



Editorial

Scand J Work Environ Health [2005;31\(2\):85-97](#)

doi:10.5271/sjweh.854

Use of routinely collected occupational exposure data in register-based studies--a trade off between feasibility and misclassification

by [Kogevinas M](#), [Hagmar L](#)

Affiliation: Respiratory and Environmental Health Research Unit, Institut Municipal d'Investigació Mèdica, E-08003 Barcelona, Spain. kogevinas@imim.es

Key terms: [editorial](#); [feasibility](#); [misclassification](#); [occupational exposure](#); [register-based study](#)

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/15864901



This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

Use of routinely collected occupational exposure data in register-based studies— a trade off between feasibility and misclassification

When the *Scandinavian Journal of Work Environment & Health* publishes a paper that is based on routinely collected data, the first author of this editorial nearly always becomes jealous, and both authors frequently become frustrated. Jealous because the wide range of registers and records that are available in the Nordic countries and that can be combined and cleverly used do not exist in south European countries or, indeed, in almost any other countries. An example of such a clever use of register data is the paper by Nordby and his co-workers, which evaluates the reproductive and cancer toxicity of the pesticide mancozeb and is published in this issue of the *Scandinavian Journal of Work, Environment & Health* (1). Frustrated because even a clever use of such registers often does not provide enough information to reach a solid conclusion. This latter problem is not specific to studies based on routinely collected data since conclusions about toxic or preventive effects would rarely be based on a single study. Conclusions are based instead on a combination of evidence from epidemiologic studies evaluating different populations and animal and experimental studies, as, for example, is done by the International Agency for Research on Cancer (IARC) (2). The question is not whether any paper provides enough evidence to reach a solid conclusion, but, rather, whether a paper provides worthwhile information at all. If IARC re-evaluated mancozeb, would the study by Nordby and his co-workers (1) count? This is what we would like to discuss within a more general context of the use of census data, disease registers, or other records based on routinely collected data. To begin with, we would like to state that the use of such data should undoubtedly continue to be widely available and promoted. However, the use and reporting of data from these registers should be optimized to enable the extraction of as much and as valid information as possible. This is not always done.

Routinely collected data can be retrieved from a wide variety of sources, such as census information, employment data, immigration data, health-related registers, and the like. Health-related registers include a wealth of information, such as cancer registers, hospital in-patient registers, data on birth outcomes, and even specialized information, such as data on cerebral palsy or infantile autism. The difference between the Nordic countries and most other countries is not only the availability of the registers, but also the possibility of reliable linkage between them.

One of the main and very efficient applications of routinely collected data is their use for the production of descriptive statistics and for surveillance. This use may cover fairly straightforward analyses, such as the evaluation of geographic differences or time trends in disease incidence. For example, the evaluation of rates of testicular cancer showed puzzling differences between Finland, Denmark, Sweden, and Norway. This difference, in combination with more limited data on male congenital malformations, has helped theories be developed on the impact of endocrine disruptors (3). The use of routinely collected data for surveillance is extensive and may cover fairly simple or more elaborate analyses of data. For example, Karjalainen and his co-workers used data on work compensation that were linked to census data to derive estimates of the incidence of asthma by occupation (4). These estimates of the incidence of occupational asthma are among the most valid statistics available worldwide because of the high quality of occupational and disease information and the large numbers.

A second main application of routinely collected data is their uses in etiologic research. The paper by Nordby and his co-workers in this issue of the *Scandinavian Journal of Work, Environment & Health* is a good example of the linking of three different sources to develop proxies for exposure to a pesticide in order to evaluate several disease outcomes in relation to exposure.

There are several advantages, but also problems, with the use of these data in etiologic research. The main advantage is that a variety of data sources is, more or less, readily available. Recovery of this information has a cost that is certainly less than the costs of collecting primary data through interviews. In addition, it can also be quicker to examine already collected data, and such a technique can probably be assumed not to have some of the biases that could potentially affect other epidemiologic designs. The availability of register-based information on both exposure and health outcome makes several studies feasible that simply could not have taken place otherwise. The numerous cancer cohorts from the Nordic countries crucially depend on the availability of high-quality cancer information. Many of the results from the large Danish National Birth Cohort (5, 6) depend on the availability of information on hospitalization or other disease registers that compliment the data recovered through questionnaires. So far so good.

What are the problems then? In our view, the main problem is that of exposure misclassification. There are numerous examples of the use of register-based data in Nordic and other countries that started from a reasonable hypothesis but that were likely to be very imprecise when specific exposures are examined, such as studies on exposure to phenoxy herbicides and dioxins in relation to various forms of cancer, studies on solvent exposure and tumors of the central nervous system, or studies evaluating occupational or environmental exposures to radio frequencies or extremely low-frequency electromagnetic fields using death certificates (7, 8). The case of phenoxy herbicides and dioxins is illustrative. Classifying exposure to phenoxy herbicides through census data on jobs would have been extremely difficult in the first place. An evaluation of exposure to the dioxin contaminants of these phenoxy herbicides is virtually impossible without the use of much more detailed information on spraying practices and, realistically, without the use of exposure biomarkers. There are simply far too many assumptions on the degree of exposure that, according to recent knowledge, makes these studies totally uninformative. Studies evaluating a specific occupation and disease risk, for example, hairdressing and cancer (9), are probably better off. Why? The main difference is the degree of misclassification that is very high in studies examining specific exposures (eg, phenoxy acids and dioxins), but that can be expected to be lower in studies evaluating occupations (eg, hairdressing). Typically these latter studies would incorporate information from more than one census. Inevitably there will be persons who worked as hairdressers between two censuses and never declared themselves to be hairdressers, and there will also be persons who worked briefly as hairdressers during the census time and had therefore minimal exposure. Thus some misclassification is unavoidable also when occupation is assessed as the exposure variable.

It is important to remember that the scientific questions asked differ between these two types of studies. The studies on phenoxy herbicides and dioxins asked whether specific exposures were associated with cancer. In the studies on hairdressers, the question is precisely whether the occupation is at a higher risk. Even though exposure misclassification is probably lower when occupations are examined rather than exposures, the information deduced may, however, be of limited use. It is of course important to know whether an occupation does have an increased or decreased risk for disease and to evaluate time trends, but, inevitably, the next step is to identify which exposures define this risk. This type of information is frequently not available in register-based studies.

The study on mancozeb by Nordby and his co-workers (1) evaluates a specific exposure using proxy information. Some of it appears fairly valid as, for example, potato farming in itself and farmers' work input per year, whereas more blunt data on the annual sale of mancozeb in Norway and district level data on meteorology (fungal forecasts) seem less precise. How does this lack of precision affect the conclusions? It is likely that the resulting misclassification will be nondifferential, leading to a bias towards the null. The main problem in this well-designed study is that we have little understanding of the degree of misclassification. If there is one take-home message when this very interesting paper is read, it is that it would have been even more interesting if the authors had managed to do a validation

study of exposure, and, more specifically, of the exposure classifications used. Epidemiologic studies are riddled with assumptions and “expert assessments” that have proved to be wrong (10, 11). The counterargument against the need for validation studies is that the registers will then lose their attractiveness as a quick way to evaluate potential health effects.

Exposure information retrieved from registers and of relevance for health outcomes may have multiple uses, and it is extremely important to have access to and use such information, but it has its limitations. How great the limitations are depends on the degree of misclassification. If the misclassification is high, as it was when dioxins were evaluated using census data for occupation, then doing the study is not worthwhile. It is actually counterproductive, since the study would only add to confusion. If misclassification is less high (and we consciously use a vague term), then the results can be of use in risk assessment, and it may thus be worthwhile doing the study. If IARC re-evaluates mancozeb in the future, the paper by Nordby and his co-workers (1) should be considered definitively. However, if feasible, it would certainly pay to try and validate the exposure classification both in that study and more generally in epidemiologic studies relying on routinely collected exposure data from registers. There are unfortunately few examples of such validations. A not very encouraging result of such a validation was seen in a Swedish case-control study of chronic myeloid leukemia, in which the exposure prevalence to organic solvents was 17% according to interview data as compared with 4% when a job-exposure matrix was applied to census data (12).

References

1. Nordby K-C, Andersen A, Irgens LM, Kristensen P. Associations between indicators of mancozeb exposure, and thyroid cancer and neural tube defects in farmers' families. *Scand J Work Environ Health* 2005;31(2):89–96.
2. International Agency for Research on Cancer (IARC). Ethylenethiourea. Lyon: IARC; 2001. IARC monographs on the evaluation of carcinogenic risks to humans, vol 79.
3. Skakkebaek NE. Testicular dysgenesis syndrome: new epidemiological evidence. *Int J Androl* 2004;27:189–191.
4. Karjalainen A, Kurppa K, Martikainen R, Klaukka T, Karjalainen J. Work is related to a substantial portion of adult-onset asthma incidence in the Finnish population. *Am J Respir Crit Care Med* 2001;164:565–8.
5. Albertsen K, Andersen AM, Olsen J, Gronbaek M. Alcohol consumption during pregnancy and the risk of preterm delivery. *Am J Epidemiol* 2004;159:155–61.
6. Basso O, Christensen K, Olsen J. Fecundity and twinning: a study within the Danish National Birth Cohort. *Hum Reprod* 2004;10:2222–6.
7. Wiklund K, Lindefors BM, Holm LE. Risk of malignant lymphoma in Swedish agricultural and forestry workers. *Br J Ind Med* 1988;45:19–24.
8. Cocco P, Heineman EF, Dosemeci M. Occupational risk factors for cancer of the central nervous system (CNS) among US women. *Am J Ind Med* 1999;36:70–4.
9. Czene K, Tiikkaja S, Hemminki K. Cancer risks in hairdressers: assessment of carcinogenicity of hair dyes and gels. *Int J Cancer* 2003;105:108–12.
10. Kauppinen T, Teschke K, Astrakianakis G, Boffetta P, Colin D, Keefe A, et al. Assessment of exposure in an international study on cancer risks among pulp, paper, and paper product workers. *Am Ind Hyg Assoc J* 2002;63:254–61.
11. Kromhout H, Oostendorp Y, Heederik D, Boleij JSM. Agreement between semiquantitative exposure estimates and quantitative exposure measurements. *Am J Ind Med* 1987;12:551–62.
12. Strömberg U, Björk J. Incorporating group-level exposure information in case-control studies with missing data on dichotomous exposures. *Epidemiology* 2004;15:494–503.

Manolis Kogevinas¹ & Lars Hagmar²

¹ Respiratory and Environmental Health Research Unit
Institut Municipal d'Investigació Mèdica
E-08003 Barcelona, Spain
kogevinas@imim.es

² Department of Occupational and Environmental Medicine
Lund University Hospital
SE-22185 Lund, Sweden
lars.hagmar@ymed.lu.se

