Scand J Work Environ Health 1994;20(5):371-375

<https://doi.org/10.5271/sjweh.1384>

Issue date: 01 Oct 1994

Comparison of information sources for case-referent data.

by [Persson B](#), [Karlsson M](#), [Axelson O](#)

Affiliation: Department of Occupational and Environmental Medicine,
University Hospital, Linköping, Sweden.

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/7863301

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Comparison of information sources for case-referent data

by Bodil Persson, MD, Magnus Karlsson, BSc, Olav Axelsson, MD¹

PERSSON B, KARLSSON M, AXELSSON O. Comparison of information sources for case-referent data. *Scand J Work Environ Health* 1994;20:371-5.

OBJECTIVES — Exposure assessment and the ascertainment of cases are topics of main concern in epidemiologic research. To investigate validity aspects of this kind practically, a comparative evaluation was performed of information sources for case-referent data.

METHODS — Cases of Hodgkin's disease and non-Hodgkin's lymphoma diagnosed in 1975-1984 were collected from the Regional Cancer Register in Linköping. The study population was restricted to men 20-80 years of age at diagnosis. Living cases included in a previous case-referent study and the complementary cases from the register were compared with regard to the exposures indicated by job title and industrial branch, as given in the National Population Census in 1975. Furthermore, questionnaire-based exposure information from the case-referent study was compared with exposures derived from census information.

RESULTS — Only 21% of all of the cases in the study base had been included in the questionnaire-based case-referent study. No major differences in the distribution of exposure were found for the two case series, except that fewer subjects were economically inactive in the previous case-referent study. No clear relations appeared between survival and particular exposures. Risk estimates derived from the present cancer register study were compared with corresponding estimates from the previous case-referent study. With restriction to only economically active cases, the risk estimates approached those of the case-referent study.

CONCLUSIONS — This comparison showed that incomplete ascertainment of cases in a case-referent study was not an important source of bias, at least not in relation to what could be achieved by more complete case ascertainment through register linkage.

KEY TERMS — exposure assessment, census, Hodgkin's disease, non-Hodgkin's lymphoma, recall bias, register linkage, restriction, validity.

Proper exposure assessment is required for valid epidemiologic research. Exposure information of various quality can be obtained from registers, questionnaires, interviews, or measurements. Recall bias can affect exposure information from questionnaires or interviews, especially in case-referent studies. Register data are more objective in character but are usually only poor surrogates for specific exposures, and they do not usually allow for taking confounding and interaction into account in the analyses. Another aspect relates to the validity of using job titles for indirectly assessing more specific exposures (ie, in principle to applying job-exposure matrices) (1).

A further and common concern in epidemiologic research relates to the problem of harvesting the cases in the base population. In a cohort study it would be deleterious not to include all cases, especially when the expected numbers are calculated from national or regional rates. The incomplete ascertainment of cases is less problematic in case-referent studies, given that there is no selective loss of cases

in relation to the risk determinants under study (2). Although this latter aspect seems theoretically clear and simple, there is nevertheless a need to obtain practical experience on how incomplete case ascertainment might influence the findings of a study. Such an evaluation is difficult to refine from the issue of exposure assessment, and also recall bias phenomena may interfere because, when all cases in the base population are enrolled as drawn from a complete register, so-called objective information (eg, on job titles) tends to become inadequate for many of them.

To investigate validity aspects of this kind, a comparison was made between earlier questionnaire-based case-referent data on living subjects (3) and cancer register information covering all cases from the same study base. This comparison involved the linkage of data from the Regional Cancer Registry to occupational information provided by the National Population and Housing Census in Sweden.

Material and methods

All cases of malignant lymphoma [codes 200.0, 200.1, 200.2 and 201 of the International Classification of Diseases (ICD), 8th revision] diagnosed in 1975-1984 among men were obtained through the Regional Cancer Registry in Linköping, a subdivi-

¹ Department of Occupational and Environmental Medicine, University Hospital, Linköping, Sweden.

Reprint requests to: Dr B Persson, Department of Occupational and Environmental Medicine, University Hospital, S-581 85 Linköping, Sweden.

sion of the National Cancer Registry. Codes 200.0, 200.1 and 200.2 were taken as non-Hodgkin's lymphoma and 201 as Hodgkin's disease. The study base came from three counties in the southeast of Sweden during the period 1975–1984, with a total population of about 0.9 million. Only men 20–80 years old at the time of diagnosis were included in the study. All told, 590 cases were registered in this age group during the study period.

The earlier case-referent study that was used for comparison was based on the same register but involved only living subjects (3), but also some other inclusion criteria had been applied, such as maximum age at the time of data acquisition, mental capacity, and a histopathological confirmation of the diagnoses. Further details have been presented elsewhere (3).

The unique personal identification numbers of Swedish citizens made it possible to link the individual cancer cases to the National Population and Housing Census. The information on job title and branch of industry used in this linking refers to conditions during 20–26 October 1975, when the census was performed. For reasons of integrity we obtained only depersonalized information from the linkage; therefore it was not possible to compare census and self-reported information directly at an individual level. Instead we had to focus more on overall (group) information and a comparison of results. When exposure frequencies were compared, the cases included in the previous study were subtracted from the total of cases obtained from the register, and the comparisons were made between the two subsets.

Using job titles and branch of industry in the census, an experienced occupational health physician devised a specific exposure classification for all cases and created different exposure categories. The classification particularly took into account the risk determinants reported in various earlier epidemiologic studies on malignant lymphomas. For example, exposure to solvents was assigned to painters, printers, gasoline station attendants, and chemical engineers, and the exposure category of electrical work included electricians, linemen, electrical engineers, and technicians. Another exposure of particular interest, such

as to fresh wood, was assigned to sawmill workers, lumberjacks, and paper pulp workers. Some specific job titles were also included, such as farming and gardening, nursing, and the like. No information on potential confounding factors or former occupational exposures was available from the census data.

The complete case material was described in terms of vital status and diagnostic subentities (table 1). The distribution of various exposures was considered, and, for the calculation of risk estimates, population referents from the previous case-referent study were used (3). For the comparison of exposure frequencies (table 2), two-tailed P-values were calculated with the use of the chi-square test or Fisher's exact test for small numbers. Odds ratios (table 3) were obtained by means of the EPI-INFO package, which also provided for significance testing (4).

Results

Of the 590 cases of malignant lymphomas, there were 76 under the ICD code 200.0, whereas 298 belonged to ICD 200.1, 86 to 200.2, and 130 to 201. Most of the men (69%) were dead in 1988, when the case subjects for the case-referent study were recruited (3). The subset of living cases included in that study constituted only 21% of the registered cases. The proportion of dead subjects varied from 65 to 83% between the various ICD codes and so did consequently also the proportion of cases included in the case-referent study (table 1).

Seventeen of the persons listed with lymphoma in the cancer register were missing in the census, and for one additional subject exposure information was lacking (table 2). Altogether, 228 subjects were reported as economically inactive, although, because of their age, only 199 were expected to be retired. The difference was likely to depend partly on early retirement pensioning.

Analyses were performed to evaluate whether there were any pertinent associations between exposures and any specific ICD codes. No clear such relations were found, however (data not shown). Lethal outcome was not related to any specified exposure in this study (data not shown).

Table 1. Cases among men in the Regional Cancer Register (total material) by diagnostic subentities. The subsets of deceased cases in 1988 and the living cases included in a case-referent study are also shown. (The living cases not fulfilling inclusion criteria for a case-referent study are not shown in the table.)

Diagnostic entity ^a	Total material (N = 590)		Deceased (N = 406)	Case-referent study (N = 124 ^b)	
	N ₁	%	N ₂	N ₃	% of N ₁
Histiocytic non-Hodgkin's lymphoma (200.0)	76	13	63	7	9
Lymphocytic non-Hodgkin's lymphoma (200.1)	298	50	199	77	26
Malignant non-Hodgkin's lymphoma, not otherwise specified (200.2)	86	15	60	9	10
Hodgkin's disease (201)	130	22	84	31	24

^a Code of the International Classification of Diseases in parentheses.

^b 21% of the total material.

The occurrence of various exposures was compared between the cases included in the case-referent study and the other cases obtained from the Regional Cancer Register (table 2). In this comparison no significant differences in exposure frequencies were found for such determinants as wood (carpenters and cabinet makers, and fresh wood), farming and gardening, or solvents. However, for non-Hodgkin's lymphoma the exposure category of welders, metal workers, and mechanics appeared significantly more often in the case-referent material. Significant differences were also noted for Hodgkin's disease and students and for non-Hodgkin's lymphoma

and those engaged in teaching. For both Hodgkin's disease and non-Hodgkin's lymphoma significantly fewer subjects were found to be economically inactive in the case-referent material (3).

Finally, odds ratios for some exposures were calculated from the present cancer register material for comparison with the risk estimates obtained in the previous case-referent study. These risk estimates were generally decreased in comparison with the crude estimates in the case-referent study (table 3). When the present material was restricted to include only cases with specified occupational information in the census (ie, when economically inactive sub-

Table 2. Exposure frequencies (as percentage) for cases of Hodgkin's disease and non-Hodgkin's lymphoma among men from a previous case-referent study compared to complementary cases within the same study population, as derived from the Regional Cancer Registry. The exposure classification was based on occupational information from a census. (The frequencies have been expressed with two figures and rounded off to 0.0 or 0.5.)

Exposure	Hodgkin's disease		non-Hodgkin's lymphoma	
	Case-referent study (N = 31)	Complementary cases (N = 99)	Case-referent study (N = 93)	Complementary cases (N = 367)
	%	%	%	%
Carpenters and cabinetmakers	6.5	2.0	7.5	4.0
Fresh wood ^a	9.5	7.0	3.0	3.0
Farming and gardening	6.5	4.0	11	6.5
Solvents	3.0	0	3.0	2.0
Welders, metal workers and mechanics	9.5	9.0	10	4.0*
Electrical work	6.5	2.0	3.0	2.0
Nursing	6.5	1.0	2.0	0.5
Other manufacturing	6.5	15	7.5	8.0
Service occupations	13	7.0	10	6.5
Students	13	1.0*	2.0	0
White-collar workers	13	10	12	9.0
Teaching	0	2.0	4.0	1.0*
Other occupations	0	2.0	2.0	3.0
Economically inactive	6.5	34*	23	47*
No exposure information	0	0	0	0
Subjects not found in census	0	4.0	0	3.5
Total	100	100	100	100

^a Sawmill workers, lumberjacks or paper pulp workers.

* $P < 0.05$, two-tailed, by chi-square or Fisher's exact test.

Table 3. Crude odds ratios (OR) for some exposures with regard to Hodgkin's disease (HD) and non-Hodgkin's lymphoma (NHL). The exposure classification for cases (HD = 130, NHL = 460) was based on census information. Population referents (N = 204) were recruited from an earlier case-referent study along with the exposure data. Excluding cases without exposure information in the census reduced the number of cases (HD = 91, NHL = 253). Risk estimates from the case-referent study based on questionnaire data are given for comparison.

	Present register study				Case-referent study	
	Exposed		Crude OR		Crude OR	Logistic OR
	Cases (N)	Referents (N)	All cases	After exclusions		
Hodgkin's disease						
Carpenters and cabinet-makers	4	25	0.2	0.3	0.2	0.2
Fresh wood ^a	10	10	1.6	2.4	2.9	3.8
Farming and gardening	6	43	0.2	0.3	0.7	0.3
Non-Hodgkin's lymphoma						
Carpenters and cabinet-makers	21	25	0.3	0.7	1.0	0.9
Fresh wood ^a	14	10	0.6	1.1	2.6	· ·
Farming and gardening	34	43	0.3	0.6	1.5	0.6

^a Sawmill workers, lumberjacks or paper pulp workers.

jects along with subjects with no exposure information or those not found in the census were excluded) the risk estimates approached those found in the earlier case-referent study. However, only the crude estimates were comparable, since the logistic estimates were adjusted for other risk factors not known for the complementary cases.

Discussion

The epidemiologic approach to study a problem is often, if not always, a result of various compromises and limitations at the operational or empirical level (5). These circumstances should be kept in mind when the external validity of a study is considered. Our comparative evaluation provides some additional knowledge about how different approaches may influence the risk estimates obtained in epidemiologic studies.

The possibility to perform register-based epidemiology is limited by the quality and completeness of registers. In spite of the fact that cancer registration in Sweden is based on compulsory notifications from both clinicians and pathologists, some cases are not reported to the cancer registry. The completeness of the National Swedish Cancer Register has been estimated to be about 96% in general, and this figure is applicable also to the Regional Cancer Register in Linköping (6). The registration deficit of fatal cases of malignant lymphomas has been estimated to be 3.7% for the National Swedish Cancer Register (6). The register-based exposure information used for this study referred to a particular week in 1975 and was obviously limited. Inadequate exposure assessment should be expected to decrease the possibility to detect existing associations (7, 8); this possibility was also clearly confirmed. In ecologic record-linkage studies, however, nondifferential misclassification of exposure may result in bias that is directed away from the no-effect hypothesis (9).

The living subjects in the case-referent study represented only a minor part of the cases in the study base, as most of the cases were deceased at the time of case ascertainment. Prospective collection of incident cases for assessing exposure by interviews or questionnaires would certainly lead to inclusion of a greater part of the cases occurring in the study base. Such a design would require a long period for collecting information, however. Information from the next-of-kin of dead cases about employment in a particular company is certainly useful. Such information has been found to be valid but would nevertheless be less detailed than self-reported specific exposure data, let alone the risk of recall bias (10, 11). Similarly, factors implying confounding and interaction would require such detailed, self-reported information even more, although smoking is reported correctly by next-of-kin (10).

In a comparison of the results from the two studies, decreased risk estimates were found throughout

the census material. Hazard evaluations based on register linkage should be expected to provide results which are inherently conservative (12). In the present example the relatively lower risk estimates in the register-based study might have been due to nondifferential misclassification and dilution because of no information. As exposure information from a census concerns the conditions during a particular week, it would be reasonable to believe that the exposure frequency to various jobs and agents is lower in relation to the questionnaire-based information that covers a considerable period of time. On the other hand, especially the skilled work force within certain occupations is rather stable, as are farmers. This argument is also usually taken as valid to support register linkage studies.

When the economically inactive cases were excluded along with those lacking exposure information or not found of the census, the risk estimates were more in agreement with the findings in the case-referent study. The remaining subgroup probably conforms more with those answering questionnaires used for assessing exposure. Any more-specific exposure classification was only possible, however, when the job titles and branch of industry pointed to a more specific and uniform exposure.

The present comparative evaluation might also serve to validate the exposure information reported in the earlier case-referent study. Since the same set of referents was used in the present study as in the previous one (3), the comparisons made reflect only aspects relating to the cases. As no major differences finally appeared regarding the risk estimates of the two studies, there was no indication of recall bias.

Register linkage might be considered as too crude a method for etiologic research (13), the main problem being the low sensitivity due to dilution in exposure assessment through the nondifferential misclassification of exposure (8). A more surprising limitation of register linkage studies has been illustrated in a study on glassworkers and brain cancer (14). Hence, two register-linkage studies had shown an increased occurrence of brain cancer among glassworkers (15, 16), but a more specific study (14) revealed a regional cluster of brain cancer in the glass producing area, simulating an occupational risk. This observation could be explained by the fact that glassworkers were living in a restricted geographic area where either an environmental exposure or some other factor might have caused this adverse health effect in the population at large.

This comparison of study results has shown that incomplete ascertainment of cases in a case-referent study is hardly any major drawback, at least not in relation to what was achieved by register linkage, even though the latter approach might lead to more complete case ascertainment. This conclusion is in agreement with what should be expected also on theoretical grounds, and it is therefore likely to be generally valid.

Acknowledgments

This study was supported by grants from the Swedish Cancer Society.

References

1. Blair A, Linos A, Stewart PA, Burmeister LF, Gibson R, Everett G, et al. Evaluations of risks for non-Hodgkin's lymphoma by occupation and industry exposures from a case-control study. *Am J Ind Med* 1993;23:301—12.
2. Rothman KJ. *Modern epidemiology*. Boston/Toronto: Little, Brown and company, 1986.
3. Persson B, Fredriksson M, Olsen K, Boeryd B, Axelson O. Some occupational exposures as risk factors for malignant lymphomas. *Cancer* 1993;72:1773—8.
4. EPI-INFO [computer program]. Atlanta (GA): Centers for Disease Control, 1990.
5. Miettinen O. *Theoretical epidemiology*. New York, NY: John Wiley & Sons Inc, 1985.
6. Mattson B. Cancer registration in Sweden: studies on completeness and validity of incidence and mortality registers [thesis]. Stockholm: Karolinska Institute, 1984.
7. Checkoway H, Pearce NE, Crawford-Brown DJ. *Research methods in occupational epidemiology*; vol 13. New York, NY: Oxford University Press, 1989.
8. Ahlbom A, Axelson O, Støttrup Hansen E, Hogstedt C, Jensen UF, Olsen J. Interpretation of "negative" studies in occupational epidemiology. *Scand J Work Environ Health* 1990;16:153—7.
9. Brenner H, Savitz DA, Jöckel K-H, Greenland S. Effects of nondifferential exposure misclassification in ecologic studies. *Am J Epidemiol* 1992;135:85—95.
10. Pershagen G, Axelson O. A validation of questionnaire information on occupational exposure and smoking. *Scand J Work Environ Health* 1982;8:24—8.
11. Pershagen G. Lung cancer mortality among men living near an arsenic-emitting smelter. *Am J Epidemiol* 1985;122:684—94.
12. Axelson O. Linking of register information on morbidity and occupation, especially with regard to quality aspects. Copenhagen: World Health Organization, 1986:27—40. WHO ICP/OCH 002.
13. Hernberg S. *Introduction to occupational epidemiology*. Chelsea, MI: Lewis Publishers Inc, 1992.
14. Wingren G, Axelson O. Cluster of brain cancers spuriously suggesting occupational risk among glassworkers. *Scand J Work Environ Health* 1992;18:85—9.
15. McLaughlin JK, Malker HSR, Blot WJ, Malker BK, Stone BJ, Weiner JA, et al. Occupational risks for intracranial gliomas in Sweden. *JNCI* 1987;78:253—7.
16. McLaughlin JK, Thomas TL, Stone BJ, Blot WJ, Malker HSR, Weiner JA, et al. Occupational risks for meningiomas of the CNS in Sweden. *J Occup Med* 1987;29:66—7.

Received for publication: 31 December 1993