

Comparison of Self-Assessment of Solvent Exposure With Measurement and Professional Assessment for Female Petrochemical Workers in China

Ye A. Hu, DSc,^{1*} Thomas J. Smith, PhD,¹ Xiping Xu, MD,¹ Lihua Wang, MD,² Hiroshi Watanabe, MD,³ and David C. Christiani, MD¹

Objective *The primary objective of this paper is to examine the validity of self-assessment of solvent exposure by comparing it with professional assessment and actual measurements.*

Methods *Self-assessment of exposures to benzene, toluene, styrene, and xylene was obtained from 132 female workers. The exposures were also estimated by an occupational hygienist and by actual measurement. Self-assessment, professional assessment, and measurement were then compared with each other.*

Results *Fair to good agreement was found between self-assessment, professional assessment, and measurement for benzene, styrene, and xylene. The agreement between self-assessment and measurement was poor for toluene, whereas the agreement between self-assessment and professional assessment was good. The latter was caused by a biased professional assessment.*

Conclusions *Workers' self-assessment and professional assessment provided useful information for benzene, styrene, and xylene exposure, but not for toluene exposure. False agreement can be obtained when professional assessment was used as reference in validity study. Am. J. Ind. Med. 41:483–489, 2002. © 2002 Wiley-Liss, Inc.*

KEY WORDS: *validity; sensitivity; specificity; overall proportion of agreement; Cohen's kappa*

¹Department of Environmental Health, Harvard School of Public Health, Huntington Avenue, Boston, Massachusetts

²Department of Environmental Sciences, School of Public Health, Beijing Medical University, Beijing, China

³Information System Department, University of Occupational and Environmental Health, Japan

This work was performed at Department of Environmental Health, Harvard School of Public Health, 665 Huntington Avenue, Boston, MA 02115

Contract grant sponsor: NIOSH; Contract grant number: OH03027; Contract grant sponsor: NIH; Contract grant number: ES00002.

*Correspondence to: Ye A. Hu, Analytical and Chemical Sciences, Research Triangle Institute, 3040 Cornwallis Road, Research Triangle Park, NC 27709. E-mail: yhu@rti.org

Accepted 20 November 2001

DOI 10.1002/ajim.10069. Published online in Wiley InterScience (www.interscience.wiley.com)

© 2002 Wiley-Liss, Inc.

INTRODUCTION

In epidemiologic studies, self-assessed exposure is a widely used, inexpensive approach for exposure assessment. However, the validity of self-assessed exposure has not been well evaluated, especially for exposures to organic solvents. The major hurdle is that the “gold standard,” measurement data, is not available in most cases and when it is, the data is of limited scope. Therefore, attempts made to investigate the validity of self-assessed exposure were mainly indirect evaluations made by comparing self-assessed exposures with professional assessment, which were based upon employer's record [Rosenstock et al.,

1984, 1987; Eskenazi and Pearson, 1988; Ahlborg, 1990; Libscomb et al., 1991; Birdsong et al., 1992]. Indirect evaluation can be potentially unreliable because professional assessment itself can be biased or inaccurate. For example, in a study of agreement between professional assessment and measurement, the R^2 value ranges from 0 to 0.67, indicating a wide range of no agreement to good agreement [Kromhout et al., 1987].

Another problem for evaluation of the self-assessment is that the statistical measures for validity study are less developed. Although sensitivity and specificity are regarded as most appropriate [Maclure and Willett, 1987; Armstrong et al., 1994], many researchers chose not to use them because they are not summary measures. Instead, the overall proportion of agreement (also called "per cent agreement") and Cohen's kappa, which were developed for studies of intramethod reliability or reproducibility, have been used [Maclure and Willett, 1987]. Cohen's kappa was meant to be an improvement to the overall proportion of agreement because it discounts the proportion of agreement that is expected by chance alone. However, because the interpretation of the result is not straightforward and Cohen's kappa is influenced by the prevalence of the investigated criterion [Li et al., 1996; Steijnmans et al., 1997], many studies still opted for the simple overall proportion of agreement as a measure for validity [Galasso et al., 1994; Riordan and Koehn, 1997; Grant et al., 1998]. Occasionally, contradictory results may be reached when different statistical measures are used.

Our prospective reproductive study in Beijing Yanshan Petrochemical Corporation (BYPC) offered an excellent opportunity to study the association between self-assessed exposure status, measurement, and professional assessment. Personal air samples were collected for 132 subjects to quantitatively assess their exposure to benzene, toluene, styrene, and xylene. Exposure questionnaires regarding their occupational exposures were also completed by the subjects. This enabled a comparison between self-assessed exposure and measurement data. In addition, plant area maps and process flow charts were collected for professional assessment of the subjects' exposures, which made it possible to examine the validity of the professional assessment.

The objectives of this paper are: (1) to examine the validity of self-assessed exposure using measurements of benzene, toluene, styrene, and xylene exposures as the gold standard; (2) to examine the validity of self-assessed exposure using an occupational hygienist's assessment of benzene, toluene, styrene, and xylene exposures as the gold standard; (3) to examine the validity of the occupational hygienist's assessed exposures relative to measured exposures; and (4) to compare the statistical measures used to assess the validity.

MATERIALS AND METHODS

General Description of the Study and the Population

One hundred thirty-two female subjects (age 20–40 years) who obtained birth permission from the family planning administration between February 1994 and August 1996 were recruited BYPC. The exposures of primary concern were benzene, toluene, styrene, and xylene. The adverse health outcomes included irregularities of menstrual cycle, prolonged time to pregnancy, spontaneous abortion, infertility, still birth, low birth weight, and birth defects [Cho et al., 2001]. All the subjects were of childbearing age. Before being employed in the plant, all the subjects were trained in the company-operated trade school for 2 years after senior high school as chemical plant operators, technicians, etc. Some job titles such as engineer require college or junior college degrees in petrochemical engineering. In general, this was a population with a reasonable knowledge of petrochemical operations.

Data collection on exposure consisted of three parts: (1) upon enrollment, each subject was interviewed using a questionnaire to assess occupational exposure in their workplace; (2) a full-shift personal air sample was collected for each subject upon enrollment; and (3) work area maps and flow charts for the manufacturing processes were collected for industrial hygiene assessment of the exposures.

Worker's Self-Assessed Exposure

Information on self-assessed exposure to benzene, toluene, styrene, and xylene was obtained by questionnaires administered to the subjects upon enrollment in the study by a doctor of occupational medicine from Beijing Medical University. The wording of the questions was, "Are you exposed to any of the listed chemicals in your current job? If yes, what are they?" For the chemicals identified by the subjects, we gave a "yes" to the exposure as the self-assessed exposure status. For those not mentioned by the subjects, we gave a "no" to the exposure as the self-assessed exposure status.

Quantitative Exposure Assessment

Quantitative exposure assessment was made by actual measurement of each subject's full-shift exposure. A full shift (6 hr) active air sample was collected from each subject using Gilair (Gilian[®]) personal sampling pumps and SKC[®] 50/100 charcoal tubes upon enrollment into the study. A small portion of subjects were sampled again after the pregnancy was confirmed. The samples were analyzed for aromatic solvents by the NIOSH 1500 method [NIOSH, 1994a]. A priori homogenous exposure zones [Corn and Esman, 1979], in which subjects were presumed to have

similar exposures, were established based upon the subjects' job activities and work area. The arithmetic mean of the exposures of the zone was assigned to all the members in the same zone as their estimate of exposure, based upon Corn and Esman method [Corn and Esman, 1979].

To make the comparisons compatible, the quantitative measurements were converted into a binary variable. Measurement results greater or equal to the limit of detection (0.004 ppm for benzene, 0.003 ppm for toluene, styrene, and xylene) were regarded as "exposed" whereas those under the limit of detection were regarded as "not exposed."

Professional Assessment

Professional assessment of exposure was based mainly on a subject's job activities and production processes without knowledge of the measurement results or self-assessed exposure status. In order to understand the production processes and work activities, each subject was interviewed to describe her job. In addition, process flow charts and work area layout maps of the facilities were obtained for each of the subject's work area.

The production processes that had sources for benzene, toluene, styrene, and/or xylene exposures were compiled by an industrial hygienist on the research team. In workshops where benzene, toluene, styrene, and xylene were present as the raw material or end products, all the subjects' exposures were rated as "yes." Examples of these workshops were Benzene Workshop in Chemical Plant #1, and the Dewaxing Workshop in the Refinery. A "yes" was also given to the subjects in workshops where benzene, toluene, styrene, and xylene were present in the mixture of the materials the subjects were handling. Examples were: Wastewater Treatment Plant, and Catalytic Cracking Process in the Refinery. Job titles such as equipment people who have uncertain working areas but who would definitely go to the exposed areas were also assigned a "yes."

Statistical Measures for Agreement

The agreement was measured by an overall proportion of agreement, Cohen's kappa, sensitivity, and specificity, which are defined as follows.

Reported exposed or professional assessment	Measurement or professional assessment	
	Yes	No
Yes	a	b
No	c	d

Sensitivity = $a/(a + c)$
 Specificity = $d/(b + d)$
 Overall proportion of agreement (P_o) = $a/(a + b + c + d) + d/(a + b + c + d)$
 Kappa = $(P_o - P_e)/(1 - P_e)$, where
 $P_e = (a + b)(a + c)/(a + b + c + d)^2 + (c + d)(b + d)/(a + b + c + d)^2$

The sensitivity and specificity assumes asymmetry between the two estimates [Maclure and Willett, 1987]. In other words, it assumes that among the exposure estimates to be compared, one is closer to the "truth" than another. When self-report or professional assessment was compared with the actual measurement, the actual measurement was assumed to be the "truth" or "gold standard." When self-report was compared with professional assessment, the professional assessment was assumed to be the "gold standard." For an ideal method, both sensitivity and specificity should be equal to 1.0, i.e., the method absolutely agrees with the truth and the misclassification items b and c are equal to 0. In reality, however, the ideal situation is very unlikely to happen. For a method to be considered as a good measure for exposure, the sum of sensitivity and specificity should be equal or greater than 1 [Armstrong et al., 1994].

The overall proportion of agreement and Kappa, however, make no such hierarchic assumptions. Originally, proposed as measures of agreement between two observers [Fleiss, 1981], the overall proportion of agreement and Kappa assume that self-report, professional assessment, and measurement are of equal validity to estimate a person's true exposure status. The overall proportion of agreement or Kappa emphasize on the agreement. Cohen's kappa can range from -1 to 1. If the agreement is greater than chance, kappa is greater than 0. To interpret the kappa value, "benchmarks" were suggested by Landis and Koch [1977]: 0.00 (poor), 0.00-0.20 (slight), 0.20-0.40 (fair), 0.40-0.60 (moderate), 0.60-0.80 (substantial), and 0.81-1.00 (almost perfect).

RESULTS

Agreement Between Self-Assessed Exposure and Measurement Results

Table I shows the measures of agreement between self-assessed exposure status and actual measurements. Except for toluene, the overall proportion of agreement between self-assessed exposure and measurement was fair (> 60%). The sensitivity of the self-assessed exposure on toluene,

TABLE I. Agreement Between Self Assessed Exposure Status and Measurement on Benzene, Toluene, Styrene, and Xylene, in Female Petrochemical Workers in China

	Overall proportion agreement	Sensitivity	Specificity	Cohen's kappa
Benzene	0.68	0.70	0.62	0.28
Toluene	0.06	0.02	0.83	-0.01
Styrene	0.85	0.39	0.94	0.38
Xylene	0.62	0.37	0.81	0.20

xylene, and styrene was very low (0.024) while the sensitivity of self-assessment of benzene exposure was good (0.70). By definition, the sensitivity is the proportion of subjects who thought they were exposed among the “truly” exposed subjects (i.e., those with detectable toluene exposures). A low sensitivity for toluene indicates that in this population, a significant portion of the subjects under-reported their exposures to toluene and xylene. The specificity, on the other hand, was relatively high for the four chemicals, ranging from 0.62 to 0.94. This implies that in this population, subjects who were not exposed to benzene, toluene, styrene, and xylene tended to regard themselves as “not exposed.”

The results also demonstrated that when self-report was compared to measurement, the statistical measures of agreement agreed with each other. All of the measures indicated that except for toluene, self-assessment for benzene, styrene, and xylene exposure agreed fairly well with the measurement results.

Agreement Between Self-Assessed Exposure and Professional Assessment

Table II shows the agreement between self-assessed exposure and professionally assessed exposure. Interestingly, the overall proportion of agreement for all the four chemicals were good, ranging from 0.58 to 0.91; whereas other measures of validity, sensitivity, specificity, and Cohen’s kappa indicated that the agreement between self-report and professional assessment was poor for toluene exposure. Compared to the agreement between self-assessment and measurement shown in Table I, the overall proportion of agreement between self-assessed toluene exposure and professional assessed toluene exposure was much higher (0.68). This implies that if professional assessment were used as “gold standard” and overall proportion of agreement were used as the measure for validity, we would conclude that self-assessment on toluene exposure was “good.” However, if we judge from sensitivity, specificity, and Cohen’s kappa, we would conclude that the agreement between self-assessment and professional assessment was poor.

TABLE II. Agreement Between Self Assessment and Professional Assessed Exposure on Benzene, Toluene, Styrene, and Xylene, in Female Petrochemical Workers in China

	Overall proportion agreement	Sensitivity	Specificity	Cohen’s kappa
Benzene	0.58	0.75	0.47	0.20
Toluene	0.68	0.03	0.97	-0.10
Styrene	0.91	0.58	0.96	0.60
Xylene	0.67	0.40	0.79	0.19

TABLE III. Agreement Between Professional Assessment and Measurement on Benzene, Toluene, Styrene, and Xylene, in Female Petrochemical Workers in China

	Overall proportion agreement	Sensitivity	Specificity	Cohen’s kappa
Benzene	0.55	0.47	0.76	0.17
Toluene	0.31	0.30	0.67	-0.004
Styrene	0.89	0.60	0.95	0.60
Xylene	0.61	0.57	0.64	0.21

Agreement Between Professional Assessment and Measurement Results

To understand the puzzling contradiction on toluene exposure, we examined the validity of the professional assessment. Table III lists the measures of the agreement between professional assessment and measurement results. The results indicated fair to good agreement between professional assessment and measurement for benzene, styrene, and xylene exposures. However, all the measures for agreement indicated a poor agreement between professional assessment and measurement of toluene exposure.

Table IV is a closer examination of the relationship between self-assessment, professional assessment, and actual measurement of the toluene exposure. A substantial number of false negative (122) was found when self-assessment was compared with actual measurements. Meanwhile, the false negative was equally significant (88) when professional assessment was compared with actual measurements. In other word, both professional assessment and the self-assessment on toluene were biased towards the same direction (underestimate). Therefore, when a biased estimate (self-assessment) was compared with a biased “reference” (professional assessment), we would obtain a misleading overall proportion of agreement.

DISCUSSION

Validity of Self-Assessment

Several previous studies by other researchers examined the validity of self-assessment on dust exposure and concluded the usefulness of self-assessment [diPico, 1982; Rom et al., 1983; Fonn et al., 1993]. However, few studies have investigated the validity of self-assessment of chemical exposures. Because the dust exposure can be easily “seen” whereas solvent exposure cannot, it is interesting to know the validity of self-assessment of solvent exposures. Such information is important for epidemiologic studies, because in most industrial settings that involve solvent exposures, workers tend to be exposed to several solvents simultaneously. Because making actual measurement for each

TABLE IV. Comparison of 2 × 2 Tables of Self Assessment Vs. Professional Assessment Vs. Measurement on Toluene

Self assessed exposure	Measured exposure		Self assessed exposure	IH assessed exposure		IH assessed exposure	Measured exposure	
	Exposed	Unexposed		Exposed	Unexposed		Exposed	Unexposed
Exposed	3	1	Exposed	1	3	Exposed	37	2
Unexposed	122	5	Unexposed	39	88	Unexposed	88	4

solvent is costly and difficult, self-administered questionnaires often become the method of choice for comprehensive exposure assessment.

This study demonstrated that most workers with reasonable knowledge of production processes could provide valid information on their exposures to benzene, styrene, and xylene. However, under-reporting of exposure appeared to be a major problem in self-assessed exposure status. The proportion of subjects who under-reported their exposure to benzene, styrene, and xylene were 12, 6, and 18%, respectively. The under-reporting rate was very higher (92%) for toluene.

Underestimation has been observed as a problem in studies where self-assessment of exposures was used to obtain data on activities that are not socially acceptable such as smoking or alcohol abuse [Wagenknecht et al., 1992; Crook et al., 1998]. Our results indicated that underestimation of exposure was also a problem for self-reported solvent exposures. A similar trend was also observed in studies that used self-assessment for chemical exposures [Ahlborg, 1990; Lindbohm et al., 1990]. In a study of validity of self-assessed exposure to various categories of chemical exposures, Fritschi et al. [1996] found relatively low sensitivity for self-assessment on solvent exposures, which also implied a potential problem for underestimating of self-assessed exposure status.

Because non-differential underestimation of exposure can lead to a bias towards or away from the null-value [Verkerk and Buitendijk, 1992], care should be taken when self-assessed exposure is used in epidemiologic studies.

Factors Influencing Validity of Self-Assessed Exposure

It is possible that the validity of self-assessed exposure was influenced by the odor and the odor threshold of the substances. Among these four substances, styrene has a characteristic sharp, unpleasant penetrating odor while benzene, toluene, and xylene have mild hydrocarbon odors. The odor thresholds for benzene, toluene, styrene, and xylene are 1.5–4.7 ppm [ATSDR, 1991], 8 ppm [ATSDR, 1994], 0.017–1.9 ppm [NIOSH, 1994b], and 1 ppm [NIOSH, 1994b], respectively. With a distinct odor and lowest odor threshold, it is not surprising that styrene achieved the best self-assessment. By the same token, the

poor performance of self-assessment on toluene may not merely be caused by chance alone, given the fact that its odor threshold is the highest among the four.

Another important factor that influenced the validity of self-assessed exposure may be the worker's awareness of the toxicity of the solvents. Among these four solvents, benzene is a carcinogen classified by International Agency for Research on Cancer [1982]. Precaution in handling benzene was a part of safety training for the workers in BYPC. As a result, the subjects might be more alert to its exposure. Furthermore, when benzene was the primary agent in the process, it was named in the area or process, such as the benzene workshop, which might also alert the workers of potential exposures.

Validity of Professional Assessment

When measurement is not available, professional assessment is widely used in epidemiologic studies and as the "gold standard" in validity studies of self-assessed exposure status. Because professional assessment was normally based upon production processes and job histories provided by the employer, the assessment can be affected by many factors such as quality of the reported information, rater's familiarity with the job and industry, the chemical properties of the substances and the actual setting of the exposure environment [Post et al., 1991; Stewart and Stewart, 1994]. Our study found that the Cohen's kappa between professional assessment and measurement ranged from –0.005 to 0.60, indicating that the misclassification for professional assessment of exposure can be substantial for some chemicals. This result agrees with the study by Kromhout et al. [1987], in which the R^2 between professional assessment and measurements ranged from 0 to 0.67, indicating no to substantial agreement. Because the professional assessment can be biased, false validity can be achieved when self-assessed exposure is compared with a professional assessment that is biased towards the same direction. We, therefore, caution the use of professional assessment as "gold standard" in validity studies.

Another point worth mentioning is that contrary to self-assessment which was biased because of underestimation, the disagreement between professional assessment and measurement in this study was caused by either underestimation or overestimation. While the disagreement for

benzene, toluene, and styrene was caused by underestimation, the disagreement for xylene was caused by overestimation of the exposures.

Choice of Measures to Quantify Agreement

There are two kinds of reliability studies: intramethod reliability studies that examine the reproducibility of a method, and the intermethod reliability studies that compare two different methods. When one of the methods is assumed to be close to the “truth,” the intermethod reliability study is also referred to as a validity study [Armstrong et al., 1994]. The study of the validity of self-assessment of exposure belongs to the later type. Unfortunately, less work has been done on the interpretation of intermethod reliability studies. In this paper, we calculated four different statistical measures for validity or reproducibility. In most of the cases, the conclusions obtained from these measures were consistent with each other. However, when self-assessed exposure status was compared with professionally assessed exposure status, a contradictory conclusion was reached by these measures. While Cohen’s kappa, sensitivity, and specificity indicated that the agreement between self-assessed exposure status and professionally assessed exposure status were poor, good agreement was indicated by the overall proportion of agreement for toluene.

This is a classic case of fallacy of overall proportion of agreement. Because it does not account for the percent of agreement expected by chance, the overall proportion of agreement does not perform well when the prevalence of investigated criterion is small [Fleiss, 1981]. In our case, the prevalence of toluene exposure reported by the workers themselves was low (only 4), a large portion of the agreement was caused by the agreement on negatives. High prevalence of the criterion of interest can also cause similar contradictions [Steinijans et al., 1997].

Cohen’s kappa, on the other hand, discounts the agreement caused by chance. Therefore, it performs better when the prevalence of the investigated criterion are small. However, interpretation of the result is not straightforward compared to the overall proportion of agreement. To compensate the disadvantages of each measure, we recommend using multiple statistical measures for a validity study, including overall proportion agreement, Cohen’s kappa, sensitivity, and specificity.

ACKNOWLEDGMENTS

This study is supported by NIOSH grant OH03027 and NIH ES00002. The authors are grateful to Dr. Robert Herrick for helpful suggestions concerning the assessment of confounding exposures, Mark Tungsheng Shih for spike samples, Dr. Koichi Kono and Dr. Yukio Orita for sample

measurement and researchers in Beijing Anti-epidemic Station for assisting sample collection.

REFERENCES

- Agency for Toxic Substance and Disease Registry (ATSDR). 1991. Toxicological Profile for Benzene. 1 p.
- Agency for Toxic Substance and Disease Registry (ATSDR). 1994. Toxicological Profile for Toluene. 2 p.
- Ahlborg GA. 1990. Validity of exposure data obtained by questionnaire—two examples from occupational reproductive studies. *Scand J Work Environ Health* 16:284–828.
- Armstrong B, Write E, Saracci R. 1994. Principles of exposure measurement in epidemiology (monographs in epidemiology and biostatistics, Vol 21). New York: Oxford University Press.
- Birdsong WH, Lash AA, Thayer S, Kumenkawa E, Becker CE. 1992. The validity of study group assignments based on occupational histories obtained from questionnaires. *J Occup Med* 34(9):940–945.
- Cho SL, Damokosh AI, Ryan LM, Chen D, Hu Y, Smith TJ, Christiani DC, Xu X. 2001. Effects of exposure to organic solvents on menstrual cycle length. *J Occup Environ Med* 43:567–575.
- Corn M, Esman NA. 1979. Workplace exposure zones for classification of employee exposure to physical and chemical agents. *Am Ind Hyg Assoc J* 40:47–57.
- Crook G, West M, Oei T. 1998. A comparison of estimates of the prevalence of heavy drinkers in local regions of Queensland. *Aust N Z J Psychiatry* 32:504–510.
- diPico GA. 1982. Epidemiologic basis for dose-response criteria. *Ann Am Conf Gov Ind Hyg* 2:189–195.
- Eskenazi B, Pearson K. 1988. Validation of a self-administered questionnaire for assessing occupational and environmental exposures of pregnant women. *Am J Epidemiol* 128(5):1117–1129.
- Fleiss J. 1981. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley and Sons.
- Fonn S, Groeneveld H, deBeer M, Becklake M. 1993. Relationship of respiratory health status to grain dust in a Witwatersrand grain mill: comparison of workers’ exposure assessments with industrial hygiene survey findings. *Am J Ind Med* 24:401–411.
- Fritschi L, Siemiatycki J, Richardson L. 1996. Self-assessed versus expert-assessed occupational exposures. *Am J Epidemiol* 144:521–527.
- Galasso R, Panico S, Celentano E, del Pezzo M. 1994. Relative validity of multiple telephone versus face-to-face 24 dietary recalls. *Ann Epidemiol* 4:332–336.
- Grant R, Gregor M, Maio R, Huang S. 1998. The accuracy of medical records and police reports in determining motor vehicle crash characteristics. *Prehosp Emerg Care* 2:23–28.
- International Agency for Research on Cancer. 1982. Monographs on the evaluation of the carcinogenic risk of chemicals to humans. Vol 29. p 93–148.
- Kromhout H, Oostendorp Y, Heederik D, Boleij J. 1987. Agreement between semiquantitative exposure estimates and quantitative exposure measurements. *Am J Ind Med* 12:551–562.
- Landis J, Koch G. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33:159–174.
- Li R, Cai J, Tegeler C, Sorlie P, Metcalf PA, Heiss G. 1996. Reproducibility of extracranial carotid atherosclerotic lesions assessed by B-mode ultrasound: the Atherosclerosis Risk in Communities Study. *Ultrasound Med Biol* 22:791–799.

- Libscomb JA, Fenster L, Wrensch M, Shusterman D, Swan S. 1991. Pregnancy outcomes in women potentially exposed to occupational solvents and women working in the electronic industry. *J Occup Med* 33(5):597–604.
- Lindbohm M, Taskinen H, Sallmen M, Hemminki K. 1990. Spontaneous abortion among women exposed to organic solvents. *Am J Ind Med* 17:449–463.
- Maclure M, Willett WC. 1987. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol* 126(2):161–169.
- National Institute for Occupational Safety and Health. 1994a. NIOSH Manual of Analytical Methods, 3rd Ed on Diskette. Hamilton: Canadian Centre for Occupational Health and Safety.
- National Institute for Occupational Safety and Health. 1994b. RTECS. CINFO disc C2 (94-4). Hamilton, Canada: Canadian Centre for Occupational Health and Safety.
- Post W, Kromhout H, Heederik D, Noy D, Smit Duijzentkunst R. 1991. Semiquantitative estimates of exposure to methylene chloride and styrene: the influence of quantitative exposure data. *Appl Occup Environ Hyg* 3:197–204.
- Riordan J, Koehn M. 1997. Reliability and validity testing of three breastfeeding assessment tools. *J Obstet Gynecol Neonatal Nurs* 26:181–187.
- Rom WN, Moshell A, Greaves W, Bang KM, Holthouser M, Campbell D, Bernstein R. 1983. A study of dermatitis in trona miners and millers. *J Occup Med* 25:295–299.
- Rosenberg CR, Mulvihill MN, Fischbein N, Blum S. 1987. A analysis of the validity of self reported occupational histories using a cohort of workers exposed to PCBs. *Br J Ind Med* 44(10):702–710.
- Rosenstock L, Logerfo J, Heyer N, Carter W. 1984. Development and validation of a self-administered occupational health history questionnaire. *J Occup Med* 26:50–54.
- Steinijans VW, Diletti E, Bomches B, Solleder P. 1997. Interobserver agreement: Cohen's kappa coefficient does not necessarily reflect the percentage of patients with congruent classifications. *Int J Clin Pharmacol Ther* 35(3):93–95.
- Stewart WF, Steward PA. 1994. Occupational case-control studies: I. Collecting information on work histories and work-related exposures. *Am J Ind Med* 26(3):297–312.
- Taskine H, Anttila A, Lindbohm M, Sallmen M, Hemminki K. 1989. Spontaneous abortions and congenital malformations among the wives of men occupationally exposed to organic solvents. *Scan J Work Environ Health* 15:345–352.
- Verkerk P, Buitendijk S. 1992. Non-differential underestimation may cause a threshold effect of exposure to appear as a dose-response relationship. *J Clin Epidemiol* 45:543–545.
- Wagenknecht L, Burke G, Perkins L, Haley N, Friedman G. 1992. Misclassification of smoking status in the CARDIA study: a comparison of self-report with serum cotinine levels. *Am J Public Health* 82:33–36.