VU Research Portal

Risk factors for ski injuries: a crash course of epidemiologic methods with emphasis on comparability in experiments and case control studies

Knipschild, P.G.; Bouter, L.M.

published in

Skiing Trauma and Safety: 7th International Symposium
1989

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Knipschild, P. G., & Bouter, L. M. (1989). Risk factors for ski injuries: a crash course of epidemiologic methods with emphasis on comparability in experiments and case control studies. In *Skiing Trauma and Safety: 7th International Symposium* (pp. 9-17). ASTM.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Risk Factors for Ski Injuries: A Crash Course of Epidemiologic Methods with Emphasis on Comparability in Experiments and Case-Control Studies

REFERENCE: Knipschild, P. G. and Bouter, L. M., "Risk Factors for Ski Injuries: A Crash Course of Epidemiologic Methods with Emphasis on Comparability in Experiments and Case-Control Studies," in *Skiing Trauma and Safety: Seventh International Symposium, ASTM STP 1022*, Robert J. Johnson, C. D. Mote, Jr., and Marc-Hervé Binet, Eds., American Society for Testing and Materials, Philadelphia, 1989, pp. 9-17.

ABSTRACT: The incidence rate of ski injuries should be studied as a function of its risk factors. Epidemiologic studies into this matter are very complicated. Central issues are comparability of baseline prognosis, comparability of measurements (effects in cohort studies and risk factors in case-control studies), and comparability of external circumstances. In experimental studies this may be achieved by randomization, blinding, and placebo intervention. The main tools in nonexperimental studies to prevent incomparability are deliberate selection and multivariate analysis. An outline is given of the design of experimental studies with increased efficiency. As for case-control studies, special attention is paid to the definition of the source population and possible ways to reduce measurement incomparability. The text provides many examples related to putative ski injury risk factors.

KEY WORDS: epidemiology, ski injuries, risk factors, comparability, experiments, cohort studies, case-control studies

Epidemiology is not restricted to epidemics of infectious diseases. The present scope is much broader. Nowadays epidemiology deals with the occurrence of all sorts of defects and diseases, such as congenital malformations, cancers, dementia, and also injuries, including ski injuries. Studying the mere occurrence of these injuries can be rather tedious. Often, it is more exciting to study the occurrence of ski injuries as a function of its determinants.

This contribution deals with epidemiologic methods to study the etiology (that is, risk factors) of ski injuries. Special attention will be paid to the theoretical background of useful study designs. It may be seen as a "crash course," because it is impossible to discuss in one paper what other epidemiologists have devoted whole books to.

First, some remarks are made on the definition of a ski injury and the difference between prevalence and incidence. Second, the rationale behind epidemiologic studies into the determinants of ski injury incidence is given. Third, the concept of comparability is explained, starting from the conventional randomized, double-blind, placebo-controlled experiment, and suggestions are given to increase its efficiency. Fourth, because experimental studies are not feasible in many situations, an introduction is given to the design of nonexperimental studies, with the same emphasis on comparability. Special attention is paid to the design of case-control studies. Finally, the problem of incomparability of measurements in case-control studies is discussed.

¹Epidemiologists, Department of Epidemiology and Health Care Research, University of Limburg, P.O. Box 616, 6200 MD Maastricht, the Netherlands.

Occurrence

It is surprising to see how often the occurrence of ski injuries is discussed without any clear definition. It is time, therefore, for ski injuries to be defined in a way that is logical, understandable, and acceptable to all people who are interested in the epidemiology of these injuries.

Our description would be the following: It must evidently concern an injury received during skiing. It is impossible to sustain a ski injury in a hotel, even when one slips on the doormat during a ski holiday. The consequences of an accident on the way to or back from the ski run cannot contribute to the occurrence as well. Slipping in a ski lift or being hit by a skier on the ski run are included in the determinants of ski injuries. It makes sense to list all potential ski injury situations and then see how much agreement there is among experts.

It is also necessary to be specific on what is called an injury. If the definition of an injury is too broad, including every scratch or hematoma, then no skier escapes injury. A statement that ski injuries occur in 100% of all skiers has no value. Our description of an injury would be stricter: all fractures and luxations, and those other injuries that require medical control. The latter include many ligament injuries, such as "ski thumb" and "ski knee," but also concussion of the brain and significant lacerations (that is, those requiring sutures). Again, it makes sense to list all ski injuries that require the help of a health-care professional and then study the degree of agreement among experts. We are not very happy with a definition that is used quite often in the literature, namely that a ski injury must restrict the victim's activities for a certain period (for instance, a day or more). When ski injuries are counted, of course, it remains useful to distinguish subcategories according to location, severity, and so on.

To sum up, ski injuries are received during skiing and medical control is necessary. Registration of ski injuries at the medical center in a ski resort leads to underreporting. Not all necessary medical control takes place, whereas some control takes place elsewhere. Investigators of the occurrence of ski injuries should complete their registration with the help of additional questionnaires. This calls for studies into the validity of these questionnaires in terms of sensitivity (underreporting, once again) and specificity (overreporting).

When a person has more than one ski injury, he or she is preferably counted as one injured person. Counting the number of injured persons is rather meaningless when there is no knowledge of the reference population. It is better to study the occurrence as a fraction with a numerator and a denominator: the number of (persons with) ski injuries in a certain population. This opens the possibility to compare the occurrence rate between groups of skiers, for instance at different times, between areas, and among groups with different characteristics.

So far, the word "occurrence" has been used in this paper. It has two time-related meanings and so it is necessary to distinguish between prevalence and incidence. Prevalence concerns the number of existing ski injuries at a certain point in time. For example, one may wonder how many skiers have a weak ankle due to a ski injury in the past. When the occurrence of ski injuries is discussed, nearly always incidence is meant: the number of new ski injuries in a certain period of time. Strictly speaking, incidence can only be studied in a population that has a 0% prevalence of (that) ski injury. It is the rate (speed) with which ski injuries arise in a population at risk. This incidence rate is best expressed as injuries per hour of skiing. Then the assumption is made, however, that the incidence is constant over the first, second, and later hours. A measure used quite often in the literature is the number of (new) ski injuries per 1000 ski person days. Comparisons between groups are still valid now if it may be assumed that the number of ski hours per day is the same between these groups, at least on the average (for example, 5 h per day).

To be less abstract, it is perhaps desirable to translate the incidence rate for a group of skiers into the risk for an individual skier. For instance, if the incidence rate of ski injuries is 3 per 1000 ski person days in a certain population, this implies a 3% risk when the holiday consists of 10 days of skiing. People on a ski holiday for 1 week run a 2% risk, and those who go for a

fortnight of ski pleasure expose themselves to a 4% risk of a ski injury. This, again, is an average among members of a group of skiers. Within that group, there may be large differences between subgroups with certain risk factor levels.

Risk Factors

No single disease is determined monocausally. Of course, some diseases occur only when a certain risk factor is present (for example, tuberculosis), but even then other factors play a very important role. As to ski injuries, the situation probably is very complicated. Potential risk factors can be subdivided into many categories, such as skill, physical condition, social habits, psychological profile, ski equipment, and the ski run environment. Within the latter category, for example, it is useful to distinguish between degree of difficulty, crowdedness, sight, temperature, snow conditions, and so forth. The many potential risk factors are, moreover, interrelated. This implies that studying the separate contribution of a certain risk factor to the level of the ski injury incidence is far from easy, even more so if the study is after the attributable risk in a quantitative sense. It is questionable whether this puzzle can ever be solved completely.

Untroubled by this complicatedness, some so-called experts pour out a large number of recommendations over the skier. These are often based on a mixture of plausibility, casuistry, and prejudice. Sometimes a quasi-empirical reasoning is given to stress the importance of certain preventive measures. For instance, suppose the "bad" habits of 100 persons with a ski injury are scored at the medical center of a ski resort, and it appears that one in three has enjoyed a very pleasant evening the day before, including a more than moderate alcohol consumption; this, of course, does not mean that one in every three ski injuries is caused by alcohol. It is possible that one in every three skiers who has received no injury also had a party the night before. A second relatively frequent fallacy is the following: Differences in the ski injury incidence at an aggregated level are translated too easily to the individual level. For example, the incidence of ski injuries appears to have decreased in the 1970s. In the same period an increasing number of educational brochures and booklets have been published. This does not necessarily mean that education has reduced the incidence. The improvement of ski equipment and of the ski run environment during the same period might be a better explanation.

Unraveling the causes of ski injuries is only possible with good empirical studies. Partly, these may be performed in the comfort of a laboratory, but in the end the answer is in the field. There, the plausibility of ideas generated in the laboratory should be tested further. If a strong association is found repeatedly, under different circumstances, between a risk factor (for example, binding adjustment) and certain ski injuries (for example, lower extremity equipment-related), this not only points to causality, but also quantifies the impact of this risk factor.

At the moment, the knowledge of many putative risk factors for ski injuries is sparse. In performing epidemiologic studies two ways are open: experimental and nonexperimental studies.

Experimental Studies

Characteristic of experimental studies is that the investigator intervenes in one (or more) putative risk factor(s) and then observes whether and to what extent this intervention changes the incidence. To measure a change, the incidence needs to be known in the situation where the manipulated risk factor is, in fact, not a risk factor—to put it differently, the incidence "under the null hypothesis." To measure this baseline incidence, one might observe the incidence twice in a group of the same persons who are subsequently exposed and not exposed to the intervention under study. For ski injury research this is not feasible, partly because an injury in the first study period makes skiers unsuitable to participate in the second study period. A solution to this

problem is to observe two groups at the same time: an intervention group and a control group, the latter not being manipulated or manipulated in a different way. It is very important now to have comparability of baseline prognosis. The groups who are to be contrasted should only differ prognostically as far as the intervention changes the incidence. Suppose, for instance, one wants to study the effect of ski lessons on the incidence of ski injuries; it is likely that ski lessons are more often taken by skiers with less experience, and that people with little experience run an increased injury risk. It is not valid now to compare the incidence of ski injuries in a group of people who have ski lessons, but who are also less experienced, with the incidence in a group of people who do not have ski lessons, but who also are more experienced. If the study goal is to see whether ski lessons prevent injuries, then the level of ski experience (and of all other risk factors) needs to be the same in both groups.

Analogously to a therapeutic experiment, a simple procedure to pursue this is randomization. First, a "cohort" of skiers is defined and then subdivided into two (or more) subcohorts purely by chance. Next, one of the subcohorts is exposed to ski lessons and the other is not. Random allocation is performed to remove any judgment by the investigators or participants as to who will have ski lessons. By following this procedure it is hoped that the subcohorts are prognostically comparable with respect to the ski injury incidence, before the intervention takes place. When the subcohorts are large enough, it may be assumed that the level of all unknown risk factors is the same as well as the distribution of known risk factors such as ski experience. An additional way to make the intervention and control group prognostically comparable is restriction. An epidemiologic study about ski lessons in relation to ski injuries might be restricted meaningfully to people without much ski experience.

A second point of attention is the comparability of (effect) measurements. Measuring the incidence in the intervention and the control group "by two standards" will invalidate the results. Of course, all effects of interest must be measured with the utmost precision, but there is a limit to everything. In practice, sometimes compromises are necessary for feasibility reasons, and one must be content with only rough information about the effects. Then it is important that the efforts to discover effects are equally strenuous in both groups. It is better to have *equally rough measurements in both than measurements which are, for instance, more valid in the control group than in the intervention group*. Otherwise, incidence differences might be found that are related purely to differences in registration, while the intervention studied has no causal meaning in reality. As an example, take an investigation that is started to see whether people with badly adjusted ski bindings run an increased injury risk. Two groups with the same baseline prognosis are observed, one with badly adjusted and the other with well adjusted ski bindings. The design looks good, but investigators do have their prejudices. They may note a less serious injury more readily when it concerns a person with badly adjusted ski bindings, or they may consider an injury not new, but already existing, when it concerns a person with well adjusted ski bindings.

Analogously to a therapeutic experiment, this problem can be solved best by blinding. The investigator who gathers the information about the effects must be unaware whether it concerns a person with or without the putative risk factor, in this example with badly or well adjusted ski bindings. This can be achieved by having the effect measurements taken by someone who is not the main investigator. The former does not even need to know the specific study goal. A self-evident supplement to blinding concerns the assessment, beforehand, of precise effect criteria. It must be stated in the study protocol how intensively the participating skiers are to be examined as to the presence of new ski injuries and which injuries are to be counted as such. When the participants need to be interviewed to complete the ski injury registration, it is desirable to blind them as well. In that case, the study is called double-blind.

Comparability of baseline prognosis and of (effect) measurements makes the study internally valid. There still is, however, a third point of attention. It is important, also, to consider the independent consequences of the intervention on other risk factors. With respect to this, a distinction must be made between a purely scientific and a more pragmatic intervention study. If

we restrict ourselves to the scientific question whether a putative risk factor has a causal meaning, then it is necessary to assign all extra maneuvers that accompany the intervention to the control group as well. The intervention may introduce a change in other risk factors that do not belong to the direct causal chain. This lack of comparability of external circumstances may confound the association between putative risk factors of interest and effects. For instance, in an experiment to study whether warming-up exercises decrease the ski injury incidence, it is easily imaginable that good warming-up, under supervision or on the advice of a ski expert, also leads to an increased attention to other putative risk factors, such as the general physical condition, night's rest, alcohol consumption, and so forth. The control group who is deprived of good warming-up does not get this extra attention. It is also possible that persons in the control group take some other preventive precaution because they are not allowed to have their warming-up.

One solution to this problem is to compare the effect of two plausible sorts of warming-up with each other. If a comparison is needed between warming-up and no warming-up, the latter should be given as a placebo intervention (that is, a dummy warming-up), which can be assumed to have no bearing on the ski injury risk. This makes the external circumstances more comparable and also provides an opportunity to blind the participants and the person who performs the effect measurements.

Now that the principles of a blinded experiment have been explained, there remains the problem of feasibility. Randomization, blinding, and placebo intervention, which are the main tools to achieve comparability, are often not easy to perform due to ethical restrictions. Moreover, there is the problem of numbers. Published epidemiologic studies show that of every 100 people who return from their ski holiday, only a few have a ski injury. At this low incidence rate very large numbers of participants are needed to detect differences between skiers with different risk factor levels. For instance, it is thought that fitness training before the ski holiday decreases the injury risk among skiers who are in a bad physical condition, perhaps by as much as 50%. In an experiment with no less than 500 participants, half having fitness training and half placebo training, the expected incidence rates per week (35 hours of skiing) will be about 2 and 4%, respectively. Unfortunately, this difference is not statistically significant at $\alpha = 0.05$.

There are several ways to increase the "power" of a study. The number of participants can be augmented further, but it is difficult and expensive to perform an experiment with many thousands of participants (even if it is a multicenter trial, which also introduces new logistic problems). However, the incidence may be increased by selection of participants. An experiment about the effect of fitness training on ski injury risk could be restricted to a group with an increased risk, for example, beginners at a ski run for more experienced skiers (just for the sake of argument!). With 2×250 skiers, this could lead to incidence rates of 10 and 20% per week (again with a risk ratio of 0.50), and there we have our beloved statistical significance.

There is also a second way to increase the incidence, namely by adopting—at least in the study—less strict criteria for the effect parameter, in this case ski injuries. Less severe injuries, which do not require medical control, could be included. Nearly always, ski injuries are the result of a fall, but not every fall leads to an injury. The number of falls may be viewed as an indication of the (much lower) incidence of real injuries. Of course, if falls are chosen as an effect parameter, clear criteria must be given beforehand as to which falls are to be counted as such. Similarly, it is very important to achieve comparability to effect measurements, for instance through blinding. If only a limited number of skiers can participate in an experiment, it may even be considered to include all near falls.

Nonexperimental Studies

With respect to many research questions concerning the etiology of ski injuries, it is impossible or preliminary to perform epidemiologic experiments because of ethical restraints and for reasons of feasibility. In that case the position may be taken that there is nothing left but to fall back to what is plausible in addition to what is learned from laboratory experiments. The alter-

native is to get involved in nonexperimental studies (sometimes called "observational"). We think that these can indicate whether certain putative risk factors may be important and which risk factors must be studied further in an experimental design. Of course, nonexperimental studies are only informative if they are performed reasonably well, mimicking the basic ideas of comparability of experiments. Randomization, blinding, and placebo intervention are nothing more than (very powerful) tools to achieve comparability. The main tools in nonexperimental studies to prevent incomparability are deliberate selection and multivariate analysis.

An experiment is a cohort study in which the investigator decides (although randomly) who is and who is not exposed to the risk factor under study. In an "observational" cohort study no intervention takes place and the investigator simply stands there beside the ski run to observe what happens to groups of skiers with certain characteristics, certain habits, under certain circumstances. In its simplest form the cohort is subdivided into two groups, with and without a certain risk factor of interest, all members of the cohort having no ski injury at the start of the study. Next, the incidence of ski injuries is measured in the contrasted groups. The difference or the ratio of the two incidence rates indicates the quantitative meaning of this risk factor in terms of etiology. Instead of studying two subcohorts, the experience of more groups can also be compared, when the putative risk factor is not dichotomous (for example, men or women), but nominal (for example, morning, early and late afternoon), ordinal (for example, blue, red, and black ski run), semicontinuous (for example, number of alcoholic drinks a day), or continuous (for example, body mass index).

Naturally, the several subcohorts should be prognostically comparable in relation to all other (external) risk factors. Where randomization and placebo intervention cannot be executed, deliberate selection with respect to certain external risk factors can solve this problem (for example, restriction to beginners). An alternative is stratification (for example, combined analysis for beginners and more advanced skiers). Because there are probably many risk factors involved in the etiology of ski injuries, the confounding effect of all known external risk factors can only be adjusted for in a multivariate analysis. A weak spot of this statistical solution is that complete adjustment is not possible when there are external risk factors that are still unknown or that cannot be measured exactly. For more detailed information about multivariate analysis techniques, such as logistic regression, the reader is referred to manuals of biostatistics.

There are no major ethical restraints to these "observational" studies and the investigator is not busy manipulating the participants. Therefore, cohort studies can often be performed with very large groups. For instance, in a cohort study 5000 (instead of 500) skiers could be studied, half of whom have fitness training before the ski holiday. In the example given in the section on experimental studies, this would lead to 50 ski injury victims among the skiers with fitness training (2%) and 100 victims in the referent group (4%).

Much of the information on the association between fitness training and the incidence of ski injuries is found in only 150 of all 5000 participants. Of these "cases," one in every three had fitness training. In the very large group of 4850 uninjured skiers about one in every two had fitness training. To estimate the frequency of fitness training among these "controls," data collection in a sample of, for example, 300 skiers would have given the same estimate of about 50% fitness training. In that case, only 450 study participants would have done the job. This last "trick" brings us to a very efficient study design, the so-called case-control study. For a fast and remunerative orientation as to which risk factors are associated with ski injuries, performing a case-control study seems to be the obvious way.

In an experimental or nonexperimental cohort study the investigator starts by defining a study population of skiers who are candidates for receiving a ski injury (Step 1). Within this population the level of all putative risk factors is recorded for every participant (Step 2). Next, the incidence of ski injuries is measured in subpopulations with different risk factor profiles (Step 3). In conformity with the logical sequence, first the putative causes and then the effects are recorded in the study population (cohort).

Also in a case-control study the investigator starts by defining a (source) population of skiers

who are candidates likely to have a ski injury (Step 1). Recording the risk factor is, however, postponed. One waits until cases of ski injury emerge from the source population. When the incidence is low (such as with ski injuries), a sample of controls is taken from the source population. So, the cases and controls are recorded first (Step 2). Next, the risk factor profile in both groups is measured "retrospectively" (that is, going back to a moment preceding any ski injury, for instance, the moment of definition of the source population) (Step 3). This is not in conformity with the logical sequence, but for reasons of efficiency, first the effects and then the putative causes are recorded in the source population.

The first step in case-controlling is often not mentioned explicitly in existing publications. It is, nevertheless, important to be fully aware which source population provides the cases. These must be restricted to those persons with new ski injuries for whom a source population can be defined reasonably well. At the same time, a sample of the same source population must be drawn from persons who could have received a ski injury, but who have escaped it for one reason or another (associated with their risk factor profile?). Uninjured skiers who are members of a different source population should not be included in the study.

Next, the risk factor profile of all cases and controls (within the source population) is to be measured "retrospectively." In a cohort study, the effects should be measured as comparably as possible, preferably blinded. In a case-control study the putative causes are measured. Now the comparability of risk factor measurements has the highest priority. This also indicates a puzzling problem for case-control studies, because blinding is nearly impossible. Those who have received a ski injury are already known. It may even be that the level of some risk factors is changed due to the injury. If the choice falls on cases who have been coping with their injury for a long time already (so-called prevalent cases), the information collected may be "biased." For instance, it is thinkable that, in a case-control study about body weight in relation to ski injury incidence, persons with longer lasting ski injuries gain weight because of their inactivity. If the case group consists mainly of skiers with "chronic" injuries and their weight (body mass index) is compared with the weight of a control group, a spurious association between body weight and ski injuries may be demonstrated. In a case-control study, information should be collected shortly after the accident leading to ski injury (that is, among incident cases and, of course, comparable controls). Direct measurements can only be taken of putative risk factors for which it may be assumed that any injury does not change their level immediately. Of course, incomparability of the measurements should be strenuously prevented.

Whenever risk factors cannot be measured directly, participants in a case-control study must be interviewed. The interview should concern the risk factor profile immediately before the ski injury incident. It is easy now to blind the investigator gathering the information. A simple way to do this is by not having personal, oral interviews with the cases and controls, but to send them a postal questionnaire. Preferably, this questionnaire is as structured as possible and contains only precoded questions. This prevents measurement incomparability by the investigator preparing the data analysis for the computer. If personal interviews are necessary, it is sometimes possible to hide the main study goals from the interviewers and the participants, but in practice this does not appear to be very successful and it still may lead to invalid information about external risk factors.

Where blinding fails, there only remains structuration of validated questionnaires. Now there is the problem whether the cases (and the controls) are willing to provide answers agreeing with reality to all questions. In general, people with a ski injury will recall the circumstances of the accident and their risky behavior relatively well. Maybe even too well, worried as they are why it all happened to them. Some of them even go and study the literature on ski injury prevention! They may give socially desirable answers to please the investigator. This all calls for cases who are as "incident" as possible. On the other hand, some cases with an excellent memory will experience a sense of shame. Afraid as they are to be laughed at or to be addressed admonitorily, these cases may conceal their risky behavior to some extent. For instance, in a study to see whether alcohol consumption is associated with a higher ski injury risk, people with a ski injury

may recall their drinking habits very well, especially if they drank too much the night before their accident. Some of them, however, may report a lower alcohol consumption because of a sense of shame. In general, people tend to fill in postal questionnaires more honestly compared with answering questions in a personal interview.

In a case-control study, the comparability of risk factor measurements is all-important. If the cases are interviewed in a certain way, the controls should be interviewed in exactly the same way. So every participant receives a postal questionnaire, for example, with a standardized introduction and the same questionnaire as to length, layout, wording, and so forth. This, however, does not prevent incomparability completely. Suppose that the cases try very hard to recall everything the investigator wants to know and that they are willing to confess every "sin." Now, comparability of measurements is only attained when the controls can be motivated to do exactly the same. It is doubtful whether a random sample of the source population can be motivated as strongly as severely injured persons.

It may, therefore, be preferable to have a nonrandom, but deliberately selected control group. If the cases recall their ski holiday with mixed feelings, this calls for a control group with similar mixed feelings about their ski holiday. Following this idea of comparability, the controls might be selected among the skiers from the same source population who have been the victims of loss or theft. Of course, the assumption is made now that these victims as a group have a risk factor profile for ski injuries that is not different from that of the source population and therefore is "representative" in that sense. Choosing this control group also has the advantage that both cases and controls can be interviewed on a comparable basis about their risk factor profile immediately before their "accident," and that the nonresponse is equally low in both groups.

There exists a tendency among investigators to pay much attention to the measurement of risk factors of interest and little or no attention to the measurement of external risk factors. This is a mistake. In a nonrandomized study the profile of these risk factors may be unequal for persons with and without the risk factors under study and, thus, may confound the association. Therefore, all external risk factors must be measured as validly and precisely as possible, and adjusted for properly to ensure comparability of baseline prognosis and of external circumstances. Everything that has been stressed about the comparability of risk factor measurements can be repeated for putative "confounders." There are many external risk factors in a study of the ski injury etiology, and every one of these deserves careful attention in the data collection and data analysis. The number and range of the external risk factors may, however, be restricted by deliberate selection. For example, in a case-control study concerning the effect of night's rest on ski injury risk, the decision may be taken to make the analysis less complicated by restricting the source population experience to good snow conditions. A disadvantage of this approach is that it is not possible anymore to study the effect of too little sleep on the ski injury risk conditional on certain snow conditions (so-called effect modification). It is thinkable that skiers who did not sleep too well only run an increased risk when the snow condition makes skiing more difficult.

A special case of deliberate selection is matching. Suppose that it is known beforehand that a group of cases will contain twice as many females as males. (Of course, the source population may have equal numbers of both, if the female/male risk ratio is two.) For efficiency reasons it may be a good idea to select twice as many female controls as well. A matched design, however, calls for a matched data analysis, but that is not a big problem with present methods of case-control data analysis. When cases and controls are matched on some risk factor, it is still possible to study effect modification by this risk factor as well. However, the contribution of the matched risk factor to the injury risk cannot be studied anymore. Because of this, only matching on age and gender may be considered and nothing more.

The Achilles' heel of "retrospective" case-control studies is the comparability of measurements of the risk factors of interest, effect modifiers included, and of all external risk factors that may confound any association of interest. Whenever possible, independent measurements must be taken as well. For instance, it would be of interest to do extra interviews with persons

related to the cases and controls. Of course, these interviews should concern the risk factor profile of the cases and controls themselves in an attempt to study whether any association found is changed dramatically using these possibly less precise, but more comparable data. If independent records exist about certain risk factors, an attempt may even be made to go through these records. Case-controlling seems easy enough, but the more you think of it, the more its execution appears to be as difficult as an experiment.

Acknowledgments

Many of the ideas in this paper were developed by Olli S. Miettinen [1] (McGill University, Montreal) and by Alvan R. Feinstein [2] (Yale University, New Haven). For those who want to start a nonexperimental epidemiologic study, we suggest the reading of Jennifer L. Kelsey's book [3]. This book also gives an introduction to data analysis. More advanced statistical methods can be found in the textbook of David G. Kleinbaum [4].

References

- [1] Miettinen, O. S., *Theoretical Epidemiology: Principles of Occurrence Research in Medicine*, John Wiley & Sons, New York, 1985.
- [2] Feinstein, A. R., *Clinical Epidemiology: The Architecture of Clinical Research*, W. B. Saunders, Philadelphia, 1985.
- [3] Kelsey, J. L., Thompson, W. D., and Evans, A. S., *Methods in Observational Epidemiology*, Oxford University Press, New York, 1986.
- [4] Kleinbaum, D. G., Kupper, L. L., and Morgenstern, H., *Epidemiologic Research: Principles and Quantitative Methods*, Lifetime Learning Publications, Belmont, CA, 1982.