HOSPITAL ADMISSION RECORDS: A SOURCE FOR IDENTIFYING OCCUPATIONAL GROUPS AT RISK OF CANCER *

Lorne Houten, Irwin D. J. Bross, and Enrico Viadana

Roswell Park Memorial Institute
Buffalo, New York 14263

In this paper we will describe the methods used in a retrospective study of cancer in relation to occupation done by the Department of Biostatistics at Roswell Park Memorial Institute under the sponsorship of the National Institute of Occupational Safety and Health. The goal of the study is to elucidate possible etiological factors in the work environment by studying the occupation and health history of patients at Roswell Park Memorial Institute.

The data were obtained from the Epidemiology schedules collected at Roswell Park Memorial Institute during the years 1956–1971, covering approximately 26,000 patients. The following variables were selected from the epidemiology schedules for study: diagnosis, age, sex, occupation, and length of exposure to the occupation. The diagnosis is subdivided by type and site of cancer or nonneoplastic disease. The occupation is classified by industry and type of work done, e.g. operative, labor, or clerical. We divided age at diagnosis into under 60 and over 60 and length of exposure by less or more than five years in a given occupation.

All cancer sites that constituted more than one percent of the study were included. Certain adjacent sites were considered together; for example, all buccal cavity sites were combined, as were colon and rectum. All leukemias were considered together, since sample sizes were too small when they were divided. Histological typing was in general not available, except that melanomas were considered separately from other skin cancers. Patients with benign neoplasms were excluded from the study in order to make a clear distinction between malignant neoplastic disease and nonneoplastic disease.

A list of potentially hazardous occupations was provided by the National Institute for Occupational Safety and Health and served as a guideline for the occupations considered. Housewives, service workers, salesworkers, and management personnel were excluded from the study, because it was assumed that workers in such occupations would be less likely to be exposed to occupational hazards.

The majority of the patients admitted to Roswell Park Memorial Institute are from the Western New York area. They were principally urban, although not as much so as would be found in the New York City area. Because of considerable previous studies done on this question, ethnic background was not made one of the variables considered, although it may be noted that blacks are underreported in the sample. This appears to reflect the fact that black cancer patients are not referred to Roswell Park by their primary physicians as frequently as whites.

* This investigation was supported by the National Institute for Occupational Safety and Health, Contract no. 41043, and Public Health Service Research grant no. CA-11531 from the National Cancer Institute.

A comparison of the occupational distribution of the Roswell Park admissions with figures published for Erie County as a whole indicate that the sample reflects the overall occupational picture of the county fairly well. Caution must be used in drawing conclusions from this, however, as biases may be introduced by varying referral practices. For example, company physicians from one factory may encourage their employees to be seen at Roswell Park, whereas another factory may discourage them.

Educational and socioeconomic levels were not considered in the initial study. For certain sites, e.g. cervix, these factors probably should be considered, and the methods used in this study to do age adjustments could also be used to do adjustments for education and socioeconomic status. Adjustment for smoking history are being done for selected sites, e.g. lung, buccal area, and bladder.

For each occupation, site, age category, and exposure time a relative risk was computed. The relative risk is a quantity used in epidemiological research that measures the chance that an individual exposed to an environmental hazard has a disease relative to the chance that an individual not exposed to the hazard has the disease.

The relative risk calculated from the data in a 2×2 contingency table of the form:

	Exposed	Not Exposed
Case	n_{11}	n_{12}
Control	n_{21}	n_{22}

is estimated by the quantity $n_{11}n_{22}/n_{12}n_{21}$.

For the purposes of this study the clerical workers were chosen as the "not exposed" population. The disease incidence of the clerical workers was compared to that of the whole hospital population and agreed quite well in almost all cases, the principal exception being a moderate difficiency in breast cancer in the clerical workers. Patients with nonneoplastic disease were chosen as the "control" population. The interviews were, of course, carried out prior to diagnosis.

For each occupation, site, and exposure time, a 2×2 contingency table was constructed for age ≤ 60 and for age 60+, namely

	Exposed	Clerical
age ≤ 60		
Cancer cases	$(n_{11})_1$	$(n_{12})_1$
Nonneoplastic cases	$(n_{21})_1$	$(n_{22})_1$

and

age 60+		
Cancer cases	$(n_{11})_2$	$(n_{12})_2$
Nonneoplastic cases	$(n_{21})_2$	$(n_{22})_2$

Age-adjusted relative risks were then calculated by the Bertell method, which we will briefly describe. We substitute for the number of controls with exposure; i.e., the quantities $(n_{21})_i$ $i = 1,2$, and the number of controls without exposure, i.e., the quantities $(n_{22})_i$ $i = 1,2$, the quantities $(e_{21})_i$ and $(e_{22})_i$, $i = 1,2$ which are computed as

$$(e_{21})_i = \frac{(n_{21})_i + ((n_{11})_i + (n_{12})_i)}{(n_{21})_i + (n_{22})_i}$$

$$(e_{22})_i = \frac{(n_{22})_i + ((n_{11})_i + (n_{12})_i)}{(n_{21})_i + (n_{22})_i},$$

which may be interpreted as the corresponding expected number of controls if the size of the control population were equal to the size of the case population.

TABLE 1

RELATIVE RISKS AND SAMPLE SIZES FOR WOMEN EVER EMPLOYED IN THE SPECIFIED OCCUPATION BY AGE AT DIAGNOSIS: OPERATIVES IN APPAREL INDUSTRY *

Cancer Diagnosis	Age	RR	A	B	C	D	Cancer Diagnosis	Age	RR	A	B
Buccal	<60	4.22	5	45	15	570	Esopha-				
Cavity &	60+	3.49	2	14	5	122	gus	<60	†	0	3
Pharynx	all	3.98						60+	†	0	1
								all	†		
Stomach	<60	6.91	2	11			Colon	<60	0.53	1	72
	60+	†	0	10			Rectum	60+	0.75	2	65
	all	2.92						all	0.66		
Pancreas	<60	9.50	1	4			Nose	<60	†	0	4
	60+	†	0	3				60+	12.20	1	2
	all	4.50						all	5.12		
Larynx	<60	30.40	4	5			Lung	<60	†	0	18
	60+	†	0	3				60+	†	0	7
	all	16.70						all	†		
Breast	<60	1.28	13	386			Cervix	<60	3.82	25	249
	60+	1.49	11	180			Uteri	60+	6.33	14	54
	all	1.37						all	4.41		
Corpus	<60	5.07	4	30			Ovary	<60	1.61	3	71
Uteri	60+	1.38	4	71				60+	†	0	17
	all	2.18						all	1.18		
Kidney	<60	†	0	14			Bladder	<60	†	0	3
	60+	†	0	6				60+	2.71	2	18
	all	†						all	2.44		
Melanoma	<60	3.17	1	12			Skin	<60	2.62	2	29
	60+	†	0	4			(other)	60+	†	0	30
	all	2.10						all	1.01		
Brain	<60	†	0	3			Lympho-				
	60+	†	0	0			mas	<60	3.04	6	75
	all	†						60+	0.94	1	26
								all	2.32		
Myeloma	<60	†	0	5			Leuke-				
	60+	†	0	3			mia	<60	†	0	6
	all	†						60+	1.52	1	16
								all	1.22		

* RR=relative risk; A=number of cases in specified occupation; B=number of cases in clerical occupations; C=number of noncancer controls in specified occupation; D=number of noncancer controls in clerical occupations.

† Quantity not defined.

TABLE 2

NUMBER OF CASES,* RELATIVE RISKS AND ASSOCIATED PROBABILITIES FOR WOMEN EVER EMPLOYED IN THE SPECIFIED OCCUPATION (ALL AGES)

Cancer Diagnosis	No. of Cases	Relative Risk	Probability
	Operatives in Apparel Industry		
Buccal cavity and pharynx	7	3.977	<.0001
Breast	24	1.367	<.1612
Corpus uteri	8	2.178	<.0010
Cervix uteri	39	4.408	<.0001
Lymphoma	7	2.315	<.0001
	Operatives in Chemical Industry		
Breast	7	0.987	<.7288
Cervix uteri	6	1.971	<.0436

* A site is included in TABLE 2 if there are at least five cases reported at that site.

The expected number can be summed over the two age categories to give a 2×2 contingency table:

	Exposed	Not Exposed
Case	$(n_{11})_1 + (n_{11})_2$	$(n_{12})_1 + (n_{12})_2$
Constructed Controls	$(e_{21})_1 + (e_{21})_2$	$(e_{22})_1 + (e_{22})_2$

The null hypothesis is that the relative risk is one. We then calculate the probability, under the null hypothesis, that the actual observation or one with a more extreme relative risk will occur. If the sample size is sufficiently large, the chi-square probability is used; otherwise, a Fisher exact probability of having this set of tables or a set with the same marginal values which would generate an even higher relative risk is calculated. The advantage of this method is that the chi-square analysis need be justified only for the summary table and not for each component subtable, as would be necessary using the Woolf-Haldane method.

TABLES 1 and 2 are representative samples of the tables obtained in the study of the cases from 1956–1965. The full document will be published as a monograph by the National Institute of Occupational Safety and Health.