

## RADIATION EXPOSURE FROM DIAGNOSTIC IMAGING: AGREEMENT BETWEEN SELF-REPORT AND MEDICAL RECORDS

Janice M. Pogoda\* and Susan Preston-Martin†

**Abstract**—Data on diagnostic imaging procedures from a highly structured interview were compared to medical records in a case-control study of radiography and acute myelogenous leukemia. Three hundred and twenty-eight cases and 315 controls (78% of participants) had medical records available from an average of 71% of providers. Proxies were used for 49% of cases because of rapid fatality. Mean agreement (number of procedures in medical records subtracted from number in interview) showed similar levels of overreporting in cases [0.6; 95% confidence interval (CI): 0.0, 1.1] and controls (0.7; CI: 0.2, 1.3) with few procedures ( $\leq 10$ ). Most participants with more procedures underreported exposure, and underreporting increased with increasing exposure, especially among cases [mean (CI) agreement =  $-2.1$  ( $-4.3, 0.0$ ) for 11–20 procedures,  $-6.4$  ( $-13.6, 0.7$ ) for  $>20$  procedures] and case proxies. High-dose, fluoroscopic, and non-routine procedures were self-reported more accurately than low-dose, non-fluoroscopic, and routine procedures, respectively ( $p < 0.01$  for each comparison), and tended to be underreported. Case-control differences in agreement were non-significant for these categories of procedures. We conclude that diagnostic imaging procedures of most interest to the AML-radiography hypothesis are self-reported accurately but that underreporting does occur and might lead to attenuated risk estimates.

Health Phys. 83(6):907–917; 2002

**Key words:** imaging; medical radiation; radiography; leukemia

### INTRODUCTION

DATA COLLECTED in interviews on history of medical procedures can be difficult to recall and may be more susceptible to inaccuracy than other types of self-reported exposures. Inaccuracies that are similar for

cases and controls, “nondifferential” inaccuracies, usually result in attenuated risk estimates and a loss of statistical power to detect true associations between exposure and risk. Inaccuracies that are systematically different in cases than in controls, “differential” inaccuracies, can either mask true associations or create spurious associations. The data collection instrument may play a key role in maximizing the quality of the data collected. In this paper, we compare self-reported data on diagnostic imaging procedures collected using highly specific interview questions and carefully devised probes to data from subjects’ medical records.

Several previous studies have compared self-reported data to medical records. Exposures typically studied include pregnancies and deliveries; chronic illnesses; hospitalizations, surgeries, and other events related to medical care; and medication use, particularly estrogens and oral contraceptives (Harlow and Linet 1989). Although results reported to date are widely varied and used different data collection instruments, there were several common observations. Unique and memorable events and unambiguous conditions were self-reported with more accuracy than frequently-occurring events and conditions with less specific criteria for diagnosis. For example, among participants in a case-control study of prostate cancer, past histories of kidney stones were more accurately reported than urinary tract infections (Zhu et al. 1999). Events that were more positive and/or more salient to respondents, i.e., more prominent or of more notable significance, were self-reported with greater accuracy than less salient and/or more negative events. For example, newborn deliveries were reported more accurately than hospitalizations (Harlow and Linet 1989). Self-report inaccuracies tended to be due to over- rather than underreporting (Paganini-Hill and Ross 1982; McKenna et al. 1992; Hiatt et al. 1995; Zhu et al. 1999), and more accurate reporting occurred when exposures were dichotomous rather than quantitative (Spitz et al. 1988). Although there was some evidence of differential accuracy, most notably in studies

---

\* Statology, 3115 Aaron Drive, Rocklin, CA 96161; † Keck School of Medicine, University of Southern California (USC), Department of Preventive Medicine, USC/Norris Comprehensive Cancer Center, 1441 Eastlake Avenue, MS 44, P.O. Box 33800, Los Angeles, CA 90033-0800.

For correspondence or reprints contact: J. M. Pogoda, Statology, 3115 Aaron Drive, Rocklin, CA 96161, or email at [jpogoda@statology.com](mailto:jpogoda@statology.com).

(Manuscript received 11 September 2001; revised manuscript received 18 February 2002, accepted 13 July 2002)

0017-9078/02/0

Copyright © 2002 Health Physics Society

of pregnancy outcome, overall there was little proof of its existence (MacKenzie and Lippman 1989; Chouinard and Walter 1995; Basso et al. 1997; Zhu et al. 1999).

Given the characteristics of exposures associated with inaccurate reporting, self-reported histories of diagnostic imaging procedures are likely to be especially problematic because of the following: 1) procedures are often not well-defined or clearly understood by patients; 2) many procedures, such as chest x rays, are routinely performed and are not usually associated with salient events, 3) diagnostic procedures are typically not associated with positive feelings, and 4) quantitative exposure is likely to be required for most hypotheses involving this type of exposure. Only a few validation or agreement studies have specifically addressed imaging procedures, and fewer still have considered imaging procedures as a quantitative variable. A very early study found that respondents failed to report 80% of all diagnostic x rays recorded in their medical records, though this was likely due to methodological limitations (Graham et al. 1963). Among Los Angeles parotid tumor cases and their matched controls, both cases and controls overreported the number of dental x-ray visits and underreported panoramic examinations compared to their dental records (Preston-Martin et al. 1985). Among new enrollees to a Harvard community health maintenance organization, accuracy of self-reported dichotomous exposure to cancer screening procedures depended on the procedure; accurate reporting was observed for chest x rays, mammograms, and electrocardiograms but not serum cholesterol tests or testicular self-examinations (Betz Brown and Adam 1992). Most inaccuracies were due to overreporting. Among respondents from a rural Japanese community, there was substantial accuracy in self-reported frequency of stomach cancer screening; inaccuracies that did occur were due to overreporting (Tsubono et al. 1994). Similarly, inaccuracies in self-reported mammography and Pap smear screening frequency compared to medical records are typically attributable to overreporting by respondents (Whitman et al. 1993).

This study used data from a Los Angeles case-control study of acute myelogenous leukemia (AML) specifically designed to evaluate risk associated with medical radiography, i.e., all diagnostic imaging procedures that expose the red bone marrow to ionizing radiation. Because radiography was the focus of the study, substantial efforts were made to maximize the validity of radiography exposure data obtained by subject interviews and to evaluate the adequacy of structured interviews for future studies involving radiography exposure. To this end, medical records were obtained from health care providers named by subjects in interviews so

that details related to past radiographic procedures could be abstracted and absence of past radiography could be noted. This report presents results from an analysis of agreement between subject interview and medical records on past exposure to diagnostic (both radiographic and non-radiographic) imaging procedures.

## MATERIALS AND METHODS

### Subjects

Subjects were participants in a population-based case-control study of AML designed to evaluate relative risk associated with medical radiography in the 10 years before diagnosis. Cases diagnosed with adult-onset AML (ICDO codes 9861, 9864, 9866, 9867 and 9891) in Los Angeles County from January 1987 through June 1994 were identified by the University of Southern California (USC) Cancer Surveillance Program, a population-based Surveillance, Epidemiology, and End Results cancer registry. Cases who met eligibility criteria were English-speaking (expanded to Spanish-speaking in 1992) non-Latino whites or African-Americans between 30 and 69 y old who resided in the U.S. during the previous 15 years and in Los Angeles County at the time of their diagnosis. A lower age limit was used because the distribution of active bone marrow changes around age 20 y (shifts from limbs to trunk). An upper age limit was used because 1) AML is more rapidly fatal in the elderly, 2) it is more difficult to find matched controls for elderly cases, and 3) recall can be more difficult for elderly respondents. The residence criteria was used because of our interest in radiography exposure specifically during the 10 years before diagnosis. Cases for whom physician consent was received and who were living and well enough to be interviewed were asked to participate in the study. Surviving spouses or other adults who had lived in cases' households for any 6 of the 10 years preceding diagnosis were used as proxies for cases who were deceased or too ill to be interviewed. Additionally, proxies were interviewed for a subset of cases who were directly interviewed to allow comparisons within pairs of direct and proxy respondents. Each case was matched to a neighborhood control by birthyear (within 5 years), race (black or white), and sex according to a previously established protocol (Preston-Martin et al. 1980). Controls were also required to speak English (or Spanish if matched to a Spanish-speaking case) and to have resided in the U.S. during the previous 15 years. Diagnosis dates of cases to whom they were matched were used as reference dates for controls (hereafter referred to as "diagnosis date" for both cases and controls). Interviews were also conducted with proxy respondents for most controls. The study proposal and method of obtaining informed consent from

study participants were approved by the USC Institutional Review Board.

### Interviews

Interviews were conducted from 1987 to 1997. Although it was not practical to blind interviewers to disease status, they were trained to conduct interviews in a standardized, highly structured manner; i.e., they followed a scripted form for asking specific, closed-ended questions so that all respondents were given the same interview. Matched pairs were interviewed by the same person. Separate questionnaires were used for personal and proxy respondents so that the scripted wording would be appropriate.

Ascertainment of diagnostic imaging history was the major component of the 45-min in-person interview that also included sections on prescribed medications, disease symptoms, radiation treatment, occupational exposures, tobacco and alcohol use, hereditary diseases, and cancer chemotherapy. Respondents were systematically queried on imaging procedures within each of 10 categories of anatomical sites (Table 1). For each procedure, they were asked 1) if they (or, for proxy respondents, if the index subject) had ever had the procedure, 2) the first year in which they had the procedure, 3) total number of times they had had the procedure prior to the diagnosis date, and 4) if they had had the procedure in the 10 years before the diagnosis year (including the diagnosis year). If they reported having had the procedure in the 10-y period before diagnosis, they were further asked 1) the exact body part examined, 2) the date of the procedure, 3) the reason for the procedure, 4) the exact name of the procedure, 5) whether or not fluoroscopy (imaging of moving structures with the aide of a contrast agent) was used, 6) the type of facility at which the procedure was performed, and 7) the name, address, and specialty of the health care provider who ordered the procedure. For the subset of subjects with diagnosis year 1991–1994, respondents were also asked to name all health care providers they had seen during the 10 years before diagnosis and were given specific prompts asking for provider names in the interview sections on prescribed medications, disease symptoms, radiation treatment, hereditary diseases, and cancer chemotherapy. After the interview, respondents were asked to sign consent forms granting us permission to obtain information from their medical records.

### Medical records

Contact was attempted with every provider named in subjects' interviews using aggressive mail and phone techniques. Each provider located was mailed the signed consent form and a form for abstracting details on

**Table 1.** Diagnostic imaging procedures queried within anatomical site categories of medical radiography history interview;<sup>a</sup> case-control study of AML, Los Angeles County, January 1987–June 1994.

Anatomical site	Diagnostic procedure	
Esophagus/stomach/intestines	Upper gastrointestinal	
	Lower gastrointestinal	
	Plain stomach/abdomen x ray <sup>b</sup>	
Pancreas/gallbladder/biliary system	Endoscopic retrograde cholangiopancreatography	
	Oral cholecystogram	
	Transhepatic cholangiogram	
	T-tube cholangiogram	
Chest/lungs	Chest x ray	
Breasts	Mammogram	
Heart/veins/arteries	Coronary angiogram	
	Pulmonary arteriogram	
	Renal arteriogram	
	Carotid arteriogram	
	Abdominal arteriogram	
	Pelvic arteriogram	
	Arteriogram of any other site	
	Venogram	
	Kidneys/ureters/bladders	Entire system
		Bladder only
		Plain x ray (no dye)
Uterus/ovaries	Hysterosalpingogram	
	Plain pelvic x ray	
	Pelvimetry	
Joints	Knee arthrogram	
	Shoulder arthrogram	
	Hip arthrogram	
Bones	Plain x ray, skull	
	Plain x ray, facial bones	
	Plain x ray, nose/sinuses	
	Plain x ray, jaw	
	Plain x ray, trachea	
	Plain x ray, collarbone	
	Plain x ray, shoulder	
	Plain x ray, ribs	
	Plain x ray, neck/spine	
	Myelogram	
	Plain x ray, hip/pelvis	
	Plain x ray, arm/wrist/hand	
	Plain x ray, leg/knee/feet	
	General <sup>c</sup>	Nuclear medicine scan
Computerized tomography scan		
Magnetic resonance imaging		
Positron emission tomography scan		
Thermogram		
Other		

<sup>a</sup> For procedures reported to have occurred during the 10 years before diagnosis, follow-up questions were asked regarding the exact body part examined, the date of the procedure, the reason for the procedure, the exact name of the procedure, whether or not a contrast medium or fluoroscopy was used, the type of facility at which the procedure was performed, and information on the health care provider who ordered the procedure.

<sup>b</sup> "Plain x ray" refers to a standard x ray examination that does not involve fluoroscopy, use of a contrast medium, or any other special procedure.

<sup>c</sup> This questionnaire section did not ask about a specific anatomical site but rather about procedures that could be performed on any site.

diagnostic imaging procedures undergone by the subject from their records. Despite the legal requirement in California of a 7-y retention period for such records, the

time period recommended by the American Health Information Management Association is 10 y (Fletcher 1999), and it has generally been observed that providers retain records for even longer time periods. The study coordinator maintained “abstract forms” for each subject’s providers specifying such details as the name of the provider or facility, the provider type (e.g., physician, hospital, health maintenance organization), the provider status (e.g., out of practice), a link to the interview questionnaire (i.e., why the provider was contacted), and specifics of all imaging procedures undergone by the subject abstracted from the provider’s records (anatomical site, procedure name, date, whether or not fluoroscopy was used). To supplement lists of providers reported by study subjects, all providers were also asked if they were aware of other providers who had treated the patient (e.g., through referrals to or from) or to whom the patient’s records had been transferred. When records had been transferred from the original provider to a new provider, the new provider was sent the consent form and asked to provide the necessary information.

### Dosimetry

An estimate of red bone marrow radiation dose associated with each diagnostic radiographic procedure reported by study participants or recorded in their medical records was developed (Preston-Martin and Pogoda 2003). Briefly, a database of dose estimates for each procedure was created through literature searches and other resources (e.g., personal communications with radiology experts), and a median red bone marrow dose was calculated for each procedure. These dose estimates were used to create categories of procedures defined by dose levels (e.g., high vs. low).

### Statistical analysis

Throughout the remainder of this manuscript, “exposure” refers to number of diagnostic imaging procedures and “agreement” refers to the difference between self-reported and medical record data. Agreement was analyzed based on total number of procedures, or subsets of procedures, rather than on a per-procedure basis. For example, one approach to assessing agreement might be to determine whether or not each procedure reported in interview was noted in medical records and to arrive at an overall description of agreement, such as a kappa statistic. However, this method was found to be impractical for two reasons: 1) Matching self-reported procedures to procedures in medical records proved to be overly cumbersome if not impossible because of subtle differences and overlap in procedure names, types, and dates; and 2) Etiologically, the risk factor of interest in the

AML-radiography hypothesis is total bone marrow radiation dose, and total number of procedures (or subsets of procedures) is a logical surrogate.

For each subject, agreement was assessed between self-reported exposure and that recorded in medical records by subtracting the number of procedures in medical records from self-reported number of procedures (Bland and Altman 1986); thus, negative differences represented “underreporting” by the respondents compared to medical record data. Throughout this manuscript, “underreporting” refers to fewer self-reported procedures than the number found in medical records and “overreporting” refers to more self-reported procedures than in medical records, even though the self-reported number may be accurate and the number from medical records may be erroneously low. Each subject’s total exposure was measured by averaging number of procedures they reported in interview and number of procedures in their medical records. Subjects for whom medical record data were incomplete were excluded only if their available medical records contained no procedures. Subjects who reported no procedures and no providers were also excluded since it was impossible to assess agreement with no medical records for comparison; i.e., assessing agreement would have entailed searching every possible provider’s records to see if these subjects were in any provider’s records. In contrast, for subjects who did report providers, medical records from the providers they reported as well as the providers themselves were informative resources for missed self-reported data; no such resources were available for subjects with no self-reported data.

Case-control differences were tested using Fisher’s exact test for proportions and *t*-tests or Wilcoxon rank sum tests for continuous variables (total exposure and agreement). Analysis of covariance with total exposure as the covariate was used for all analyses of agreement except for those stratified by total exposure. For analyses involving more than one observation per subject, such as comparisons of agreement by different types of procedures, “subject” was also used as a covariate to account for within-subject correlation. Covariate-adjusted results are reported as least-squares means and 95% confidence intervals (CIs) (SAS/STAT User’s Guide 1988). Outcome variables were evaluated for normality and, when group comparisons were required, for equality of variance between groups. When these assumptions were not met, results were compared between standard parametric and rank-transformation analyses (Conover and Iman 1981, 1982). If standard and rank-based results substantially differed, the 25th and 75th percentiles (the “mid-spread”) of the outcome variable were reported as well as

means and CIs for the ranks. All tests were two-sided with 0.05 significance levels.

## RESULTS

Of 726 eligible cases, 188 (26%) were deceased or too ill for interview and had no available proxy, 31 (4%) were not contacted as advised by their physicians, 21 (3%) were lost to follow-up, and 74 (10%) refused to participate. Therefore, the study included 57% (412/726) of originally identified cases or 85% (412/487) of cases invited to participate. Interviews with proxy respondents were conducted for 201 (49%) deceased cases; further, proxy interviews were conducted for 101 cases and 174 controls who were also interviewed directly (comparisons between direct and proxy respondents to be published later). This analysis uses interview data from "best respondents;" i.e., direct respondents except for those cases for whom proxies were the only available respondents.

1,574 and 1,446 providers were reported by 395 case and 389 control respondents, respectively (Table 2). Complete records were successfully received from 1,093 case providers (69%) and 974 control providers (67%). On a per-subject basis, record retrieval was successful from an average of 71% of a subject's providers for both cases and controls. The most common reasons for unsuccessful record retrieval were 1) the provider claimed that they had no records for the subject (13% of all providers) and 2) inadequate information was available to locate the provider (13%). Record retrieval status for providers did not generally differ by case-control status; the largest discrepancy was that more control than case providers were retired or deceased or their facilities had

been closed (Table 2; 2.1% of control providers vs. 0.9% of case providers).

Medical record data were ultimately available for 330 cases and 316 controls. Among the 330 cases, 94% of 2,987 procedures reported in interview and 98% of 3,042 procedures recorded in medical records were in the 10-y period before diagnosis that was specifically asked about. Among the 316 controls, 94% of 2,666 procedures reported in interview and 98% of 2,485 procedures recorded in medical records were in the 10-y period before diagnosis. All procedures were included in the agreement analysis, whether or not they were dated within the 10-y pre-diagnosis period. This was done to prevent falsely inflated disagreement due to minor disparate dating of procedures. Dates of procedures ranged from 2.8 y after diagnosis to 12.3 y before diagnosis in self-reported data and from 2.25 y after diagnosis to 14.5 y before diagnosis in data from medical records. Two cases and one control were excluded from analyses due to extreme disparity between number of self-reported procedures ( $\geq 70$ ) and number of procedures recorded in medical records ( $\leq 8$ ). All three of these subjects were personally interviewed and there was no obvious reason for the discrepancies. Thus, the agreement analysis ultimately included 328 cases and 315 controls (Table 3).

Agreement between self-reported and medical record exposure depended on total exposure (Fig. 1). 222 cases and 233 controls with low exposure ( $\leq 10$  procedures) tended to overreport exposure [mean (CI) agreement = 0.6 (0.0, 1.1) for cases, 0.7 (0.2, 1.3) for controls;  $p = 0.72$  for case-control difference]. Among 66 cases and 81 controls with moderate exposure (11 to 20 procedures), cases were more likely than controls to underreport exposure [mean (CI) agreement = -2.1

**Table 2.** Final record retrieval status for providers named by 395 case and 389 control respondents; case-control study of AML, Los Angeles County, January 1987–June 1994.

Provider status	Case providers		Control providers	
	#	%	#	%
Complete record received from provider	1,093	69.4	974	67.4
Incomplete/inadequate file received from provider	13	0.8	22	1.5
Not enough information available to locate records	196	12.5	186	12.9
Provider retired/deceased or facility closed	14	0.9	30	2.1
Provider's records destroyed	24	1.5	31	2.1
Provider has no record of patient	205	13.0	189	13.1
Provider refused or charged excessive fee for file	17	1.1	8	0.6
Unable to obtain current patient consent	12	0.8	5	0.3
Unresolved	0	0.0	1	0.1
Total providers	1,574	100.0	1,446	100.0

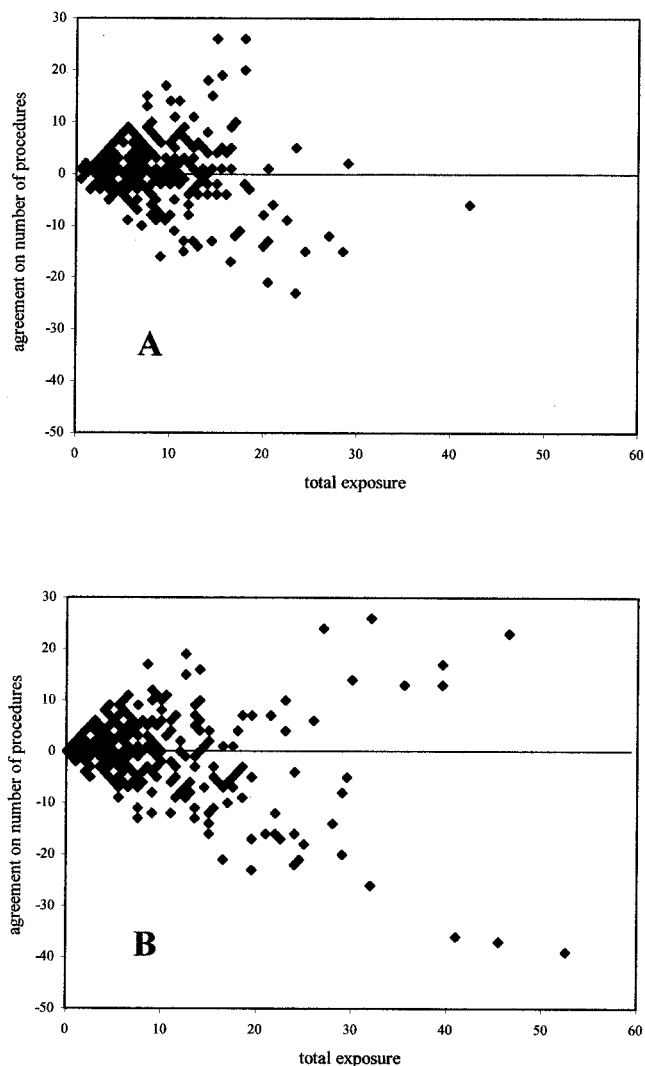
**Table 3.** Number of cases and controls included in analysis of agreement between self-report and medical records on diagnostic imaging procedures in the 10 years before diagnosis; case-control study of AML, Los Angeles County, January 1987–June 1994.

	# Cases	# Controls
Interviewed	412	412
Reported no exams and no providers in interview, thus no medical records	17	23
Reported exams in interview but no medical records received	6	18
No exams recorded in medical records due to incomplete files	59	55
Extreme disparity between number of procedures in interview compared to medical records	2	1
Total included in agreement analysis	328	315

(-4.3, 0.0) for cases, 0.9 (-1.1, 2.8) for controls;  $p = 0.06$  for case-control difference]. Only 29 cases and 12 controls had more than 20 procedures and largely under-reported exposure [mean (CI) agreement = -6.4 (-13.6, 0.7) for cases, -9.3 (-15.0, -3.7) for controls;  $p = 0.83$  for case-control difference]. Similar patterns of agreement by exposure level were evident for both direct and proxy case respondents, although proxies under-reported to a greater extent at the higher exposure levels (Table 4). Predictably, the subset of subjects with diagnosis year 1991–1994 for whom respondents were asked to name all health care providers (rather than just those associated with diagnostic imaging procedures) had higher total exposure and greater underreporting than subjects with earlier diagnoses [median (midspread) exposure = 7.0 (4.0, 12.5) for 1991–1994, 6.5 (3.5, 10.5) for earlier years,  $p = 0.08$ ; mean (CI) agreement = -0.5 (-1.3, 0.3) for 1991–1994, 0.2 (-0.5, 1.0) for earlier years,  $p = 0.20$ ]. The pattern of increasing underreporting with increasing exposure was the same for both groups defined by diagnosis year.

As shown in Fig. 2, agreement varied by type of procedure ( $p < 0.0001$ ) and anatomical site ( $p < 0.0001$ ). Mammograms, chest x rays, and procedures involving the blood or lymphatic systems were most likely to be overreported by respondents. Gastrointestinal (GI) series and regular x rays not involving the spine were also overreported but to a lesser degree. For most procedures and sites, agreement was similar for cases and controls. Cases were somewhat more likely than controls to overreport blood/lymphatic system procedures after controlling for numbers of these types of procedures [LS mean (CI) agreement = 0.6 (0.1, 1.1) for cases, -0.1 (-0.6, 0.4) for controls], particularly among direct case respondents [LS mean (CI) agreement = 1.0 (0.1, 1.9) for direct, 0.4 (-0.3, 1.1) for proxies].

Procedures were divided into five subsets that are most relevant to the AML-radiography hypothesis and,



**Fig. 1.** Agreement between self-report and medical records on number of diagnostic imaging procedures in the 10 years before diagnosis by total exposure, case-control study of AML, Los Angeles County, January 1987–June 1994. Total exposure was measured as the average of the number of self-reported procedures and the number of procedures reported in medical records. Agreement was measured as the number of procedures reported in medical records subtracted from the number of self-reported procedures. Each marker may represent more than one observation. A = controls ( $n = 315$ ), B = cases ( $n = 328$ ).

for each of the five subsets, an “opposing” subset of procedures was created: non-zero bone marrow radiation dose vs. zero bone marrow dose; fluoroscopic vs. non-fluoroscopic; high-dose (fluoroscopic imaging procedures; computerized tomography (CT) scans of the head, neck, and/or trunk; high-dose nuclear medicine scans that involve injection of radioactive isotopes) vs. low- or no-dose; regular x rays of spine or trunk (excluding chest) vs. all other regular x rays (including chest); and procedures done because of illness or injury vs. procedures done for routine reasons. Within procedure subsets,

**Table 4.** Means and 95% confidence intervals (CI) for agreement on number of diagnostic procedures<sup>a</sup> among cases by proxy status and exposure level; case-control study of AML, Los Angeles County, January 1987–June 1994.

Average number of procedures <sup>b</sup>	Direct respondents			Proxy respondents		
	<i>n</i>	mean	CI	<i>n</i>	mean	CI
0–10	118	0.7	–0.1, 1.5	115	0.4	–0.4, 1.3
11–20	39	–0.8	–3.4, 1.7	27	–4.0	–7.9, –0.1
>20	15	–2.5	–13.6, 8.7	14	–10.6	–20.4, –0.9

<sup>a</sup> Agreement = number of procedures reported in medical records subtracted from self-reported number of procedures; i.e., negative agreement represents underestimation by self-report.

<sup>b</sup> Average of number of procedures reported in interview and in medical records.

and after controlling for numbers of procedures within subsets, the only case-control difference in agreement was for non-zero bone marrow dose imaging procedures for which cases tended to underreport and controls tended to overreport exposure [LS mean (CI) = –0.6 (–1.0, –0.1) for cases, 0.3 (–0.1, 0.8) for controls;  $p = 0.004$ ].

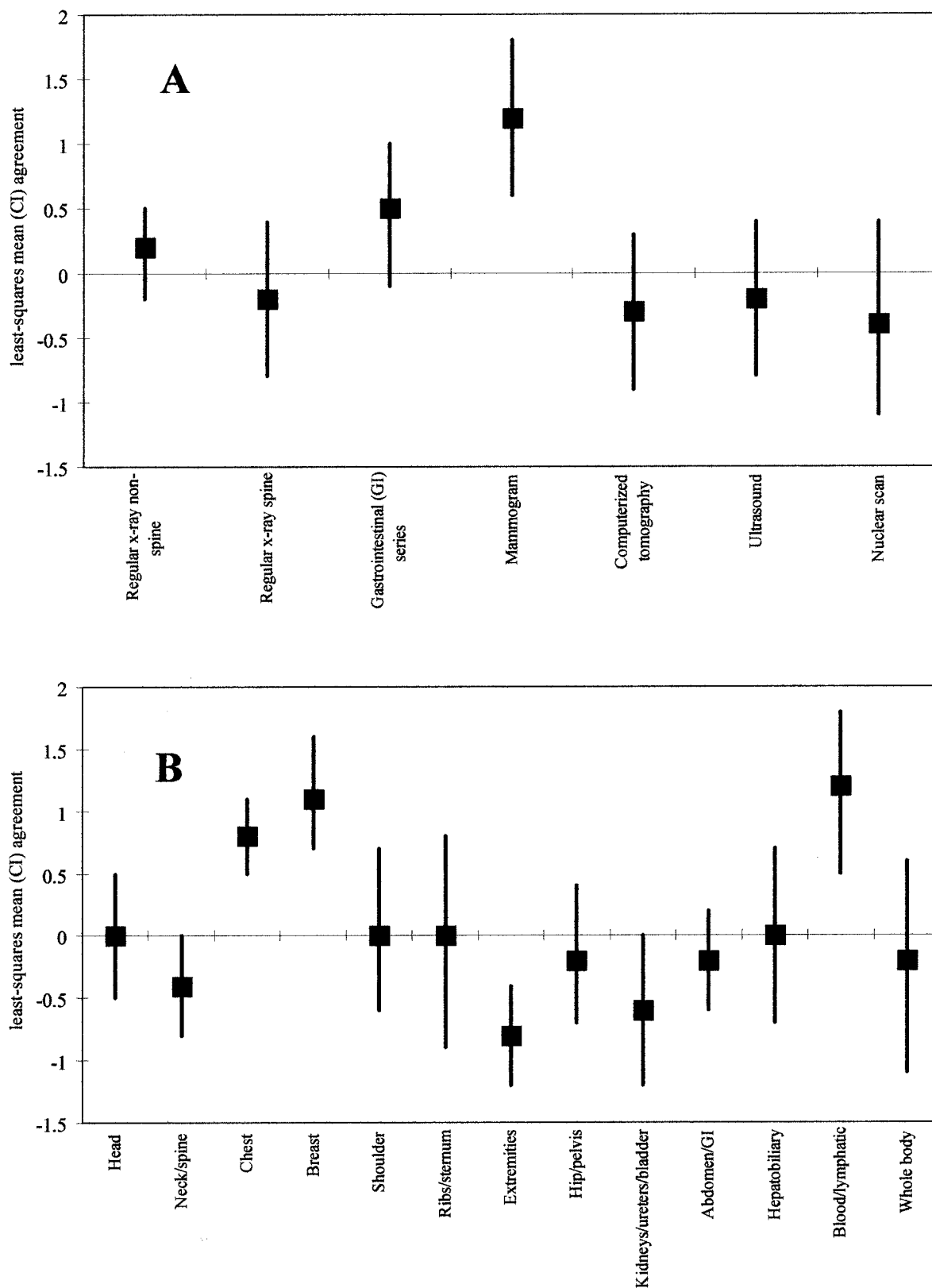
Because of substantial differences in variability of agreement between opposing pairs of procedure subsets, standard parametric analyses of agreement differences by subset were inappropriate and, in some instances, produced quite different results than rank-based analyses. Fig. 3 shows the midspread of agreement for each pair of procedures. Because midspreads could not be adjusted for within-subject correlation or number of procedures, we also included LS means and CIs for ranks of agreement scores (e.g., for non-zero bone marrow dose vs. zero bone marrow dose, agreement scores for each were ranked relative to each other within subjects). Agreement for regular spine/trunk x rays was similar to agreement for all other regular x rays ( $p = 0.66$ ), and agreement for non-zero dose procedures was similar to agreement for zero dose procedures ( $p = 0.89$ ). High-dose, fluoroscopic, and non-routine procedures tended to be underreported while low-dose, non-fluoroscopic, and routine procedures tended to be overreported ( $p < 0.01$  for each comparison).

Table 5 shows differences in agreement by disease status for varying latency (time since exposure) periods, adjusted for number of procedures within each latency period. Among controls, agreement was generally the same regardless of latency. Among cases, agreement was particularly poor in the 2 years before diagnosis, in which exposure was more underreported than in any other 2-y interval in the 10 years before diagnosis. During these 2 years before diagnosis, cases had more procedures than in any other 2-y interval in the 10 years before diagnosis [mean (CI) number of procedures = 3.9 (3.4, 4.4) for the interval [0–2] years before diagnosis, 2.7 (2.3, 3.1) for (2–4] y, 2.3 (2.0, 2.7) for (4–6] y, 2.5 (2.1, 3.0) for (6–8] y, 2.2 (1.7, 2.7) for (8–10] y;  $p < 0.0001$ ].

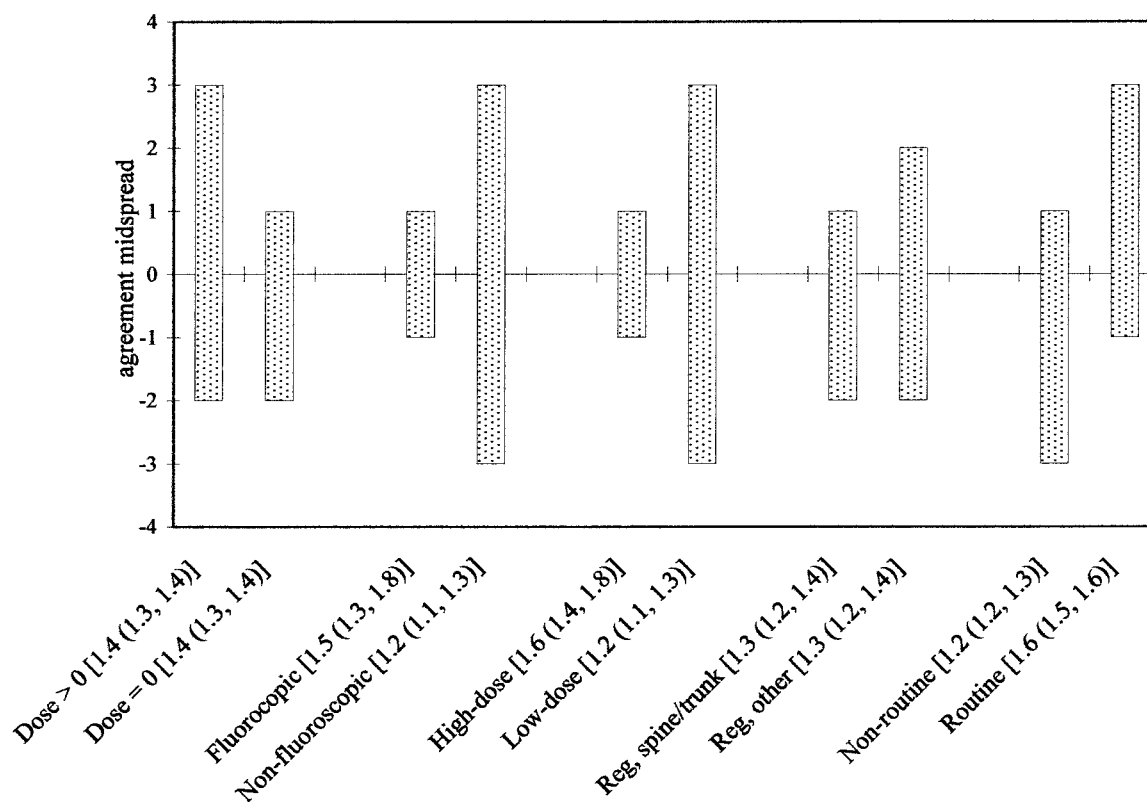
## DISCUSSION

In this study, use of medical records to determine exposure was not entirely free of influences from subject recall since health providers who were contacted to provide records were reported to us in subject interviews (supplemented by providers' input on additional providers who had treated their patients). Nonetheless, neither number of total providers nor proportion of providers from whom complete records were obtained differed by case-control status. A relatively high average percentage (71%) of subjects' medical records were successfully retrieved from their past health care providers. The majority of subjects had no more than 10 total procedures and overreported their exposure compared to medical records; i.e., fewer procedures were self-reported than were found in their medical records. In general, agreement between self-report and medical records did not depend on case-control status. These findings are generally consistent with past studies of agreement between self-report and medical records on prior diagnostic procedures and other health-related events (Paganini-Hill and Ross 1982; Preston-Martin et al. 1985; MacKenzie and Lippman 1989; Betz Brown and Adam 1992; McKenna et al. 1992; Whitman et al. 1993; Tsubono et al. 1994; Hiatt et al. 1995; Chouinard and Walter 1995; Basso et al. 1997; Zhu et al. 1999).

In this study, greater total exposure was associated with greater disagreement, particularly among cases and case proxy respondents. Unlike results from most other studies [as reviewed by Harlow and Linet (1989)] and contrary to prevailing theory on recall of frequently-occurring events (Sudman and Bradburn 1973; Bradburn et al. 1987), respondents were more likely to under-rather than overreport exposure as total exposure increased. Most subjects in this study with high total exposure (an average of at least 20 procedures) had between five and nine different kinds of procedures; i.e., imaging procedures in general were common occurrences for this group, but individual types of procedures were not



**Fig. 2.** Least-squares mean (95% CI) agreement between self-report and medical records on number of diagnostic imaging procedures in the 10 years before diagnosis by procedure type (A) and anatomical site (B), adjusted for "subject" and number of procedures within type/site, case-control study of AML, Los Angeles County, January 1987–June 1994; perfect agreement = 0, agreement <0 = respondent underestimation, agreement >0 = respondent overestimation.



**Fig. 3.** Midspreads (25th to 75th percentiles) of agreement between self-report and medical records on number of diagnostic imaging procedures in the 10 years before diagnosis by subsets of procedures of most interest to the AML-radiography hypothesis and subsets of their opposing procedures, case-control study of AML, Los Angeles County, January 1987–June 1994. Least-squares means (95% CIs) of ranks of agreement, adjusted for “subject” and number of procedures within subsets, are annotated next to each label along the x-axis.

**Table 5.** Least-squares (LS) means and 95% confidence intervals (CI) for agreement<sup>a</sup> between self-report and medical records on diagnostic imaging procedures in the 10 years before diagnosis, by case-control status and 2-y intervals since procedure,<sup>b</sup> adjusted for number of procedures within the 2-y interval; case-control study of AML, Los Angeles County, January 1987–June 1994.

Years since procedure <sup>c</sup>	Cases			Controls		
	<i>n</i>	LS mean	CI	<i>n</i>	LS mean	CI
[0–2] <sup>d</sup>	327	–1.6	–2.0, –1.2	313	–0.9	–1.4, –0.5
(2–4]	284	–0.8	–1.2, –0.4	274	–0.9	–1.3, –0.5
(4–6]	240	–1.0	–1.3, –0.6	220	–0.9	–1.3, –0.5
(6–8]	173	–0.7	–1.1, –0.3	171	–0.7	–1.1, –0.3
(8–10]	107	–1.1	–1.6, –0.7	103	–0.9	–1.4, –0.4

<sup>a</sup> Agreement = number of procedures reported in medical records subtracted from self-reported number of procedures; i.e., negative agreement represents underestimation by self-report.

<sup>b</sup> Diagnosis year used as the reference year.

<sup>c</sup> Brackets indicate that the endpoint is included in the interval, parentheses indicate that it is excluded.

<sup>d</sup>  $p < 0.05$ ; analysis of covariance with adjustment for number of procedures within the interval.

necessarily frequently occurring. Therefore, the over-reporting that is usually observed for frequently-occurring events may not apply in this setting since it was the general exposure of imaging procedures that occurred frequently rather than a specific exposure, such as a chest x ray. In fact, mammograms, regular chest x rays, and GI series were among the most overreported procedures in

our analysis and, in this population, were among the most frequently-occurring specific procedures.

Because the primary hypothesis of the case-control study from which these data originated was that medical radiography increases risk of AML, accuracy of self-reports of procedures associated with relatively high bone marrow doses of radiation were of more interest

than procedures delivering little or no dose. Among the highest-dose procedures are CT scans, fluoroscopic procedures, and certain high-dose nuclear medicine scans. It seems possible that these procedures would be more salient to respondents than lower-dose procedures, such as chest x rays. Some studies suggest that salient events, such as these types of procedures, are more susceptible to being reported as having occurred more recently and with more frequency than they were actually experienced (Sudman and Bradburn 1973; Bradburn et al. 1987). However, other studies have generally found that saliency is associated with greater accuracy (Harlow and Linet 1989; Zhu et al. 1999). In the study published by Graham et al. in the early 1960's, fluoroscopic procedures were reported more accurately than other types of procedures (Graham et al. 1963). Interestingly, our study found that procedures involving the blood or lymphatic systems were more likely to be overreported by cases for whom the blood or lymphatic systems might be more salient. But agreement for fluoroscopic as well as high-dose procedures in general was better than for non-fluoroscopic and low- or no-dose procedures, suggesting that saliency was generally associated with greater report accuracy. The only significant case-control difference in agreement was for all non-zero bone marrow dose procedures combined, which were underreported by case and overreported by control respondents. These findings suggest that self-reported exposure is more accurate for procedures of most interest to the AML-radiography hypothesis and that this exposure is reported with a similar level of accuracy by both cases and controls.

AML is a disease of rapid onset and most patients are symptomatic for less than six weeks before diagnosis (Goldberg 1975). Nonetheless, it is possible that medical radiography in the year or two before diagnosis may be related to preleukemic symptoms (e.g., joint pain). Therefore, for this unique exposure, it is especially prudent to exclude recent time periods in analyses of AML risk estimation. Analysis of agreement by latency period showed relatively poor agreement, specifically underreporting, among cases for this time interval. It is surmised that this is because cases underwent more procedures during these 2 years than they had in any previous 2-y interval in the 10 years before diagnosis, thus making this time period more susceptible to underreporting error, as noted above.

Some limitations of this study must be considered in interpreting these results. First, because AML is such a rapidly fatal disease, a relatively large proportion of eligible cases were unable to participate either directly or through proxies. It seems unlikely that availability of providers' records would depend on case vital status or degree of illness. However, it is possible that recall by

cases would be adversely affected by disease severity. The direction of potential reporting error (under- vs. overreporting) by severely ill respondents is not obvious, but consideration must be given to the fact that the cases who participated in this study may not have been representative of AML cases in general. Second, again because of the rapid case fatality and morbidity, half of the case respondents were proxies. Proxies tended to underreport exposure, especially as exposure increased. Direct respondents also underreported high-end exposure but to a lesser degree than proxies. Thus, fewer proxies would likely result in more accurate, albeit underreported, self-reported exposure. Third, about 30% of provider records per subject, on average, were unobtainable. Complete provider records would almost certainly have increased the numbers of procedures in medical records; thus, underreporting may be greater with complete medical record data than that observed in this study. Since there appear to be no case-control differences in proportion of or reasons for unavailable provider records, this suggests a further attenuation of risk estimates from studies with self-reported data.

## CONCLUSION

This study of agreement between self-report and medical records on diagnostic imaging has important implications for the study of radiography as a risk factor for AML. Spurious associations with increased risk seem unlikely, whether self-reported exposure data are used or exposure data are abstracted from medical records. Such associations can occur when cases overreport more than controls or when controls underreport more than cases. There was no evidence of such reporting error in this study. If anything, cases tended to underreport compared to controls. Further, there was no evidence of more complete record ascertainment from cases' health care providers, and it seems improbable that cases' providers would have used more diligence in documenting diagnostic procedures over a 10-y period because time from first symptoms to AML diagnosis is usually very short. Because "unexposed" subjects were defined as those with no self-reported procedures and no procedures in medical records (since "total exposure" was defined as the sum of self-reported and medical record procedures), it could not be determined whether there was differential underreporting among these subjects. Therefore, a conclusion cannot be drawn as to how this type of reporting error might affect risk estimates for "exposed" categories relative to the reference group of "unexposed" subjects.

Medical record data from roughly one-third of all providers named in interview were unobtainable, and discrepancies between self-report and medical records

tended towards underreporting for frequently-exposed subjects despite the use of highly specific questions that included interviewer prompts and explicit descriptions of the procedures being queried. A less thorough data collection instrument, such as a self-administered questionnaire, is likely to result in even greater underreporting. Adding to the underreporting was the use of proxies for nearly half of all cases, a necessity for a rapidly fatal disease such as AML. The cost-efficient approach of a highly structured interview appears acceptable for ascertaining radiography exposure for case-control studies. The greatest limitation of a study of radiography as a risk factor may be nondifferential underreporting of true exposure, regardless of data source, leading to attenuated risk estimates.

*Acknowledgments*—This work was supported by NIH grant CA17054 and by the National Institute of Environmental Health Sciences grant 5 P30 ES07048-06. Cases were identified by the Cancer Surveillance Program with support from SEER contract NO1-CN-25403 of the Division of Cancer Prevention and Control, National Cancer Institute, National Institutes of Health, Department of Health and Human Services and contracts 050 (C-J)-8709 from the State of California Department of Health Services. We thank Wendy J. Mack for her review of the manuscript and Maria Paul, Kristina Paoff, and Jennifer Martin Borja for their assistance in data collection and management.

## REFERENCES

- Basso O, Olsen J, Bisanti L, Karmaus W. The performance of several indicators in detecting recall bias. *Epidemiology* 8:269–274; 1997.
- Betz Brown J, Adam ME. Patients as reliable reporters of medical care processes. *Med Care* 30:400–411; 1992.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 1:307–310; 1986.
- Bradburn NM, Rips LJ, Shevell SK. Answering autobiographical questions: The impact of memory and inference on surveys. *Science* 236:157–161; 1987.
- Chouinard E, Walter S. Recall bias in case-control studies: an empirical analysis and theoretical framework. *J Clin Epidemiol* 48:245–254; 1995.
- Conover WJ, Iman RL. Rank transformations as a bridge between parametric and nonparametric statistics. *Am Statistician* 35:124–133; 1981.
- Conover WJ, Iman RL. Analysis of covariance using the rank transformation. *Biometrics* 38:715–724; 1982.
- Fletcher DM. Practice brief. Retention of health information (updated). *American Health Information Management Association. J AHIMA* 70(Suppl 27):64; 1999.
- Goldberg JD. The effects of misclassification on the bias in the difference between two proportions and the relative odds in the fourfold table. *J Am Stat Assn* 70:561–567; 1975.
- Graham S, Levin ML, Lilienfeld AM, Dowd JE, Schuman LM, Gibson R, Hempelmann LH, Gerhardt P. Methodological problems and design of the tristate leukemia survey. *Ann NY Acad Sci* 107:557–569; 1963.
- Harlow SD, Linet MS. Agreement between questionnaire data and medical records. *Am J Epidemiol* 129:233–248; 1989.
- Hiatt RA, Perez-Stable EJ, Quesenberry C Jr., Sabogal F, Otero-Sabogal R, McPhee SJ. Agreement between self-reported early cancer detection practices and medical audits among Hispanic and non-Hispanic white health plan members in northern California. *Prev Med* 24:278–285; 1995.
- MacKenzie SG, Lippman A. An investigation of report bias in a case-control study of pregnancy outcome. *Am J Epidemiol* 129:65–75; 1989.
- McKenna MT, Speers M, Mallin K, Warnecke R. Agreement between patient self-reports and medical records for pap smear histories. *Am J Prev Med* 8:287–291; 1992.
- Paganini-Hill A, Ross RK. Reliability of recall of drug usage and other health-related information. *Am J Epidemiol* 116:114–122; 1982.
- Preston-Martin S, Pogoda JM. Estimation of radiographic doses in a case-control study of acute myelogenous leukemia (AML). *Health Phys* 2003 (in press).
- Preston-Martin S, Paganini-Hill A, Henderson BE, Pike MC, Wood C. Case-control study of intracranial meningiomas in women in Los Angeles County. *J Natl Cancer Inst* 65:67–73; 1980.
- Preston-Martin S, Bernstein L, Andrews Maldonado A, Henderson BE, White SC. A dental x-ray validation study. Comparison of information from patient interviews and dental charts. *Am J Epidemiol* 121:430–439; 1985.
- SAS/STAT User's Guide, Release 6.03 Edition. Cary, NC: SAS Institute; 1988: 943.
- Spitz MR, Fueger JJ, Newell GR. The development of a comprehensive, institution-based patient risk evaluation program: II. Validity and reliability of questionnaire data. *Am J Prev Med* 4:188–193; 1988.
- Sudman S, Bradburn NM. Effects of time and memory factors on response in surveys. *J Am Stat Assoc* 68:805–815; 1973.
- Tsubono Y, Fukao A, Hisamichi S, Hosokawa T, Sugawara N. Accuracy of self-report for stomach cancer screening. *J Clin Epidemiol* 47:977–981; 1994.
- Whitman S, Lacey L, Ansell D, Chen EH, Dell J, Phillips CW. Do chart reviews and interviews provide the same information about breast and cervical cancer screening? *Int J Epidemiol* 22:393–397; 1993.
- Zhu K, McKnight B, Stergachis A, Daling JR, Levine RS. Comparison of self-report data and medical records data: results from a case-control study on prostate cancer. *Int J Epidemiol* 28:409–417; 1999.

