RISK IDENTIFICATION BY REGISTER EPIDEMIOLOGY – A SWEDISH EXPERIENCE

H. Malker
National Board of Occupational Safety and Health, Solna, Sweden

Population registration in Sweden dates back to 1749. The personal identification number, unique for each individual, was introduced in the late 1940s. This number facilitates the linkage, at the individual level, of information from effect registers (e.g. the Cause of Death Registry or the Cancer Registry) and registers containing information on possible exposures or on occupation and industry. A linkage between an effect register and a census can be used to screen for high risk groups in different work environments. This usage is exemplified with studies on malignant pleural mesotheliomas and nasal cancer. The results were in agreement with previous knowledge about risk environments but also yielded some new hypotheses. Some problems in register epidemiology are reviewed. Comments are made on the use and usefulness of register epidemiology and also on possible study designs for the follow-up of the hypotheses arising from such studies.

More than a hundred years ago, Farr and Ogle in England started to use mortality data to detect high risk groups. One of their purposes was to isolate high risk groups in the work environment. In the U.K. that tradition has been kept up and descriptive statistics on mortality by occupational groups have been regularly published since.

In the 1950s the same technique as the one used in the U.K. was applied in the U.S. Today there are more than 15 reports, most of them from the U.S. dealing with descriptive statistics for screening high risk groups in the work environment (1).

Sweden has a tradition of population registration dating back to 1749. The unique personal identification number, which was introduced in the late 40s, provides, together with computer techniques, facilities for easy linkage of information systems.

The first Swedish census stored on magnetic tape is that of 1960. In the last decade data from linked nation-wide registries including effect data, e.g. cancer cases, and census data, e.g. occupation and industry, have become available for analysis (2, 3).

The purpose of this paper is to describe briefly the relevant registers in Sweden for linking occupational data with different outcomes, as well as the possibilities and
constraints of linked registers, using examples with special emphasis on asbestos and wood-dust induced cancer.

THE USE OF LINKED REGISTERS

Register epidemiology may be used to generate, refine and verify hypotheses. The possible use of the information system is determined by the specificity of the exposure information (4).

Only crude information, e.g. that on occupation or industry, is needed in order to use registers to screen for high risk groups. The purpose is to describe similarities or differences in disease patterns in order to generate hypotheses regarding possible risk factors. These can include hitherto unknown hazards as well as known hazards in new environments. To refine a hypothesis, one needs exposure information at least on a group level. To be able to verify a hypothesis one must have detailed exposure information regarding the time and level of exposure for individuals, and often also other information, e.g. on smoking and drinking habits and diet.

INFORMATION SYSTEMS IN SWEDEN

Sweden, like some other countries in Northern Europe, has nation-wide population-based "effect registers". The effect registers most commonly used in epidemiological studies are: the Cause of Death Registry (5), the Swedish Cancer Registry (6), the Medical Birth Records and the Registry of Congenital Malformations (7) and Inpatients Statistics (the Hospital Discharge Registry) (8), the latter covering the whole country from 1984.

The methods and quality of data collection are fairly similar in all the Nordic countries. The effect registers at national level usually contain fairly scanty information, e.g. identification number (including the year, month and date of birth), name, address, information about the disease and references to hospital records. At central level there should be only "hard" information that can be verified either by linkage to other registers or by reference back to the hospital records.

More detailed information is available at regional or local level. A priori knowledge and additional information such as concerning stages of cancer, smoking habits and drug usage, occupational history and estimates of occupational exposure at regional and local levels will probably be utilized more and become more important in future. These data sources are especially suitable for smaller studies when the effect studied is fairly common. The computerisation process at these levels is very rapid in Sweden, making linkage to other sources of information easier.

Census data in Sweden include the personal identification number as well as other variables such as occupation, industry, place of residence, rank etc. The unique personal identification number makes it possible to link census and effect data at individual level. This facilitates screening for risk groups.
Specific, individually based national exposure registers exist only for known hazardous agents such as quartz, asbestos, lead and to some extent trichlorethylene and cadmium (9). These registers can be linked to effect registers to ascertain that the TLVs are appropriate but also to detect or refute hypotheses about the etiology of other diseases, e.g. lung cancer and silica exposure.

The registers of workers exposed to lead, cadmium and trichlorethylene are examples of biological monitoring. More specific information on exposure can sometimes be obtained through trade unions and firms or within the occupational health care system. Using membership lists from the trade unions, specific job categories can very often be defined. These categories, together with information about exposure, may be used to refine hypotheses by linkage to the effect registers (10, 11). An intra-cohort case-control study can then be performed to verify the hypothesis. A company's administrative files may be used for specific follow up studies when linked to the effect registers. These files usually include particulars on when and where an individual has worked. This information, together with crude information about exposure or sometimes also accompanied by exposure evaluation based on hygiene measurements for different work environments within a firm, can then be linked to evaluate differences in health outcome. This type of study design is probably the most common type of cohort design (12). The use of personal identification numbers in such files is very important for the convenience of future linkage.

The Chemicals Commission has recommended to the Swedish Government to introduce legislation empowering the Government or the National Board of Occupational Safety and Health to force employers to save records on the workers in certain specific work environments (13).

In occupational health care network there are some computerized information systems including both effect and exposure data. One such computerized information system is the register initiated in the early 1970s by the Organization for Safety and Health in the Construction Industry (Bygghälsan). This data base has lately been linked to the cancer register and to the cause of death register, with the intention of evaluating possible long-term effects of the work environment.

There has been a demand for computerized information systems within the Swedish occupational health care service which include medical, hygienic and ergonomical aspects. These thoughts have been systematized in a nation-wide project (14). If this project is implemented, data will be gathered in the same way in every ward, making aggregation from different centres and linkage to long-term effect registers easier.

The Swedish Government has recommended that essential information in the patient records should be saved until the person dies or until he/she is 100 years of age. For a 10 per cent population sample everything should be stored permanently. The law requires only three years' storage. This recommendation includes occupational health care data, and it also covers ways of storing data in the event of closure (15 - 17).

There are also nation-wide data bases which contain effect data as well as data about possible hazards. One such register is the Information System on Occupationally Induced Injuries (ISA). This register includes work accidents, commuting accidents and work-induced diseases as well as possible causes. The most common cases in the
work-induced diseases group are musculoskeletal diseases, diseases of the central nervous system and sensory organs, diseases of respiratory organs and skin diseases (18).

In registers like ISA, hazard-specific cohorts can be aggregated and may then be linked to other effect registers in order to evaluate possible risks for other outcomes than those originally reported to ISA. This has been done with the information systems for accepted cases of work-induced silicosis and asbestosis.

SCREENING REGISTERS

A special kind of information systems available in Sweden are the population-based, nation-wide linked registers which can be used for screening purposes, e.g. the Cancer Environment Registry (CER) and the Death Registry. (To screen means to examine in order to separate into different groups.)

The CER contains data from the Cancer Registry, 1961 – 1979, and the 1960 census (19). It has also been suggested that the Cancer Registry should be linked to the 1970 census, in order to obtain census data e.g. occupations from at least two different points in time, but nothing has yet been decided (20).

The Death Registry actually contains two different parts – one of data for deaths from 1961 to 1970 and the 1960 census and the other containing data for deaths from 1971 to 1980 and the 1970 census. Work is in progress to merge these two registers.

The In-patient Register, the Medical Birth Records and the Register of Congenital Malformations are also linked to the 1975 and 1980 censuses.

It should be emphasized that the original purposes of the constituent registers on the national level are mostly administrative, very often as part of short- and long-term planning and administrative follow-up. This might influence the quality of the data and should always be kept in mind when drawing conclusions from screening surveys on this type of data set.

The screening registers are used:

1. To describe the disease pattern within a specified group

Most screening studies describe the diseases pattern either in an occupation, an industry or in an occupation within an industry (1, 21).

One example: in a study on CER 1961 – 1973 we found an excess of skin cancer of the upper limbs among physicians (21). This excess rate could be followed up to 1979. There are several possible hypotheses for the etiology. One concerns a disinfectant for the hands and upper limbs, containing a suspected carcinogen, which had been in use in Swedish hospitals for many years. Another has to do with X-ray exposure, although according to the literature the hazard to radiologists has been minimal for many decades. When following up the specialties eight out of the twelve cases turned out to be radiologists. Also, an unexpectedly large number of physicians had the double specialty of radiologist and gynaecologist (22). This study is now continuing in search for further clues.

2. To describe disease patterns in different groups within a specific field

The purpose is to describe the disease patterns for specific subgroups and to evaluate similarities and differences which may lead to etiological hypotheses.
A study was performed on CER data, for the metal industry (21). This industry included iron and steel plants, pig iron and steel foundries, other metal plants, iron and steel making and other metal manufacturing. The cancer risks in each of these industries and 15 subgroups of job categories were analysed (21).

3. To describe different risk groups for a specific disease

The CER has been used to study the patterns of risk groups for a number of different cancer sites. Screening for a certain site may be performed for specific histopathological types e.g. pleural mesotheliomas, nasal cancer (adenocarcinomas, squamous cell carcinoma and other histology types) or for different subsites e.g. biliary tract cancer (gall bladder and other biliary tract cancers (23–25).

One way to do this is to start with the major one-digit classification of occupation and industry and, where excess rates are found at this level, to continue to the general two-digit number and then to follow those excess risks to the specific three-digit classification numbers. But in order not to miss any excess rates on the specific three-digit level a crude analysis for every occupation, industry and occupation within industry with more than 500 workers is also performed and then evaluated for excess risks.

The papers 23 and 24 are examples of such systematic assessments of pleural mesothelioma and nasal cancer occurrence in the Swedish working force 1961–79. The mesothelioma study was consistent with the data published earlier, relating pleural mesothelioma to occupational exposure, e.g. shipbuilding and repair (Ind = 360, SIR = 3.0, 25 cases), manufacturing of railroad equipment (Ind = 362, SIR = 7.2, 11 cases), and plumbers and pipe fitters (Occ = 754, SIR = 4.8, 13 cases). But there were also some unexpected associations, not reported in the literature earlier, e.g. sugar factories and refineries (Ind = 206, SIR = 6.3, 7 cases) and paper and allied products (Ind = 26, SIR = 2.9, 25 cases). In these cases too, the excess rate was related to heavy asbestos exposure (25–28).

The results in the study of nasal cancer were consistent with earlier findings, with the associations often specific for the histologic type. The highest risks were observed for nasal adenocarcinoma among wood workers, especially among furniture and cabinet makers. Twenty-five per cent of all nasal adenocarcinomas over the 19-year study period were among furniture makers, an industry accounting for less than two per cent on the male work force in Sweden in 1960.

As in other studies, the association with woodworking appears specific for adenocarcinoma, with little or no excess risk of squamous or other cell types. The excess was predominantly among furniture makers, although there was some suggestion of a residual increase in risk linked to woodworking trades in related industries suggesting that exposures responsible for the excess nasal cancer risk among furniture makers may be found also in other woodworking jobs, although at lower levels.

An approximate two-fold increase in nasal squamous cell carcinoma was observed among textile workers, with an excess risk among both men and women. The finding is noteworthy because of previously published reports of the same results.
These studies could then be used to point out work-environments where the Labour Inspectorate should be especially vigilant concerning asbestos exposure and to refute the hypothesis of paper dust causing mesotheliomas of the pleura.

4. To determine disease pattern for an exposure

To refine a hypothesis, e.g. from animal studies or where the hypothesis is fairly incomplete, a first crude screening study can be done on such inadequate exposure data in occupation or industry.

Formaldehyde is one example where animal studies have been positive and linked register studies have been used to set a background for further epidemiological studies (29).

Silica dust exposure is another example of a study on linked registers from all the Nordic countries being used in an attempt to clarify an earlier hypothesis (30).

One difficulty is the problem of evaluating negative results obtained in screening studies. To be able to make at least rough evaluations, one needs a priori information about the proportion exposed in every group and also some estimate of the exposure level in the different groups.

Another circumstance that never should be forgotten in any epidemiological study is the multifactorial etiology of most diseases. This is even more important to remember when evaluating screening studies that most often are analysed unfactorially.

As underlined in the examples, descriptive screening studies have to be verified by specific surveys, e.g. intra-cohort case-control studies. As mentioned earlier, one way is to use half of the cases to create specific hypotheses about hazardous agents (refining) and to use the other half for an intra-cohort case-control study to verify the hypothesis.

5. For model assessment, education, etc.

Very often the choice of a proper comparative population is essential for the evaluation of an epidemiological study. With large linked registers, different reference populations and standardisations can easily be used. An evaluation of the outcome when different reference populations have been used may demonstrate which type(s) of risk factor may be involved, e.g. lifestyle or occupational factors (31).

Data from large linked registers can also be used to implement different types of models or in education to get a feeling for the impact of different factors or analytical tools.

As example, one can employ models so as to be able to distinguish between time-age and birth-cohort-effects (32).

PROBLEMS IN REGISTER EPIDEMIOLOGY

Register epidemiology on population-based screening registers shares most of the well-known problems of traditionally performed epidemiological studies, but in addition there are some others, such as the multiple comparison problem, the problem of diluted and inexact exposure information, the problem of the population base being defined at only one or a few points in time and the problem of intercurrent mortality of other diseases when the outcome variable is morbidity in a certain disease.
The multiple comparison problem

When analysing the data in the CER there are more than 50 sites, 300 histopathology codes and about 6,000 occupational, industrial, or single-industry occupation codes that are worth analysing. In a traditional statistical sense this would lead to a multiple comparison problem with many ‘differences’ by chance.

When one looks at the results of screening studies based on registers, this does not seem to happen. One explanation might be that most of the major sources of error involved tend to exceed the ‘random errors’ and therefore this problem does not seem to occur to an extent that would be expected.

The problem of diluted exposure

The purpose of ‘fishing expeditions’, i.e. screening for high risk groups, is partly to describe a certain pattern. If only a fraction of the target population is exposed, this will lead to an underestimation of the real risk. If any information about exposure is available, this may form a basis for further studies regarding etiology. Furthermore, with information about the proportion exposed, crude limits for risk estimations may be performed. Even so one should be very careful about drawing any conclusions about the etiology from descriptive analysis alone.

The problem of the population base defined at discrete points in time

Where it is possible for descriptive purposes to link nation-wide registers, usually in countries with unique personal identification numbers, this is usually done with one or a few causes. This fact may impose a number of limitations on the analysis, e.g.: a. trend analysis can only be carried out under certain, very often inappropriate assumptions, b. a cut off for latency time for long term effects like cancer or heart diseases can very seldom be used in the analysis, c. if there is selective migration for instance between occupations over time, wrong conclusions may be drawn, d. it is very hard to draw any definite conclusions even on the basis of good a priori information about the level of dilution or mixed exposure, because of migration between occupations and industries. However, in the few studies done in Sweden following clues from descriptive studies the impact of this fault has been very limited. The explanation might be that this problem will tend to give negative results (where no clues are generated), so that they will hardly be followed by any ad hoc study.

COMPARISONS BETWEEN STUDIES

As is often said about epidemiological studies, ‘consistency is more valuable than statistical significance’. This is especially true when screening for high risk groups. The results of one screening study can very seldom be evaluated on their own. It is the consistency between several studies that makes the results interesting. This is also a way to solve the multiple comparison problem. Two examples of this are the consistency of
findings for lung cancer among 'butchers' in Denmark, England and Sweden (33) and the excess risk of lung cancer among welders (34).

When presenting the results from large screening studies a comparison should always be made with other studies, as in Dubrow and Wegman's work in 'Cancer Victims in Massachusetts' (35). In assessing occupational health hazards in the work environment, consistency between the results from screening registers from different countries together with ad hoc studies with specific exposure information is essential to the risk evaluation (1).

Confounders factors

The screening registers by themselves very seldom include any data e.g. on smoking habits. Sometimes there are data or a priori information to evaluate the impact of such a factor on the results. In Sweden, for example, there is a study from 1963 on the smoking habits of a random sample of the Swedish population. This survey has been used together with the CER to obtain rough estimates of the impact of smoking habits on lung cancer in occupational groups (31).

The example mentioned above refer mainly to the Cancer Environment Registry (CER). The same type of study has also been performed with the other linked registers mentioned earlier.

Although the diagnosis in some of those other registers is not always as good as the cancer diagnosis in the CER, the exposure data are still the limiting factor.

Exposure data

Exposure data are usually kept only for known or suspected hazardous agents. The dilemma is that good exposure information is needed to verify new leads. For diseases with a short latency time, where the number exposed is large enough or the disease fairly common, this can be solved by creating cohorts followed prospectively and then link them to effect registers.

The real problem is to get exposure information when the outcome is a disease with a long latency time. One possibility is to use a questionnaire and a case-control study design. Another is to use the administrative files from different companies together with lists of agents used. Today the administrative files from companies, trade unions etc. are fairly easy to obtain. The problem is to evaluate which agents the individuals were exposed to. Very often one has to rely on what workers, hygienists, engineers and health-physicians remember. Fairly often, though, account books are kept of the types and quantities of chemicals the firm has bought, and these can be used. From some psychosocial or ergonomic hazards or some general factor in the work environment, the job itself may be the only identifying variable. Very little has been done in this field, a field that will probably be important in the future.

Register problems

No linked register is better than the constituent registers.
The target population

One problem is to define a proper target or study population e.g. the occupation may be defined as the usual, the latest, the current or the occupation at the time of a certain census. Usually the occupational codes are taken from the International Standard for Classification of Occupation (ISCO), but there are several revisions and special national versions of the classification. The work (and exposure) represented by one occupational code may also differ between countries. These facts must be clarified before any comparisons can be made.

Information about exposure is sometimes also needed when looking at the description of what is included in a code. For example, embalmers in the U.S. and Canada are exposed to formaldehyde, which is not the case in Sweden.

Dropout

The influence of dropout, e.g. missing individuals or missing variables, can be estimated if the dropout is known. Very often, though, there is a certain level of unknown dropouts that cannot be quantified, especially in nation-wide registers. The dropout should always be treated as selective until proved otherwise.

The quality

In Sweden the quality of the data in the effect registers is most often superior to that of the exposure registers. Essential to all linkage is the identification on which the linkage is done. There are always cases where the matching is unsuccessful, for instance because of a false identification number or differences of target population caused by migration or missing persons etc. In countries with unique personal identification numbers these faults are usually so small that they do not affect the results to any great extent, even if they have been shown to be selective (36).

There are also some general factors affecting the quality of a register, namely the purpose of the register and demand for it (2).

The usefulness of register epidemiology in risk identification

The discussion about problems in register epidemiology and the comparison with traditionally performed cohort studies suggest good reliability in the register-based results, provided the analyses are well performed and the results carefully interpreted. The results from analysis of large linked nation-wide registers must often need in-depth studies to confirm or reject hypotheses on causality between work environment and disease. To be able to interpret the results, one must be aware of the problems and faults that may exist. In most cases, these tend to equalize the results. Very seldom will an epidemiological study, by itself, be sufficient for making any conclusions about causality. According to Hill (37), at least nine points need to be considered, such as consistency of the association, strength of the association, specificity in the results, time dependence, dose-response and, last but not least, biological plausibility.
REFERENCES

14. Foretagshaldsoardseeklinationen: Arbetarskyddsfonden 1984 (In Swedish)
16. Huvudbetankande av journalutredningen: Socialdepartementet, SOU 1984/73. (In Swedish)
18. National Board of Occupational Safety and Health Statistics Sweden, Stockholm 1983. (Summary in English)

Sažetak

PROCJENA RIZIKA IZLOŽENOSTI PREMA EPIDEMIOLOŠKOM REGISTRU

Registri u Švedskoj postoje od 1749. godine. Pored nekoliko godina ovog delovanja Švedske je osnovni identifikacijski hrv jedinstven za svaku osobu, koji olakšava povezivanje podataka iz različitih registara. Tako se mogu povezati podaci za pojedinca iz npr. Registra uzroka smrti i Registra malignih bolesti i registara koji sadrže podatke o mogućoj izloženosti ili uticaju zaposlenja. Ovakvim povezivanjem mogu se dobiti podaci u visokoriznim skupinama u različitim radnim okolinama. Izneseni su primjeri povezivanja malignog pleuralnog mezioteliom i raka nosa. Rezultati se služe u dosadašnjim saznanjima o mogućim okolinskim faktorima nastanka ovih bolesti, no iznose se i neke nove pretpostavke. Pregledno su prikazani neki problemi vezani uz epidemiološke registre. Iznesene su neke upute o mogućnosti upotrebe i korisnosti ovakvih registara, kao i o planiranju prospektivnih studija temeljenih na dobivenim pretpostavkama.

Nacionalno ministerstvo za sigurnost na radu i medicinu rada, Solna, Švedska