.50
Gas welding fumes	100.0	1.0	1.0	100.0	1.0	1.0	90.0	0.46	1.0	100.0	1.0	1.0	90.0	0.46	1.0	90.0	0.46	1.0	95.0	0.73
Arc welding fumes	90.0	0.0	u	90.0	0.0	u	90.0	0.44	0.44	100.0	1.0	1.0	90.0	0.0	u	90.0	0.0	u	91.7	0.24
Metal oxide fumes	90.0	0.61	0.61	85.0	0.32	0.42	85.0	0.32	0.42	95.0	0.77	1.0	85.0	0.32	0.42	80.0	−0.11	−0.11	86.7	0.52
Lead fumes	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	96.7	0.33
Inorganic acid solution	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.50
Caustic solution	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	[illegible]	1.0	100.0	1.0	1.0	97.5	0.50
Paints and varnishes	95.0	0.77	1.0	95.0	0.77	1.0	95.0	0.83	1.0	100.0	1.0	1.0	90.0	[illegible]	1.0	90.0	0.62	1.0	94.2	0.77
Wood stains and varnishes	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.00
Dyes and pigments	95.0	0.64	1.0	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.64	1.0	95.0	0.64	1.0	100.0	1.0	1.0	97.5	0.82
Organic dyes and pigments	95.0	0.64	1.0	90.0	0.0	u	95.0	0.64	1.0	95.0	0.0	u	100.0	1.0	1.0	95.0	0.0	u	95.0	0.38
Adhesives	90.0	0.61	0.61	95.0	0.77	1.0	90.0	0.46	1.0	95.0	0.77	1.0	90.0	0.46	1.0	95.0	0.64	1.0	92.5	0.55
Solvents	70.0	0.37	0.64	65.0	0.26	0.55	65.0	0.27	0.39	85.0	0.57	0.66	85.0	0.63	7.2	80.0	0.47	0.64	75.0	0.43
Waxes and polishes	90.0	0.0	u	90.0	0.0	u	85.0	−0.07	−0.11	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	92.5	0.16
Gasoline	100.0	1.0	1.0	95.0	0.64	1.0	90.0	0.46	1.0	95.0	0.64	1.0	90.0	0.46	1.0	95.0	0.77	1.0	94.2	0.66
Diesel oil	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.0	u	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	97.5	0.50
Mineral spirits	90.0	0.62	1.0	95.0	0.83	1.0	95.0	0.83	1.0	95.0	0.77	1.0	95.0	0.77	1.0	100.0	1.0	1.0	95.0	0.80
Aromatic naphtha	90.0	−0.05	−0.05	95.0	0.0	u	90.0	0.46	1.0	95.0	0.0	u	90.0	0.46	1.0	85.0	0.0	u	90.8	0.15
Lubricating oils and grease	75.0	0.31	0.58	90.0	0.74	1.0	80.0	0.47	0.64	85.0	0.48	0.58	85.0	0.48	0.58	90.0	0.69	0.69	84.2	0.53
Cutting fluids	95.0	0.64	1.0	100.0	1.0	1.0	95.0	0.0	u	95.0	0.64	1.0	90.0	0.0	u	95.0	0.0	u	95.0	0.38
Coal tar and pitch	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.00
Creosote	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	96.7	0.50
Pitch	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.50
Glues	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.0	u	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	97.5	0.50

Table A.7.5, continued

Substance	Chem 1–Chem 2			Chem 1–Chem 4			Chem 1–Chem 6			Chem 2–Chem 4			Chem 2–Chem 6			Chem 4–Chem 6			Average	
	p_o %	κ	κ' (2)	p_o %	κ	κ' (2)	p_o %	κ	κ' (2)	p_o %	κ	κ' (2)	p_o %	κ	κ' (2)	p_o %	κ	κ' (2)	p_o %	κ
Turpentine	95.0	0.64	1.0	90.0	0.0	u	90.0	0.44	0.44	95.0	0.0	u	95.0	0.64	1.0	90.0	0.0	u	92.5	0.29
Linseed oil	90.0	0.44	0.44	95.0	0.64	1.0	100.0	1.0	1.0	95.0	0.64	1.0	90.0	0.44	0.44	95.0	0.64	1.0	94.2	0.63
Methanol	95.0	0.64	1.0	95.0	0.0	u	90.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	95.0	0.27
Ethanol	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.50
Ethylenеglycol	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.50
Carbon tetrachloride	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.50
Methylene chloride	90.0	0.0	u	95.0	0.64	1.0	100.0	1.0	1.0	95.0	0.0	u	90.0	0.0	u	95.0	0.64	1.0	94.2	0.38
Tetraethyl lead	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.0	u	100.0	0.0	u	95.0	0.0	u	95.0	0.0	u	97.5	0.33
Trichloroethylene	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.50
Benzene	75.0	0.0	u	75.0	0.0	u	75.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	87.5	0.50
Toluene	85.0	0.0	u	85.0	0.0	u	85.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	92.5	0.50
Xylene	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.50
Phenol	90.0	0.0	u	90.0	0.0	u	90.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.50
Benzidine	90.0	0.0	u	90.0	0.0	u	90.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.50
Naphtylamine	85.0	0.0	u	85.0	0.0	u	85.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	92.5	0.50
O-toluidine	90.0	0.0	u	90.0	0.0	u	90.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.50
Aluminum compounds	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	90.0	−0.05	−0.05	95.0	0.16
Chrome compounds	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.50
Iron compounds	85.0	−0.02	−0.11	95.0	0.0	u	90.0	−0.05	−0.50	90.0	0.0	u	85.0	−0.07	−0.11	95.0	0.0	u	90.0	−0.03
Zinc compounds	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.50
Cadmium compounds	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.50
Lead compounds	90.0	0.46	1.0	90.0	0.46	1.0	85.0	0.0	u	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	92.5	0.32
Aliphatic hydrocarbons	80.0	0.0	u	85.0	0.48	0.58	75.0	0.29	0.34	85.0	0.0	u	75.0	0.0	u	90.0	0.69	1.0	81.7	0.15
Aliphatic alcohols	85.0	−0.07	−0.11	90.0	0.0	u	90.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	92.5	0.16
Aliphatic acids	90.0	0.0	u	90.0	0.0	u	90.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.50
Aliphatic aldehydes	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.50
Aliphatic saturated halogens	90.0	0.0	u	90.0	0.0	u	95.0	0.64	1.0	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	94.2	0.27
Aliphatic unsaturated halogens	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.50
Aromatic hydrocarbons	75.0	0.0	u	80.0	0.27	1.0	75.0	0.29	0.34	95.0	0.0	u	80.0	0.0	u	75.0	−0.09	−0.26	80.0	0.08
Aromatic alcohols	95.0	0.0	u	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	97.5	0.50
Rubber	95.0	0.64	1.0	100.0	1.0	1.0	95.0	0.64	1.0	95.0	0.64	1.0	100.0	1.0	1.0	95.0	0.64	1.0	97.5	0.82
Plastics	90.0	−0.05	−0.05	95.0	0.64	1.0	100.0	1.0	1.0	95.0	0.64	1.0	90.0	−0.05	−0.05	95.0	0.64	1.0	94.2	0.47
Leather	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0
Foodstuff	95.0	0.77	1.0	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.77	1.0	95.0	0.77	1.0	100.0	1.0	1.0	97.5	0.89

Table A.7.5, continued

Substance	Chem 1–Chem 2			Chem 1–Chem 4			Chem 1–Chem 6			Chem 2–Chem 4			Chem 2–Chem 6			Chem 4–Chem 6			Average	
	P_o %	κ	κ' (2)	P_o %	κ	κ' (2)	P_o %	κ	κ' (2)	P_o %	κ	κ' (2)	P_o %	κ	κ' (2)	P_o %	κ	κ' (2)	P_o %	κ
Cleaning agents	95.0	0.77	1.0	95.0	0.64	1.0	100.0	1.0	1.0	90.0	0.46	1.0	95.0	0.77	1.0	95.0	0.64	1.0	95.0	0.71
Pharmaceuticals	95.0	0.0	u	100.0	1.0	1.0	100.0	1.0	1.0	95.0	0.0	u	95.0	0.0	u	100.0	1.0	1.0	97.5	0.50
Non-specific dust	95.0	0.0	u	80.0	0.0	u	95.0	0.0	u	75.0	-0.09	-0.26	90.0	-0.05	-0.05	75.0	-0.09	-0.26	85.0	-0.04
Metal oxide dust	85.0	0.32	0.42	85.0	0.32	0.42	75.0	0.17	0.34	90.0	0.61	0.61	90.0	0.69	1.0	90.0	0.69	1.0	85.0	0.47
Average (3):	94.5	0.43	-	94.5	0.44	-	97.3	0.38	-	99.3	0.69	-	95.3	0.51	-	93.1	0.56	-	94.2	0.50
Average (4):	96.1	0.78	-	96.1	0.78	-	95.5	0.76	-	97.9	0.88	-	96.4	0.81	-	95.6	0.83	-	97.7	0.80
Average (5):	96.1	0.50	-	96.1	0.49	-	95.5	0.50	-	97.9	0.64	-	96.4	0.51	-	95.6	0.50	-	96.3	0.52

(1) The table is based on the assessment of exposure to 20 job descriptions using a dichotomous scale: exposed, at any level of confidence, and not exposed.
(2) u means undefined (i.e., 0 ÷ 0).
(3) Calculated for the 105 substances listed on the table.
(4) Calculated for all 172 substances assuming the P_o and κ equal to 1 for the remaining 67 (172 - 105) substances.
(5) Calculated using the collapsed table method (Table 7.1.4).

Appendix 8

Supplementary Tables for the Rubber Industry Trial

Table A.8.1

Occupations, Industries, and Time Periods of Employment for Job Descriptions used in the Rubber Industry Trial

Occupation Description	Standard Industrial Code(1)	Occupation Code (2)	Time Period of Employment
Moulding	162	8573-166	1941-43
Cutting and finishing	162	8575-130	1924-30
Dye mixer	162	8161-186	1949-50
Inspector	162	8576-118	1948-50
Moulding	162	8573-299	1940-40
Cutting and finishing	619	8575-178	1949-64
Carpenter	162	8781-114	1946-54
Bonding and cementing	619	8571-118	1957-64

Table A.8.1, continued

Occupation Description	Standard Industrial Code(1)	Occupation Code (2)	Time Period of Employment
Supervisor	162	4150-110	1937-80
Mechanic and millwright	162	8584-122	1963-80
Coating	162	8173-234	1973-81
Misc work printing	289	9518-199	1927-29
Coating	162	8173-234	1968-81
Bonding and cementing	162	8571-158	1943-47
Supervisor	162	4150-110	1935-39

(1) Refer to SIC, 1970.

(2) Refer to OCC, 1971.

Table A.8.2

Statistical Tests of Equality of the Proportion of Exposure Attributed Present at a Given Level of Confidence between all Raters in the Rubber Industry Trial (1)

Rater Pair	Chi-Square Value for Exposure Categorized at Confidence Level(2): High	Medium or High	Any
Chem 1 - Chem 2	0.3	8.0**	2.5
Chem 1 - Chem 3	0.9	32.0**	31.4**
Chem 2 - Chem 3	0.1	9.0**	16.4**
Final - Chem 1	23.3**	23.0**	23.0**
Final - Chem 2	30.0**	50.0**	34.8**
Final - Chem 3	36.5**	92.4**	93.1**
Overall (3)	0.8	33.4**	34.1**

Table A.8.2, continued

(1) A total of 4050 exposures = 15 job descriptions x 270 substances were used in the calculations.

(2) The tests in the first part of the table are derived using McNemar's Test (1 df). Note that: * indicates $p < .05$; ** indicates $p < .01$.

(3) This test was evaluated using Cochran's Q statistic on 2 df. The assessment of exposure in the Final coding was excluded in this calculation.

Table A.8.3

Inter-Rater Agreement using the Collapsed Table Method for the Four Exposure Variables and the Synthetic Index in the Rubber Industry Trial (1)

Rater Pair	Contact			Frequency			Concentration			Level of Confidence			Synthetic Index (2)		
	P_o %	κ	κ'	P_o %	κ	κ'	P_o %	κ	κ'	P_o %	κ	κ'	P_o %	κ	κ'
Chem 1 - Chem 2	97.0	0.60	0.67	96.5	0.53	0.60	95.7	0.43	0.51	96.6	0.55	0.61	96.6	0.55	0.62
Chem 1 - Chem 3	96.7	0.50	0.64	96.8	0.51	0.65	96.1	0.40	0.56	96.8	0.51	0.66	96.9	0.52	0.67
Chem 2 - Chem 3	96.7	0.45	0.54	96.8	0.47	0.57	96.3	0.39	0.50	96.7	0.47	0.57	96.9	0.49	0.54
Final - Chem 1	96.7	0.64	0.74	96.5	0.62	0.72	96.4	0.60	0.70	96.7	0.63	0.73	96.6	0.62	0.75
Final - Chem 2	96.4	0.58	0.74	96.4	0.58	0.72	95.7	0.51	0.68	96.3	0.57	0.74	96.3	0.58	0.73
Final - Chem 3	96.3	0.52	0.80	96.3	0.52	0.80	96.0	0.48	0.74	96.5	0.55	0.86	96.6	0.56	0.87
Average (3):	96.8	0.52	0.62	96.7	0.50	0.61	96.0	0.41	0.52	96.7	0.51	0.61	96.8	0.52	0.61

(1) A total of 4050 exposures = 15 job descriptions x 270 substances were used in the analysis.
The original scales of measurement (four categories) were used in each of the four exposure variables.
(2) The synthetic index is defined as frequency x concentration x level of confidence, grouped into 3 categories: no , "medium", and "high" exposure.
(3) The averages exclude comparisons between the individual raters and the Final Coding.

Table A.8.4

Frequency of Exposure for all Attributed Substances in the Rubber Industry Trial (1)

Substance	Prevalence of Exposure as Attributed by:				
	Chem 1 %	Chem 2 %	Chem 3 %	Final %	Average (2) %
Abrasive dust	0.0	6.7	0.0	20.0	2.2
Metallic dust	0.0	20.0	26.7	53.3	15.6
Asbestos (Chrysotile)	53.3	40.0	26.7	60.0	40.0
Talc dust	33.3	46.7	13.3	53.3	31.1
Clay dust	26.7	20.0	40.0	26.7	28.9
Inorganic Pigments	0.0	0.0	0.0	20.0	0.0
Sulphur dust	53.3	73.3	46.7	80.0	57.8
Calcium carbonate	40.0	13.3	0.0	33.3	17.8
Titanium dioxide	40.0	20.0	0.0	33.3	20.0
Iron dust	0.0	0.0	0.0	13.3	0.0
Zinc oxide	0.0	13.3	6.7	33.3	6.7
Lead dust	0.0	6.7	20.0	20.0	8.9
Carbon black	86.7	80.0	80.0	100.0	82.2
Rubber dust	0.0	0.0	0.0	53.3	0.0
Floor dust	0.0	6.7	0.0	0.0	2.2
Ammonia	0.0	0.0	0.0	6.7	0.0
Metal oxide fumes	6.7	26.7	6.7	26.7	13.4
Tin fumes	0.0	0.0	0.0	6.7	0.0
Lead fumes	0.0	20.0	0.0	13.3	6.7
Pyrolysis & Combustion fumes	80.0	80.0	40.0	80.0	66.7
Combustion products of natural gas	0.0	0.0	0.0	13.3	0.0
Benzene	53.3	20.0	0.0	53.3	24.4
Toluene	53.3	13.3	0.0	53.3	22.2
Xylene	53.3	13.3	0.0	53.3	22.2
Adhesives	20.0	33.3	33.3	26.7	28.9
Solvents	73.3	80.0	53.3	86.7	68.9
Mineral spirits	20.0	0.0	46.7	60.0	22.2
Aromatic napthas	53.3	80.0	33.3	73.3	55.5
Lubricating oils & grease	13.3	26.6	13.3	33.3	17.7

Table A.8.4, continued

Substance	Prevalence of Exposure as Attributed by:				
	Chem 1 %	Chem 2 %	Chem 3 %	Final %	Average (2) %
Zinc compounds	20.0	0.0	13.3	20.0	11.1
Lead compounds	26.7	6.7	26.7	20.0	20.0
Aliphatic Hydrocarbons	66.7	33.3	20.0	0.0	40.0
Aromatic Hydrocarbons	66.7	80.0	0.0	6.7	48.9
Aromatic amines	59.3	26.7	53.3	100.0	44.4
Rubber	100.0	100.0	100.0	100.0	100.0

(1) The table is based on the assessment of exposure to 15 job descriptions using a dichotomous scale: exposed, at any level of confidence, and not exposed.
(2) The average excludes the Final Coding.

Table A.8.5

Inter-Rater Agreement for all Attributed Substances in the Rubber Industry Trial (1)

Substance	Chem 1–Chem 2			Chem 1–Chem 3			Chem 2–Chem 3			Final–Chem 1			Final–Chem 2			Final–Chem 3			Average (2)	
	P_o %	κ	κ' (3)	P_o %	κ	κ' (3)	P_o %	κ	κ' (3)	P_o %	κ	κ' (3)	P_o %	κ	κ' (3)	P_o %	κ	κ' (3)	P_o %	κ
Abrasive dust	93.3	0.0	u	100.0	1.0	1.0	93.3	0.0	u	80.0	0.0	u	86.7	0.44	1.0	80.0	0.0	u	95.5	0.33
Metallic dust	80.0	0.0	u	73.3	0.0	u	93.3	0.82	1.0	47.6	0.0	u	66.7	0.36	1.0	73.3	0.48	1.0	82.2	0.27
Asbestos (Chrysotile)	73.3	0.47	0.64	73.3	0.48	1.0	86.7	0.71	1.0	80.0	0.60	0.69	66.7	0.36	0.63	66.7	0.39	1.0	77.7	0.55
Talc dust	73.3	0.46	0.63	66.7	0.12	0.26	66.7	0.30	1.0	80.0	0.61	1.0	93.3	0.87	1.0	60.0	0.24	1.0	68.9	0.29
Clay dust	80.0	0.44	0.54	73.3	0.41	0.58	80.0	0.55	1.0	86.7	0.66	0.66	80.0	0.44	0.53	73.3	0.41	0.58	77.8	0.47
Inorganic Pigments	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	80.0	0.0	u	80.0	0.0	u	80.0	0.0	u	100.0	1.0
Sulphur dust	66.7	0.31	0.53	80.0	0.60	0.69	60.0	0.22	0.46	73.3	0.44	1.0	93.3	0.82	1.0	66.7	0.36	1.0	68.9	0.38
Calcium carbonate	73.3	0.38	1.0	60.0	0.0	u	86.7	0.0	u	93.3	0.86	1.0	80.0	0.47	1.0	66.7	0.0	u	73.3	0.13
Titanium dioxide	80.0	0.55	1.0	60.0	0.0	u	80.0	0.0	u	93.3	0.86	1.0	86.7	0.67	1.0	66.7	0.0	u	73.3	0.18
Iron dust	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	86.7	0.0	u	86.7	0.0	u	80.7	0.04	u	100.0	1.0
Zinc oxide	87.7	0.0	u	93.3	0.0	u	80.0	−0.10	−0.16	66.7	0.0	u	80.0	0.47	1.0	73.3	0.25	1.0	87.0	−0.03
Lead dust	93.3	0.0	u	80.0	0.0	u	86.7	0.44	1.0	80.0	0.0	u	86.7	0.44	1.0	100.0	1.0	1.0	86.7	0.15
Carbon black	66.7	−0.19	−0.25	80.0	0.29	0.38	73.3	0.17	0.17	86.7	0.0	u	80.0	0.0	u	80.0	0.0	u	73.3	0.09
Rubber dust	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	46.7	0.0	u	46.7	0.0	u	46.7	0.0	u	100.0	1.0
Flock dust	93.3	0.0	u	100.0	1.0	1.0	93.3	0.0	u	100.0	1.0	1.0	93.3	0.0	u	100.0	1.0	1.0	95.5	0.33
Ammonia	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	93.3	0.0	u	93.3	0.0	u	93.3	0.0	u	100.0	1.0
Metal oxide fumes	80.0	0.33	1.0	100.0	1.0	1.0	80.0	0.33	1.0	80.0	0.33	1.0	100.0	1.0	1.0	80.0	0.33	1.0	86.7	0.55
Tin fumes	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	93.3	0.0	u	93.3	0.0	u	93.3	0.0	u	100.0	1.0
Lead fumes	80.0	0.0	u	100.0	1.0	1.0	80.0	0.0	u	86.7	0.0	u	80.0	0.29	0.38	86.7	0.0	u	86.7	0.33
Pyrolysis & Combustion fumes	73.3	0.17	0.17	60.0	0.29	1.0	46.7	0.05	0.17	86.7	0.58	0.58	73.3	0.17	0.17	60.0	0.29	1.0	60.0	0.17
Combustion products of natural gas	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	86.7	0.0	u	86.7	0.0	u	86.7	0.0	u	100.0	1.0
Benzene	66.7	0.36	1.0	46.7	0.0	u	80.0	0.0	u	86.7	0.73	0.73	53.3	0.10	0.28	46.7	0.0	u	64.5	0.12
Toluene	60.0	0.24	1.0	46.7	0.0	u	86.7	0.0	u	86.7	0.73	0.73	46.7	−0.02	−0.08	46.7	0.0	u	66.7	0.12
Xylene	60.0	0.24	1.0	46.7	0.0	u	86.7	0.0	u	86.7	0.73	0.73	46.7	−0.02	−0.08	46.7	0.0	u	64.5	0.08
Adhesives	73.3	0.35	0.49	86.7	0.67	1.0	60.0	0.10	0.10	93.3	0.82	1.0	66.7	0.21	0.25	93.3	0.84	1.0	73.3	0.37
Solvents	53.3	−0.30	−0.32	80.0	0.59	1.0	46.7	−0.11	−0.25	73.3	0.19	0.32	80.0	0.29	0.38	66.7	0.30	1.0	60.0	0.06
Mineral spirits	80.0	0.0	u	23.3	0.44	1.0	53.3	0.0	u	60.0	0.29	1.0	40.0	0.0	u	86.7	0.74	1.0	68.9	0.15
Aromatic napthas	73.4	0.44	1.0	66.7	0.35	0.57	53.3	0.22	1.0	80.0	0.59	1.0	93.3	0.82	1.0	60.0	0.31	1.0	64.5	0.34
Lubricating oils & grease	86.7	0.60	1.0	100.0	1.0	1.0	86.7	0.60	1.0	80.0	0.47	1.0	93.3	0.84	1.0	80.0	0.47	1.0	61.1	0.73

Table A.8.5, continued

Substance	Chem 1–Chem 2			Chem 1–Chem 3			Chem 2–Chem 3			Final–Chem 1			Final–Chem 2			Final–Chem 3			Average (2)	
	p_o %	κ	κ' (3)	p_o %	κ	κ' (3)	p_o %	κ	κ' (3)	p_o %	κ	κ' (3)	p_o %	κ	κ' (3)	p_o %	κ	κ' (3)	p_o %	κ
Zinc compounds	80.0	0.0	u	66.7	-0.19	-0.25	86.7	0.0	u	73.3	0.17	0.17	80.0	0.0	u	66.7	-0.19	-0.25	77.8	-0.06
Lead compounds	80.0	0.33	1.0	100.0	1.0	1.0	80.0	0.33	1.0	93.3	0.82	1.0	86.7	0.44	1.0	93.3	0.82	1.0	86.7	0.55
Aliphatic Hydrocarbons	53.3	0.16	0.40	53.3	0.22	1.0	73.3	0.33	0.49	33.3	0.0	u	66.7	0.0	u	80.0	0.0	u	60.0	0.24
Aromatic Hydrocarbons	73.3	0.33	0.49	33.3	0.0	u	20.0	.0.0	u	40.0	0.07	1.0	76.7	0.04	1.0	93.3	0.0	u	42.2	0.11
Aromatic amines	73.3	0.48	1.0	60.0	0.20	0.20	46.7	-0.04	-0.08	53.3	0.0	u	26.7	0.0	u	53.3	0.0	u	60.0	0.21
Rubber	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0
Average (4):	80.4	0.38	-	78.9	0.50	-	78.4	0.34	-	78.5	0.36	-	75.4	0.30	-	75.2	0.26	-	78.4	0.41
Average (5):	97.4	0.92	-	97.3	0.94	-	97.2	0.91	-	97.2	0.92	-	96.8	0.91	-	96.8	0.90	-	97.2 (7)	0.92
Average (6):	97.4	0.65	-	97.3	0.57	-	97.2	0.54	-	97.2	0.69	-	96.8	0.63	-	96.8	0.58	-	97.3 (7)	0.59

(1) The table is based on the assessment of exposure to 15 job descriptions using a dichotomous scale: exposed, at any level of confidence, and not exposed.
(2) The average excludes the exposure judgements made in the Final Coding
(3) u means undefined value (i.e., 0 ÷ 0).
(4) Calculated for the 35 substances listed in the table.
(5) Calculated for all 270 substances assuming that p_o and κ equal 1 for the remaining 235 (270–35) substances.
(6) Calculated using the collapsed table method (see Table 7.1.11).
(7) The difference in these two numbers is due solely to rounding errors.

Appendix 9

Supplementary Tables for the Paint Manufacturing Trial

Table A.9.1

Occupations, Industries, and Time Periods of Employment for Job Descriptions used in the Paint Manufacturing Trial

Occupation Description	Standard Industrial Code(1)	Occupation Code (2)	Time Period of Employment
Kettle Operator	378	8161-150	1938-39
Mixer	375	8161-218	1945-51
Mixer	375	8161-218	1963-79
Hand Packager	375	9318-122	1951-66
Lead-oxide maker	378	8167-178	1943-52

(1) Refer to SIC, 1970.

(2) Refer to OCC, 1971.

Table A.9.2

Statistical Tests of Equality of the Proportion of Exposure Attributed Present at a Given Level of Confidence between all Raters in the Paint Manufacturing Trial (1)

Rater Pair	Chi-Square Value for Exposure Categorized at Confidence Level (2): High	Medium or High	Any
Judge 1 -Chem 1	6.3*	25.1**	31.7**
Judge 1 -Chem 2	0.8	0.7	3.4
Chem 1 -Chem 2	2.5	22.4**	18.7**
Overall (3)	6.10	32.60**	38.10**

(1) A total of 865 exposures = 5 job descriptions x 173 substances were used in the calculations.

(2) The tests in the first part of the table are derived from using McNemar's Test (1 df). Note that: * indicates $p < .05$; ** indicates $p < .01$.

(3) This test was evaluated using Cochran's Q statistic on 2 df.

Table A.9.3

Inter-Rater Agreement using the Collapsed Table Method for the Four Exposure Variables and the Synthetic Index in the Paint Manufacturing Trial (1)

Rater Pair	Contact			Frequency			Concentration			Level of Confidence			Synthetic Index (2)		
	P_o %	κ	κ'	P_o %	κ	κ'	P_o %	κ	κ'	P_o %	κ	κ'	P_o %	κ	κ'
Judge 1 - Chem 1	92.0	0.39	0.77	92.9	0.45	0.76	91.9	0.38	0.59	93.1	0.47	0.75	93.4	0.49	0.78
Judge 1 - Chem 2	94.3	0.42	0.50	94.2	0.41	0.52	93.5	0.34	0.43	94.8	0.47	0.55	94.8	0.47	0.55
Chem 1 - Chem 2	92.7	0.49	0.78	92.6	0.48	0.63	92.4	0.47	0.62	93.6	0.55	0.72	94.0	0.57	0.75
Average (3):	93.2	0.41	0.64	93.6	0.43	0.64	92.7	0.36	0.51	94.0	0.47	0.65	94.1	0.48	0.67
Average (4):	93.0	0.43	0.68	93.2	0.45	0.64	92.6	0.40	0.55	93.8	0.50	0.67	94.1	0.51	0.69

(1) A total of 865 exposures = 5 job descriptions x 173 substances were used in the calculations.
The original scales of measurements (four categories) were used for each of the four exposure variables.
(2) The synthetic index is defined as frequency x concentration x level of confidence, grouped into 3 categories: no , "medium", and "high" exposure.
(3) The average excludes the Chem 1 - Chem 2 comparison.
(4) The average includes the Chem 1 - Chem 2 comparison.

Table A.9.4

Inter-Rater Agreement for all Attributed Substances in the Paint Manufacturing Trial (1)

Substance	Prevalence of Exposure as Attributed by: Judge 1	Chem 1	Chem 2	Average	Agreement Statistics for Rater Pair: Judge 1–Chem 1			Judge 1–Chem 2			Chem 1–Chem 2			Average (2)	
					P_o	κ	κ' (3)	P_o	κ	κ' (3)	P_o	κ	κ' (3)	P_o	κ
	%	%	%	%	%			%			%			%	
Metallic Dust	20.0	20.0	20.0	20.0	100.0	1.0	1.0	60.0	−0.25	−0.25	60.0	−0.25	−0.25	80.0	0.38
Metal Oxide Dust	40.0	80.0	60.0	60.0	60.0	0.29	1.0	80.0	0.62	1.0	80.0	0.55	1.0	70.0	0.46
Asbestos (Amphibole)	0.0	0.0	20.0	6.7	100.0	1.0	1.0	80.0	0.0	u	80.0	0.0	u	90.0	0.50
Silica Dust	0.0	20.0	0.0	6.7	80.0	0.0	u	100.0	1.0	1.0	80.0	0.0	u	90.0	0.50
Talc	0.0	20.0	70.0	13.3	80.0	0.0	u	80.0	0.0	u	100.0	1.0	1.0	80.0	0.0
Calcium Oxide	0.0	0.0	20.0	6.7	100.0	1.0	u	80.0	0.0	u	80.0	0.0	u	90.0	0.50
Titanium Dioxide	0.0	70.0	40.0	20.0	80.0	0.0	u	60.0	0.0	u	80.0	0.55	1.0	70.0	0.0
Metal Oxide Fumes	20.0	20.0	40.0	26.7	100.0	1.0	1.0	80.0	0.55	1.0	80.0	0.55	1.0	90.0	0.78
Paints, Varnishes	100.0	100.0	80.0	93.3	100.0	1.0	1.0	80.0	0.0	u	80.0	0.0	u	90.0	0.50
Wood Stains & Varnishes	40.0	60.0	0.0	33.3	80.0	0.62	1.0	60.0	0.0	u	40.0	0.0	u	70.0	0.31
Organic Dyes & Pigments	20.0	60.0	40.0	40.0	60.0	0.29	1.0	40.0	−0.36	−0.66	40.0	−0.15	0.62	50.0	−0.04
Solvents	100.0	80.0	80.0	86.7	80.0	0.0	u	80.0	0.0	u	100.0	1.0	1.0	80.0	0.0
Mineral Spirits	40.0	80.0	80.0	66.7	60.0	0.29	1.0	60.0	0.29	1.0	100.0	1.0	1.0	60.0	0.29
Linseed Oil	20.0	80.0	60.0	53.3	40.0	0.12	1.0	60.0	0.29	1.0	80.0	0.55	1.0	50.0	0.56
Dyes	0.0	20.0	20.0	13.3	80.0	0.0	u	80.0	0.0	u	60.0	−0.25	−0.25	80.0	0.0
Benzene	0.0	60.0	0.0	23.3	40.0	0.0	u	100.0	1.0	1.0	40.0	0.0	u	70.0	0.50
Toluene	80.0	60.0	0.0	46.7	40.0	−0.36	−0.66	70.0	0.0	u	40.0	0.0	u	30.0	−0.18
Xylene	80.0	60.0	0.0	46.7	40.0	−0.36	−0.66	70.0	0.0	u	40.0	0.0	u	30.0	−0.18
Chrome Compounds	0.0	40.0	0.0	13.3	60.0	0.0	u	100.0	1.0	1.0	60.0	0.0	u	80.0	0.50
Cobalt Compounds	0.0	20.0	0.0	6.7	80.0	0.0	u	100.0	1.0	1.0	80.0	0.0	u	90.0	0.50
Copper Compounds	0.0	20.0	0.0	6.7	80.0	0.0	u	100.0	1.0	1.0	80.0	0.0	u	90.0	0.50
Zinc Compounds	0.0	40.0	0.0	13.3	60.0	0.0	u	100.0	1.0	1.0	60.0	0.0	u	80.0	0.50
Cadium Compounds	0.0	40.0	0.0	13.3	60.0	0.0	u	100.0	1.0	1.0	60.0	0.0	u	80.0	0.50
Lead Compounds	40.0	100.0	40.0	60.0	40.0	0.0	u	100.0	1.0	1.0	40.0	0.0	u	70.0	0.50
Aliphatic Hydrocarbons	0.0	80.0	80.0	53.3	20.0	0.0	u	20.0	0.0	u	100.0	1.0	1.0	20.0	0.0

Table A.9.4, continued

Substance	Prevalence of Exposure as Attributed by: Judge 1 %	Chem 1 %	Chem 2 %	Average %	Agreement Statistics for Rater Pair: Judge 1- Chem 1 P_o %	κ	κ' (3)	Judge 1-Chem 2 P_o %	κ	κ' (3)	Chem 1-Chem 2 P_o %	κ	κ' (3)	Average (2) P_o %	κ
Aliphatic Alcohols	0.0	0.0	20.0	46.7	100.0	1.0	1.0	80.0	0.0	u	80.0	0.0	u	90.0	0.50
Aromatic Hydrocarbons	80.0	80.0	80.0	80.0	60.0	-0.25	-0.25	60.0	-0.25	-0.25	100.0	1.0	1.0	60.0	-0.25
Non-specific Dust	0.0	20.0	20.0	13.3	80.0	0.0	u	80.0	0.0	u	100.0	1.0	1.0	80.0	0.0
Floor Dust	0.0	20.0	20.0	13.3	80.0	0.0	u	80.0	0.0	u	100.0	1.0	1.0	80.0	0.0
Kerosene	0.0	20.0	20.0	13.3	80.0	0.0	u	80.0	0.0	u	100.0	1.0	1.0	80.0	0.0
Heating Oil	20.0	20.0	20.0	20.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0
Turpentine	0.0	40.0	40.0	26.7	60.0	0.0	u	60.0	0.0	u	100.0	1.0	1.0	60.0	0.0
Methylene Chloride	0.0	20.0	0.0	6.7	80.0	0.0	u	100.0	1.0	1.0	80.0	0.0	u	90.0	0.50
Benzidine	0.0	20.0	0.0	6.7	80.0	0.0	u	100.0	1.0	1.0	80.0	0.0	u	90.0	0.50
Naphtylamines	0.0	20.0	0.0	6.7	80.0	0.0	u	100.0	1.0	1.0	80.0	0.0	u	90.0	0.50
O-toluidine	0.0	20.0	0.0	6.7	80.0	0.0	u	100.0	1.0	1.0	80.0	0.0	u	90.0	0.50
Aliphatic Saturated Halogens	0.0	20.0	0.0	6.7	80.0	0.0	u	100.0	1.0	1.0	80.0	0.0	u	90.0	0.50
Pesticides	0.0	20.0	0.0	6.7	80.0	0.0	u	100.0	1.0	1.0	80.0	0.0	u	90.0	0.50
Paper Dust	20.0	20.0	20.0	20.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0
Coal Dust	20.0	20.0	20.0	20.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0
Carbon Monoxide	0.0	20.0	20.0	13.3	80.0	0.0	u	80.0	0.0	u	100.0	1.0	1.0	80.0	0.0
Combustion Products of Coal and Coke	20.0	20.0	20.0	20.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0	1.0	100.0	1.0
			Average (4):		74.8	0.25	-	80.0	0.45	-	78.6	0.37	-	77.4	0.36
			Average (5):		93.8	0.82	-	95.1	0.87	-	94.8	0.85	-	94.5(7)	0.85
			Average (6):		93.8	0.52	-	95.1	0.50	-	94.8	0.62	-	95.5(7)	0.51

(1) The table is based on the assessment of exposure to 5 job descriptions using a dichotomous scale: exposed, at any level of confidence, and not exposed.
(2) Average calculated for comparisons between the external judge and the two chemists.
(3) u means undefined value (i.e., 0 ÷ 0).
(4) Calculated for the 42 substances listed in the table.
(5) Calculated for all 173 substances assuming that p_o and κ=1 for 131 (173-42) substances.
(6) Calculated using the collapsed table method (see Table 7.2.5).
(7) The difference-in these 2 values is due solely to rounding errors.

Appendix 10

Supplementary Tables for the Welding Trade Trial

Table A.10.1

Occupations, Industries, and Time Periods of Employment for Job Descriptions used in the Welding Trade Trial

Industry	Standard Industrial Code(1)	Occupation Code (2)	Time Period of Employment
Appliance Mfg	332	8335-166	1947-50
Fruit & Vegetable Mfg	103	8581-122	1937-66
Boiler & Plate Works	301	8335-150	1967-72
Government Administration	931	2791-122	1938-79
Metal stamping	304	8335-142	1954-56
Railroad car mfg.	326	8511-114	1969-80
Shipbuilding	327	8335-138	1939-46
	327	8335-138	1950-51
Advertising	862	8335-138	1947-48

Table A.10.1, continued

Industry	Standard Industrial Code(1)	Occupation Code (2)	Time Period of Employment
Railway transport	503	8335-114	1940-45
Truck mfg	324	8335-126	1940-46
Bridge construction	406	8335-138	1950-60
Fabricated metal	302	8335-126	1960-67
Concrete products	354	8584-122	1945-47
Electric lamps & shades	268	8335-162	1967-71
Pulp and paper	271	8335-126	1943-51
Metal mines	059	8335-126	1951-57
Blacksmithing	896	8335-126	1957-76

(1) Refer to SIC, 1970.

(2) Refer to OCC, 1971.

Table A.10.2

Tests of Equality of the Proportion of Exposure Attributed Present at a Given Level of Confidence between all Raters in the Welding Trade Trial (1)

Rater Pair	Chi-Square Value for Exposure Categorized at Confidence Level (2): High	Medium or High	Any
Judges 2 - Chem 1	0.2	47.6**	68.3**
Judges 2 - Chem 3	50.2**	2.1	15.1**
Chem 1 - Chem 3	39.5*	68.1**	29.6**
Overall (3)	48.2**	19.2**	17.3**

Table A.10.2, continued

(1) A Total of 4662 exposures = 18 job descriptions x 259 substances were used in the calculations.

(2) The tests in the first part of the table are derived using McNemar's Test (1 df). Note that: * indicates $p < .05$; ** indicates $p < .01$.

(3) This test was evaluated using Cochran's Q statistic on 2 df.

Table A.10.3

Inter-Rater Agreement using the Collapsed Table Method for the Four Exposure Variables and the Synthetic Index in the Welding Trade Trial (1)

Rater Pair	Contact			Frequency			Concentration			Level of Confidence			Synthetic Index (2)		
	P_o %	κ	κ'	P_o %	κ	κ'	P_o %	κ	κ'	P_o %	κ	κ'	P_o %	κ	κ'
Judges 2 - Chem 1	92.0	0.49	0.54	88.5	0.29	0.43	89.4	0.34	0.41	90.6	0.41	0.49	90.1	0.38	0.44
Judges 2 - Chem 3	93.3	0.55	0.58	90.6	0.38	0.48	90.5	0.37	0.44	92.2	0.48	0.53	90.8	0.39	0.53
Chem 1 - Chem 3	95.0	0.69	0.73	92.3	0.53	0.64	91.1	0.46	0.61	93.0	0.57	0.70	92.5	0.54	0.66
Average (3):	92.7	0.52	0.56	89.6	0.34	0.46	90.0	0.36	0.43	91.4	0.45	0.51	90.5	0.39	0.49
Average (4):	93.4	0.58	0.62	90.5	0.40	0.52	90.3	0.39	0.49	91.9	0.49	0.57	91.1	0.44	0.54

(1) A total of 4662 exposures = 18 job descriptions x 259 substances were used in the analysis.
The original scales of measurement (four categories) were used for each of the four exposure variables.
(2) The synthetic index is defined as frequency x concentration x level of confidence, grouped into 3 categories: no , "medium" and "high" exposure.
(3) The average excludes the Chem 1 - Chem 3 comparison.
(4) The average includes the Chem 1 - Chem 3 comparison.

Table A.10.4

Inter-Rater Agreement for all Attributed Substances in the Welding Trade Trial (1)

Substance	Prevalence of Exposure as Attributed by:				Agreement Statistics for Rater Pair:									Average (2)	
					Judge 2-Chem 1			Judge 2-Chem 3			Chem 1-Chem 3				
	Judge 2 %	Chem 1 %	Chem 3 %	Average %	p_o %	κ	κ' (3)	p_o %	κ	κ' (3)	p_o %	κ	κ' (3)	p_o %	κ
Abrasive dust	66.7	83.3	88.9	79.6	83.3	0.57	1.0	77.8	0.40	1.0	94.4	0.77	1.0	80.6	0.49
Insulation material dust	0.0	5.6	0.0	1.9	94.4	0.0	u	100.0	1.0	1.0	94.4	0.0	u	97.2	0.50
Construction site dust	5.6	0.0	5.6	3.7	94.4	0.0	u	88.9	-.06	-.06	94.4	0.0	u	91.7	-0.03
Mine dust	5.6	0.0	5.6	3.7	94.4	0.0	u	100.0	1.0	1.0	94.4	0.0	u	97.2	0.50
Metallic dust	0.0	94.4	83.3	59.2	5.6	0.0	u	16.7	0.0	u	88.9	0.46	1.0	11.2	0.0
Asbestos (Chrysotile)	0.0	94.4	16.7	37.0	5.6	0.0	u	83.3	0.0	u	22.2	0.02	1.0	44.5	0.0
Asbestos (Amphibole)	0.0	94.4	16.7	37.0	5.6	0.0	u	83.3	0.0	u	22.2	0.02	1.0	44.5	0.0
Silica dust	5.6	0.0	27.8	11.1	94.4	0.0	u	77.8	0.27	1.0	72.2	0.0	u	86.1	0.14
Cement dust	5.6	0.0	0.0	1.9	94.4	0.0	u	94.4	0.0	u	100.0	1.0	1.0	94.4	0.0
Glass fibres	0.0	5.6	0.0	1.9	94.4	0.0	u	100.0	1.0	1.0	94.4	0.0	u	97.2	0.50
Concrete dust	5.6	0.0	0.0	1.9	94.4	0.0	u	94.4	0.0	u	100.0	1.0	1.0	94.4	0.0
Brass dust	5.6	16.7	11.1	11.1	88.9	0.46	1.0	94.4	0.64	1.0	94.4	0.77	1.0	91.7	0.55
Stainless steel dust	0.0	27.8	5.6	11.1	72.2	0.0	u	94.4	0.0	u	77.8	0.27	1.0	83.3	0.0
Mild steel dust	72.2	77.8	33.3	61.1	83.3	0.56	0.66	50.0	0.13	0.41	55.6	0.25	1.0	66.7	0.40
Solder alloy dust	0.0	0.0	11.1	3.7	100.0	1.0	1.0	88.9	0.0	u	88.9	0.0	u	94.5	0.50
Aluminum dust	22.2	22.2	11.1	18.5	100.0	1.0	1.0	88.9	0.61	1.0	88.9	0.61	1.0	94.5	1.0
Alumina	5.6	77.8	38.9	40.8	27.8	0.03	1.0	55.6	-0.11	-0.65	50.0	0.11	0.36	41.7	-0.04
Silicon carbide	0.0	77.8	55.6	44.5	22.2	0.0	u	44.4	0.0	u	66.7	0.29	0.55	33.3	0.0
Calcium oxide	5.6	0.0	0.0	1.9	94.4	0.0	u	94.4	0.0	u	100.0	1.0	1.0	94.4	0.0
Calcium carbide	0.0	5.6	0.0	1.9	94.4	0.0	u	100.0	1.0	1.0	94.4	0.0	u	97.2	0.50
Chrome dust	0.0	0.0	5.6	1.9	100.0	1.0	1.0	94.4	0.0	u	94.4	0.0	u	97.2	0.50
Iron dust	72.2	83.3	72.2	75.9	77.8	0.37	0.54	66.7	0.17	0.17	77.8	0.37	0.54	72.3	0.27
Iron oxides	77.8	0.0	55.6	44.5	22.2	0.0	u	66.7	0.29	0.55	44.4	0.0	u	44.5	0.15

Table A.10.4, continued

Substance	Prevalence of Exposure as Attributed by:				Agreement Statistics for Rater Pair:									Average (2)	
	Judge 2	Chem 1	Chem 3	Average	Judge 2-Chem 1			Judge 2-Chem 3			Chem 1-Chem 3				
	%	%	%	%	P_o %	κ	κ^1 (3)	P_o %	κ	κ^1 (3)	P_o %	κ	κ^1 (3)	P_o %	κ
Nickel dust	5.6	0.0	5.6	3.7	94.4	0.0	u	88.9	-0.06	-0.06	94.4	0.0	u	91.7	-0.03
Copper dust	0.0	5.6	5.6	3.7	94.4	0.0	u	94.4	0.0	u	88.9	-0.06	-0.06	94.4	0.0
Zinc dust	0.0	5.6	5.6	3.7	94.4	0.0	u	94.4	0.0	u	88.9	-0.06	-0.06	94.4	0.0
Silver dust	0.0	0.0	5.6	1.9	100.0	1.0	1.0	94.4	0.0	u	94.4	0.0	u	97.2	0.50
Cadmium dust	0.0	0.0	5.6	1.9	100.0	1.0	1.0	94.4	0.0	u	94.4	0.0	u	97.2	0.50
Tin dust	0.0	5.6	11.1	5.6	94.4	0.0	u	88.9	0.0	u	94.4	0.64	1.0	91.7	0.0
Lead dust	0.0	16.7	11.1	9.3	83.3	0.0	u	88.9	0.0	u	94.4	0.77	1.0	86.1	0.0
Soot	0.0	55.6	0.0	18.5	44.4	0.0	u	100.0	1.0	1.0	44.4	0.0	u	72.2	0.50
Coke dust	0.0	5.6	5.6	3.7	94.4	0.0	u	94.4	0.0	u	88.9	-0.06	-0.06	94.4	0.0
Graphite dust	27.8	0.0	0.0	9.3	72.2	0.0	u	72.2	0.0	u	100.0	1.0	1.0	72.2	0.0
Carbon monoxide	88.9	94.4	83.3	88.9	94.4	0.64	1.0	94.4	0.77	1.0	88.9	0.46	1.0	94.4	0.71
Nitrogen oxide	83.3	94.4	100.0	92.6	88.9	0.64	1.0	83.3	0.0	u	94.4	0.0	u	86.1	0.23
Ozone	77.8	94.4	94.4	88.9	83.3	0.34	1.0	72.2	-0.10	-0.29	88.9	-0.06	-0.06	77.8	0.12
Natural gas	0.0	5.6	5.6	3.7	94.4	0.0	u	94.4	0.0	u	88.9	-0.06	-0.06	94.4	0.0
Propane	5.6	27.8	5.6	13.0	77.8	0.27	1.0	100.0	1.0	1.0	77.8	0.27	1.0	88.9	0.64
Acetylene	77.8	55.6	61.1	64.8	77.8	0.53	1.0	83.3	0.62	0.62	94.4	0.89	1.0	80.6	0.58
Phosgene	0.0	61.1	38.9	33.3	38.9	0.0	u	61.1	0.0	u	33.3	-0.27	-0.47	50.0	0.0
Gas welding fumes	72.2	50.0	72.2	64.8	27.8	0.56	1.0	88.9	0.72	0.72	77.8	0.56	1.0	83.4	0.64
Arc welding fumes	83.3	88.9	94.4	88.9	94.4	0.72	1.0	88.9	0.46	1.0	94.4	0.64	1.0	91.7	0.62
Soldering fumes	38.9	5.6	16.7	20.4	55.6	-0.11	-0.65	77.8	0.48	1.0	77.8	-0.09	-0.20	66.7	0.19
Metal oxide fumes	88.9	94.4	100.0	94.4	83.3	-0.08	-0.13	88.9	0.0	u	94.4	0.0	u	86.1	-0.04
Aluminum fumes	22.2	22.2	27.8	24.1	100.0	1.0	1.0	94.4	0.85	1.0	94.4	0.85	1.0	97.2	0.93
Chrome fumes	0.0	27.8	61.1	29.6	72.2	0.0	u	38.9	0.0	u	66.7	0.39	1.0	55.6	0.0
Manganese fumes	83.3	61.1	83.3	75.9	77.8	0.48	1.0	100.0	1.0	1.0	77.8	0.48	1.0	88.9	0.74
Iron fumes	83.3	88.9	83.3	85.2	94.4	0.77	1.0	100.0	1.0	1.0	94.4	0.77	1.0	97.2	0.89
Nickel fumes	5.6	27.8	50.0	27.8	77.8	0.27	1.0	55.6	0.11	1.0	66.7	0.33	0.59	66.7	0.19
Copper fumes	27.8	33.3	38.9	33.3	83.3	0.61	0.70	77.8	0.51	0.68	83.3	0.64	0.73	80.6	0.56
Zinc fumes	55.6	50.0	33.3	46.3	61.1	0.22	0.25	77.8	0.57	1.0	83.3	0.67	1.0	69.5	0.40
Silver fumes	0.0	0.0	5.6	1.9	100.0	1.0	1.0	94.4	0.0	u	94.4	0.0	u	97.7	0.50

Table A.10.4, continued

Substance	Prevalence of Exposure as Attributed by:				Agreement Statistics for Rater Pair:									Average (2)	
	Judge 2	Chem 1	Chem 2	Average	Judge 2-Chem 1			Judge 2-Chem 3			Chem 1-Chem 3				
	%	%	%	%	p_o %	κ	κ^1 (3)	p_o %	κ	κ^1 (3)	p_o %	κ	κ^1 (3)	p_o %	κ
Cadmium	11.1	5.6	16.7	11.1	94.4	0.64	1.0	83.3	0.31	0.40	88.9	0.46.	1.0	88.9	0.48
Tin fumes	38.9	38.9	33.3	37.0	77.8	0.53	0.53	72.2	0.40	0.46	83.3	0.64	0.73	75.0	0.47
Lead fumes	44.9	72.2	50.0	55.5	61.1	0.62	0.55	72.2	0.44	0.50	55.6	0.11	0.20	66.7	0.35
Pyrolysis	5.6	94.4	77.8	59.3	11.1	0.01	1.0	27.8	0.03	1.0	72.2	-0.10	0.29	19.5	0.02
Combustion products of coke	11.1	5.6	0.0	5.6	94.4	0.64	1.0	88.9	0.0	u	94.4	0.0	u	91.7	0.32
Combustion products of natural gas	0.0	0.0	5.6	1.9	100.0	1.0	1.0	94.4	0.0	u	94.4	0.0	u	97.7	0.50
Inorganic acid solution	0.0	5.6	0.0	1.9	94.4	0.0	u	100.0	1.0	1.0	94.4	0.0	u	97.7	0.50
Caustic solutions	0.0	5.6	0.0	1.9	94.4	0.0	u	100.0	1.0	1.0	94.4	0.0	u	97.7	0.50
Other inorganic liquids	38.9	0.0	0.0	13.0	61.1	0.0	u	61.1	0.0	u	100.0	1.0	1.0	61.1	0.0
Hydrofluoric acid	38.9	77.8	100.0	72.2	50.0	0.11	0.36	38.9	0.0	u	77.8	0.0	u	44.5	0.06
Hydrochloric acid	38.9	83.3	94.4	72.2	44.4	0.03	0.13	44.4	0.07	1.0	77.8	-0.10	-0.22	44.4	0.05
Acrolein	0.0	0.0	16.7	5.6	100.0	1.0	1.0	83.3	0.0	u	83.3	0.0	u	91.7	0.50
Carbon tetrachloride	0.0	72.2	5.6	25.9	27.8	0.0	u	94.4	0.0	u	33.3	0.04	1.0	61.1	0.0
Trichloroethylene	11.1	5.6	5.6	7.4	83.3	-0.08	-0.13	94.4	0.64	1.0	88.9	-0.06	-0.06	88.9	0.28
Perchloroethylene	0.0	5.6	5.6	3.7	94.4	0.0	u	94.4	0.0	u	88.9	-0.06	-0.06	94.4	0.0
Toluene	0.0	0.0	5.6	1.9	100.0	1.0	1.0	94.4	0.0	u	94.4	0.0	u	97.2	0.50
Turpentine	11.1	5.6	5.6	7.4	94.4	0.64	1.0	94.4	0.64	1.0	100.0	1.0	1.0	94.4	0.64
Solvents	66.7	72.2	33.3	57.4	83.3	0.61	0.70	66.7	0.40	1.0	61.1	0.32	1.0	75.0	0.51
Mineral spirits	0.0	72.2	33.3	35.2	27.8	0.0	u	66.7	0.0	u	61.1	0.32	1.0	47.3	0.0
Oils & grease	66.7	11.1	11.1	29.6	44.4	0.12	1.0	44.4	0.12	1.0	100.0	1.0	1.0	44.4	0.12
Cutting fluids	5.6	0.0	0.0	1.9	94.4	0.0	u	94.4	0.0	u	100.0	1.0	1.0	94.4	0.0

Table A.10.4, continued

Substance	Prevalence of Exposure as Attributed by: Judge 2 %	Chem 1 %	Chem 2 %	Average %	Agreement Statistics for Rater Pair: Judge 2–Chem 1 p_o %	κ	κ^1 (3)	Judge 2–Chem 3 p_o %	κ	κ^1 (3)	Chem 1–Chem 3 p_o %	κ	κ^1 (3)	Average (2) p_o %	κ
Paints & varnishes	0.0	0.0	5.6	1.9	100.0	1.0	1.0	94.4	0.0	u	94.4	0.0	u	97.2	0.50
Cyanides	0.0	0.0	5.6	1.9	100.0	1.0	1.0	94.4	0.0	u	94.4	0.0	u	97.2	0.50
Fluorides	38.9	0.0	22.2	20.4	61.1	0.0	u	50.0	-0.14	-0.23	77.8	0.0	u	55.6	-0.07
Aluminum compounds	22.2	0.0	0.0	7.4	77.8	0.0	u	77.8	0.0	u	100.0	1.0	1.0	77.8	0.0
Manganese compounds	83.3	0.0	0.0	27.8	16.7	0.0	u	16.7	0.0	u	100.0	1.0	1.0	16.7	0.0
Iron compounds	83.3	0.0	0.0	27.8	16.7	0.0	u	16.7	0.0	u	100.0	1.0	1.0	16.7	0.0
Nickel compounds	5.6	0.0	0.0	1.9	94.4	0.0	u	94.4	0.0	u	100.0	1.0	1.0	94.4	0.0
Copper compounds	27.8	0.0	0.0	9.3	72.2	0.0	u	72.2	0.0	u	100.0	1.0	1.0	72.2	0.0
Zinc compounds	55.6	0.0	0.0	18.5	44.4	0.0	u	44.4	0.0	u	100.0	1.0	1.0	44.4	0.0
Cadmium	11.1	0.0	0.0	3.7	88.9	0.0	u	88.9	0.0	u	100.0	1.0	1.0	88.9	0.0
Tin compounds	38.9	0.0	5.6	14.8	61.1	0.0	u	66.7	0.17	1.0	94.4	0.0	u	66.4	0.09
Lead compounds	44.4	0.0	11.1	18.5	55.6	0.0	u	55.6	0.03	0.11	88.4	0.0	u	55.6	0.02
Aliphatic alcohols	0.0	0.0	5.6	1.9	100.0	1.0	1.0	94.4	0.0	u	94.4	0.0	u	97.2	0.50
Ionizing radiation	16.7	16.7	0.0	11.1	100.0	1.0	1.0	83.3	0.0	u	83.3	0.0	u	91.7	0.50
Bronze dust	5.6	16.7	11.1	11.1	88.9	0.46	1.0	94.4	0.64	1.0	94.4	0.77	1.0	91.7	0.55
				Average (4):	80.3	0.29	-	87.1	0.24	-	87.9	0.33	-	77.6	0.27
				Average (5):	93.3	0.76	-	95.1	0.74	-	95.9	0.77	-	92.4 (7)	0.75
				Average (6):	93.3	0.55	-	95.6	0.61	-	95.9	0.69	-	95.5 (7)	0.58

(1) The table is based on the assessment of exposure to 18 job descriptions using a dichotomous scale: exposed, at any level of confidence, and not exposed.
(2) Average based on agreement between the external rater and the two chemists only.
(3) u means undefined value (i.e., 0 ÷ 0).
(4) Calculated for the 88 substances listed in the table.
(5) Calculated for all 259 substances assuming that 171 substances (259–88) have p_o and κ=1.
(6) Calculated using the collapsed table method (see Table 7.2.12).
(7) The difference in these 2 values is due solely to rounding errors.

Appendix 11

Supplementary Tables for the Metal Industry Trial

Table A.11.1

Occupations, Industries, and Time Periods of Employment for Job Descriptions used in the Metal Industry Trial

Occupation Description	Standard Industrial Code(1)	Occupation Code (2)	Time Period of Employment
Sheet Metal Worker	304	8333-118	1981-84
Metal Polisher	304	8393-298	1953-84
Welder	304	8335-126	1979-84
Metal Forming Operator	309	8337-138	1973-84
Drill-press Operator	309	8315-150	1980-80
Arc Welder	309	8335-150	1981-84
Spray Painter	309	8595-142	1981-84

(1) Refer to SIC, 1970.

(2) Refer to OCC, 1971.

Table A.11.2

Statistical Tests of Equality of the Proportion of Exposure Attributed Present at a Given Level of Confidence in the Metal Industry Trial (1)

Chi-Square Value for Exposure Categorized at Confidence Level:		
High	Medium or High	Any
1.3	7.5**	6.7**

(1) A total of 2100 exposures = 7 job descriptions x 300 substances were used in the calculations. The values in the table are obtained from McNemar's Test (1 df). Note that: * indicates $p < .05$; ** indicates $p < .01$.

Table A.11.3

Inter-Rater Agreement for all Attributed Substances in the Metal Industry Trial (1)

Substance	Prevalence of Exposure as Attributed by: Consensus	Judges-3	Average	P_o	κ	κ' (2)
	%	%	%	%		
Solvents	28.6	57.1	42.9	71.4	0.46	1.0
Xylene	14.3	14.3	14.3	100.0	1.0	1.0
Toluene Di-isocyanate	14.3	14.3	14.3	100.0	1.0	1.0
Aliphatic esters	14.3	14.3	14.3	100.0	1.0	1.0
Aliphatic ketones	14.3	42.9	28.6	71.4	0.36	1.0
Filler pigments	14.3	0.0	7.2	85.7	0.0	u
Polyurethanes	14.3	0.0	7.2	85.7	0.0	u
Talc	0.0	14.3	7.2	85.7	0.0	u
Titanium dioxide	0.0	14.3	7.2	85.7	0.0	u
Iron oxides	14.3	14.3	14.3	71.4	-0.17	-0.17
Phosphoric acid	0.0	14.3	7.2	85.7	0.0	u
Mineral spirits	0.0	28.6	14.3	71.4	0.0	u
Toluene	0.0	57.1	28.6	42.9	0.0	u
Iron compounds	0.0	14.3	7.2	85.7	0.0	u
Cobalt compounds	0.0	14.3	7.2	85.7	0.0	u
Zinc compounds	0.0	28.6	14.3	71.4	0.0	u
Chromates	0.0	28.6	14.3	71.4	0.0	u
Abrasive dust	57.1	57.1	57.1	71.4	0.42	1.0
Metallic dust	85.7	71.4	78.6	85.7	0.59	1.0
Silicon carbide	0.0	57.1	28.6	42.9	0.0	u
Aluminum dust	42.9	28.6	35.8	85.7	0.70	1.0
Mild steel dust	42.9	71.4	57.2	42.9	-0.08	-0.17
Stainless steel dust	71.4	71.4	71.4	71.4	0.30	0.3
Carbon monoxide	57.1	57.1	57.1	100.0	1.0	1.0
Nitrogen oxides	57.1	57.1	54.1	100.0	1.0	1.0
Ozone	28.6	57.1	42.9	71.4	0.46	1.0
Pyrolysis & combustion fumes	28.6	28.6	28.6	100.0	1.0	1.0
Arc welding fumes	71.4	57.1	64.3	85.7	0.70	1.0
Soldering fumes	14.3	14.3	14.3	100.0	1.0	1.0
Metal oxide fumes	57.1	57.1	57.1	71.4	0.42	0.42

Table A.11.3, continued

Substance	Prevalence of Exposure as Attributed by: Consensus	Judges-3	Average	P_o	κ	κ' (1)
	%	%	%	%		
Aluminum fumes	14.3	14.3	14.3	100.0	1.0	1.0
Chromium fumes	28.6	57.1	42.9	71.4	0.46	1.0
Manganese fumes	0.0	57.1	28.6	42.9	0.0	u
Iron fumes	57.1	57.1	57.1	71.4	0.42	0.42
Nickel fumes	28.6	57.1	42.0	71.4	0.46	1.0
Copper fumes	14.3	28.6	21.5	85.7	0.59	1.0
Silver fumes	28.6	28.6	28.6	100.0	1.0	1.0
Tin fumes	14.3	28.6	21.5	85.7	0.59	1.0
Lead fumes	28.6	28.6	21.5	100.0	1.0	1.0
Caustic solution	0.0	42.9	21.5	57.1	0.0	u
Mineral oil	0.0	57.1	28.9	42.9	0.0	u
Ultraviolet radiation	28.6	42.9	35.8	85.7	0.70	1.0
Alumina	71.4	57.1	64.3	85.7	0.70	1.0
Iron oxides	14.3	0.0	7.2	85.7	0.0	u
Propane	14.3	0.0	7.2	85.7	0.0	u
Zinc fumes	28.6	14.3	21.5	85.7	0.59	1.0
Inorganic acid solution	42.9	0.0	21.5	57.1	0.0	u
Hydrochloric acid	14.3	0.0	7.2	85.7	0.0	u
Lubricating oils & grease	28.6	14.3	21.5	85.7	0.59	1.0
Phosgene	0.0	14.3	7.2	85.7	0.0	u
Gas welding fumes	28.6	14.3	21.5	85.7	0.59	1.0
Other inorganic gas	14.3	0.0	7.2	85.7	0.0	u
Natural gas	14.3	0.0	7.2	85.7	0.0	u
Acetylene	28.6	0.0	14.3	71.4	0.0	u
Combustion products of natural gas	14.3	0.0	7.2	85.7	0.0	u
Trichloroethylene	14.3	0.0	7.2	85.7	0.0	u
Acrolein	14.3	0.0	7.2	85.7	0.0	u
Sulfuric acid	28.6	0.0	14.3	71.4	0.0	u
Cotton dust	0.0	14.3	7.2	85.7	0.0	u
Hydrofluoric acid	0.0	14.3	7.2	85.7	0.0	u

Table A.11.3, continued

Substance	Prevalence of Exposure as Attributed by: Consensus	Judges-3	Average	p_o	κ	κ' (2)
	%	%	%	%		
Animal, vegetable glues	0.0	14.3	7.2	85.7	0.0	u
Inorganic pigments	14.3	14.3	14.3	100.0	1.0	1.0
Carbon black	14.3	14.3	14.3	100.0	1.0	1.0
Paints, varnishes lacquers	14.3	14.3	14.3	100.0	1.0	1.0
Zinc dust	14.3	0.0	7.2	85.7	0.0	u
			Average (3):	81.1	0.35	-
			Average (4):	95.9	0.86	-
			Average (5):	95.9	0.59	-

(1) The table is based on the assessment of exposure to 7 job descriptions using the dichotomous scale: exposed, at any level of confidence, and not exposed.
(2) u means undefined value (i.e., 0÷0).
(3) Calculated for the 65 substances listed in the table.
(4) Calculated for all 300 substances assuming that p_o and $\kappa = 1$ for the remaining 235 (300-65) substances.
(5) Calculated using the collapsed table method (Table 7.3.4).

Appendix 12

Supplementary Tables for the Chemical Manufacturing Trial

Table A.12.1

Occupations, Industries, and Time Periods of Employment for Job Descriptions used in the Chemical Manufacturing Trial

Occupation Description	Standard Industrial Code(1)	Occupation Code (2)	Time Period of Employment
Stationary Engineer	373	9533-122	1975-78
Chemical Process Operator	373	8179-122	1977-82
Flaker Tender	373	8167-342	1975-80
Chemical Labourer	373	8178-110	1970-76
Filter Press Tender	373	8163-155	1977-80

(1) Refer to SIC, 1970.

(2) Refer to OCC, 1971.

Table A.12.2

Statistical Tests of Equality of the Proportion of Exposure Attributed Present at a Given Level of Confidence among Chemical Manufacturing Trial (1)

Chi-Square value for Exposure Categorized at Confidence Level:		
High	Medium or High	Any
6.8**	4.8*	6.4*

(1) A total of 1500 exposures = 5 job descriptions x 300 substances were used in the calculations. The tests are derived using McNemar's Test (1 df). Note that: * indicates $p \leq .05$; ** indicates $p < .01$.

Table A.12.3

Inter-Rater Agreement for all Attributed Substances in the Chemical Manufacturing Trial (1)

Substance	Prevalence of Exposure as Attributed by:			Agreement Statistics		
	Consensus %	Judge 4 %	Average %	p_o %	κ	κ' (2)
Inorganic Pigments	0.0	20.0	10.0	80.0	0.0	u
Asbestos (Amphibole)	0.0	20.0	10.0	80.0	0.0	u
Other Inorganic Dust	0.0	20.0	10.0	80.0	0.0	u
Titanium Dioxide	20.0	20.0	20.0	100.0	1.0	1.0
Paper Dust	0.0	40.0	20.0	60.0	0.0	u
Synthetic Fibres	0.0	40.0	20.0	60.0	0.0	u
Carbon Monoxide	0.0	20.0	10.0	80.0	0.0	u
Natural Gas	20.0	20.0	20.0	100.0	1.0	1.0
Ozone	20.0	20.0	20.0	100.0	1.0	1.0
Formaldehyde	20.0	20.0	20.0	100.0	1.0	1.0
Combustion Products of Natural Gas	20.0	20.0	20.0	100.0	1.0	1.0
Inorganic Acid Solution	20.0	0.0	10.0	80.0	0.0	u
Acetic Acid	60.0	100.0	80.0	60.0	0.0	u
Sulfuric Acid	20.0	20.0	20.0	60.0	-0.25	-0.25
Mineral Spirits	40.0	20.0	30.0	40.0	-0.36	-0.66
Lubricating Oil & Grease	0.0	40.0	20.0	60.0	0.0	u
Acetone	40.0	40.0	40.0	60.0	0.17	0.17
Alkali Caustic Solution	40.0	40.0	40.0	100.0	1.0	1.0
Methanol	0.0	20.0	10.0	80.0	0.0	u
Formic Acid	0.0	20.0	10.0	80.0	0.0	u
Solvents	20.0	20.0	20.0	60.0	-0.25	-0.25
Benzene	20.0	20.0	20.0	100.0	1.0	1.0
Cellulose Acetate	60.0	60.0	60.0	100.0	1.0	1.0
Cellulose	40.0	0.0	20.0	60.0	0.0	u
Insulation Material	0.0	40.0	20.0	60.0	0.0	u
Asbestos (Chrysotile)	0.0	40.0	20.0	60.0	0.0	u
Concrete Dust	0.0	20.0	10.0	80.0	0.0	u
Aliphatic Alcohols	40.0	40.0	40.0	100.0	1.0	1.0
Aliphatic Aldehydes	20.0	20.0	20.0	100.0	1.0	1.0
Aliphatic Acids	0.0	20.0	10.0	80.0	0.0	u
Aliphatic Ketones	20.0	40.0	30.0	80.0	0.55	1.0

Table A.12.3, continued

Substance	Prevalence of Exposure as Attributed by: Consensus %	Judge 4 %	Average %	Agreement Statistics p_o %	κ	κ' (2)
Other Inorganic Vapors	20.0	0.0	10.0	80.0	0.0	u
Aliphatic Esters	20.0	20.0	20.0	100.0	1.0	1.0
Cleaning Agents	20.0	0.0	10.0	80.0	0.0	u
		Average (3):		79.4	0.34	-
		Average (4):		97.7	0.93	-
		Average (5):		97.7	0.52	-

(1) The table is based on the assessment of exposure to 5 job descriptions using a dichtomous scale: exposed, at any level of confidence, and not exposed.
(2) u means undefined value (i.e., 0 ÷ 0).
(3) Calculated for the 34 substances listed in the table.
(4) Calculated for all 300 substances assuming that p_o and κ = 1 for the remaining 266 (300-34).
(5) Calculated using the collapsed table method (Table 7.3.11).

Appendix 13

Supplementary Tables for the Code/Recode Trial

Table A.13.1

Occupations, Industries, and Time Periods of Employment for Job Descriptions used in the Code/Recode Trial

Industry/ Occupation Description	Standard Industrial Code(1)	Occupation Code (2)	Time Period of Employment
Drug and Toilet/ Salesman	616	5141-110	1930-37
Defense/ Truck Driver	902	9175-118	1939-61
Motor Vehicle Repairs/ Salesman	658	5131-118	1962-70
Local Administration/ Fire Fighter	951	6111-126	1971-80
Chemical Industry/ Cleaner	379	6191-114	1930-70
Farming/ Farm Hand	017	7181-110	1925-46 1947-50

Table A.13.1, continued

Occupation Description	Standard Industrial Code(1)	Occupation Code (2)	Time Period of Employment
Trade Contractor/ Roofer	421	8787-118	1943-60
Trade Contractor/ Foreman	421	8780-154	1961-81
Signs & Display/ Stripper	397	8595-334	1947-49
Signs & Display/ Sales Manager	397	1137-118	1950-78
Fruit & Vegetable Proces./ Manager	103	1143-110	1982-84
Pulp and Paper/ Logger	271	7513-122	1937-42

Table A.13.1, continued

Occupation Description	Standard Industrial Code(1)	Occupation Code (2)	Time Period of Employment
Defense/ Aircraft Mechanic	902	8582-110	1942-49
Appliance Manufacture/ Appliance Assembler	332	8531-142	1947-56
Communication Eqpt Mfg/ Braider Tender	335	8149-166	1956-71
Communications Eqpt Mfg/ Shipping Supervisor	335	4150-114	1971-75
Communication Eqpt Mfg/ Stationary Engineer	335	9533-122	1978-84
Chemical Manufacture/ Labourer	378	8178-110	1944-77

Table A.13.1, continued

Occupation Description	Standard Industrial Code(1)	Occupation Code (2)	Time Period of Employment
Trade Contractor/ Painter	421	8785-110	1950-77
Textile Manufacture/ Machine Cleaner	189	8278-174	1980-84
Trade Contractor/ Sheet Metal Worker	421	8333-118	1946-56
Railway Transport/ Instructor	503	2797-199	1956-64

(1) Refer to SIC, 1970.

(2) Refer to OCC, 1971.

Table A.13.2

Statistical Tests of Equality of the Proportion of Exposure Attributed Present at a Given Level of Confidence in the Code/Recode Trial (1)

Chi-Square Value for Exposure Categorized at Confidence Level:		
High	Medium or High	Any
0.04	0.01	0.17

(1) A total of 6210 exposures = 23 job descriptions x 270 substances were used in the calculations. The values in the table are obtained from McNemar's Test (1 df). Note that: * indicates $p < .05$; ** indicates $p < .01$.

Table A.13.3

Inter-Rater Agreement for all Attributed Substances in the Code/Recode Trial (1)

Substance	Prevalence of Exposure as Attributed in the: First Code %	Recode %	p_o %	κ	κ' (2)
Abrasive Dust	0.0	13.0	87.0	0.0	u
Insulation Material Dust	21.7	13.0	91.3	0.70	1.0
Construction Site Dust	4.3	4.3	100.0	1.0	1.0
Mine Dust	8.7	4.3	96.7	0.65	1.0
Metallic Dust	17.4	8.7	82.6	0.25	0.40
Asbestos (Chrysotile)	8.7	8.7	100.0	1.0	1.0
Asbestos (Amphibole)	8.7	8.7	100.0	1.0	1.0
Silica Dust	4.3	8.7	95.7	0.65	1.0
Glass Fibres	4.3	0.0	95.7	0.0	u
Inorganic Pigments	4.3	4.3	100.0	1.0	1.0
Alumina	0.0	8.7	91.3	0.0	u
Sulfur	4.3	0.0	95.7	0.0	u
Calcium Oxide	8.7	8.7	100.0	1.0	1.0
Zinc Oxide	0.0	4.3	95.7	0.0	u
Gypsum	0.0	4.3	95.7	0.0	u
Titanium Dioxide	0.0	4.3	95.7	0.0	u
Iron Oxides	4.3	4.3	100.0	1.0	1.0
Ultraviolet Radiation	4.3	4.3	100.0	1.0	1.0
Lead Oxides	8.7	0.0	91.3	0.0	u
Aluminum Dust	4.3	0.0	95.7	0.0	u
Iron Dust	4.3	0.0	95.7	0.0	u
Copper Dust	8.7	4.3	95.7	0.65	1.0
Zinc Dust	4.3	4.3	100.0	1.0	1.0
Tin Dust	0.0	4.3	95.7	0.0	u
Stainless Steel Dust	4.3	4.3	100.0	1.0	1.0
Mild Steel Dust	8.7	8.7	100.0	1.0	1.0
Coal Dust	4.3	4.3	100.0	1.0	1.0
Carbon Black	4.3	0.0	95.7	0.0	u

Table A.13.3, continued

Substance	Prevalence of Exposure as Attributed in the: First Code %	Recode %	p_o %	κ	κ' (2)
Soot	4.3	0.0	95.7	0.0	u
Fabric Dust	4.3	0.0	95.7	0.0	u
Synthetic Fibres	4.3	0.0	95.7	0.0	u
Rubber Dust	4.3	8.7	95.7	0.65	1.0
Polystyrene Dust	4.3	4.3	100.0	1.0	1.0
Cotton Dust	4.3	4.3	100.0	1.0	1.0
Wool Fibres	4.3	4.3	100.0	1.0	1.0
Wood Dust	13.3	13.0	91.3	0.62	0.62
Cork Dust	4.3	4.3	100.0	1.0	1.0
Floor Dust	4.3	0.0	95.7	0.0	u
Carbon Monoxide	13.0	8.7	95.7	0.78	1.0
Hydrogen Cyanide	4.3	4.3	100.0	1.0	1.0
Ammonia	8.7	8.7	100.0	1.0	1.0
Nitrogen Oxides	8.7	8.7	100.0	1.0	1.0
Ozone	4.3	4.3	100.0	1.0	1.0
Sulfur Dioxide	8.7	4.3	95.7	0.65	1.0
Hydrogen Sulphide	4.3	4.3	100.0	1.0	1.0
Chlorine	4.3	4.3	100.0	1.0	1.0
Spray Gases	4.3	4.3	100.0	1.0	1.0
Freons	4.3	4.3	100.0	1.0	1.0
Propane	8.7	4.3	95.7	0.65	1.0
Formaldehyde	8.7	8.7	91.3	0.45	0.45
Acetylene	0.0	4.3	95.7	0.0	u
Phosgene	4.3	4.3	100.0	1.0	1.0
Pyrolysis and Combustion Fumes	8.7	8.7	100.0	1.0	1.0
Engine (Diesel) Emissions	4.3	8.7	95.7	0.65	1.0

Table A.13.3, continued

Substance	Prevalence of Exposure as Attributed in the: First Code %	Recode %	P_o %	κ	κ' (2)
Combustion Products of Coke and Coal	4.3	4.3	100.0	1.0	1.0
Combustion Products of Liquid Fuel	4.3	13.0	91.3	0.47	1.0
Combustion Products of Wood	4.3	8.7	95.7	0.65	1.0
Gas Welding Fumes	0.0	4.3	95.7	0.0	u
Arc Welding Fumes	4.3	4.3	100.0	1.0	1.0
Soldering Fumes	4.3	4.3	100.0	1.0	1.0
Metal Oxide Fumes	4.3	4.3	100.0	1.0	1.0
Aluminum Fumes	4.3	4.3	100.0	1.0	1.0
Chrome Fumes	4.3	4.3	100.0	1.0	1.0
Iron Fumes	4.3	4.3	100.0	1.0	1.0
Nickel Fumes	4.3	4.3	100.0	1.0	1.0
Copper Fumes	4.3	4.3	100.0	1.0	1.0
Zinc Fumes	4.3	4.3	100.0	1.0	1.0
Tin Fumes	4.3	4.3	100.0	1.0	1.0
Lead Fumes	4.3	4.3	100.0	1.0	1.0
Inorganic Acid Solution	4.3	0.0	95.7	0.0	u
Javex	0.0	4.3	95.7	0.0	u
Hydrofluoric Acid	0.0	4.3	95.7	0.0	u
Sulfuric Acid	0.0	4.3	95.7	0.0	u
Hydrochloric Acid	8.7	8.7	100.0	1.0	1.0
Engine Emissions	34.8	43.5	91.3	0.82	1.0
Leaded Gasoline	8.7	8.7	91.3	0.45	0.45
Paints, Varnishes, Lacquers	8.7	8.7	100.0	1.0	1.0
Wood Stains & Varnishes	0.0	4.3	95.7	0.0	u

Table A.13.3, continued

Substance	Prevalence of Exposure as Attributed in the: First Code %	Recode %	p_o %	κ	κ' (2)
Organic Dyes	0.0	4.3	95.7	0.0	u
Adhesives	4.3	4.3	100.0	1.0	1.0
Solvents	17.4	13.0	87.0	0.50	0.60
Waxes, Polishes	4.3	0.0	95.7	0.0	u
Kerosene	8.7	13.0	95.7	0.78	1.0
Heating Oil	8.7	0.0	91.3	0.0	u
Mineral Spirits	21.7	8.7	78.3	0.18	0.35
Lubricating Oils & Greases	26.1	26.1	82.6	0.55	0.55
Hydraulic Fluid	8.7	8.7	100.0	1.0	1.0
Turpentine	8.7	4.3	95.7	0.65	1.0
Linseed Oil	8.7	8.7	100.0	1.0	1.0
Polychlorinated Biphenyl	4.3	0.0	95.7	0.0	u
Mineral Spirits with BTX	0.0	13.0	87.0	0.0	u
Isopropanol	4.3	4.3	100.0	1.0	1.0
Acetic Acid	0.0	4.3	95.7	0.0	u
P-dioxan	0.0	4.3	95.7	0.0	u
Methylen Chloride	0.0	4.3	95.7	0.0	u
Trichlorethylene	0.0	4.3	95.7	0.0	u
Lead Compounds	0.0	4.3	95.7	0.0	u
Chromates	4.3	0.0	95.7	0.0	u
Aliphatic Hydrocarbons	8.7	0.0	91.3	0.0	u
Aromatic Hydrocarbons	4.3	0.0	95.7	0.0	u
Foodstuff	0.0	4.3	95.7	0.0	u
Cleaning Agents	8.7	4.3	95.7	0.65	1.0
Mineral Wool	0.0	8.7	91.3	0.0	u
AVGAS	0.0	4.3	95.7	0.0	u

Table A.13.3, continued

Substance	Prevalence of Exposure as Attributed in the: First Code %	Recode %	p_e %	κ	κ' (2)
Poly (acrylonitrile) butadiene	4.3	0.0	95.7	0.0	u
Coal Tar & Pitch	13.0	21.7	91.3	0.70	1.0
Asphalt	17.4	26.1	91.3	0.75	1.0
Polyesters	4.3	0.0	95.7	0.0	u
		Average (3):	95.5	0.53	-
		Average (4):	98.5	0.83	-
		Average (5):	98.5	0.67	-

(1) The table is based on the assessment of exposure to 23 job descriptions using a dichotomous scale: exposed, at any level of confidence, and not exposed.
(2) u means undefined value (i.e., 0 ÷ 0).
(3) Calculated for the 108 substances listed in the table.
(4) Calculated for all 270 substances assuming that p_e and κ=1 for the remaining 162 (270-108) substances.
(5) Calculated using the collapsed table method (see Table 7.4.3).