

A tool for assessing the usefulness of prevalence studies done for surveillance purposes: the example of hypertension

Luis Carlos Silva,¹ Pedro Ordúñez,² María Paz Rodríguez,²
and Sylvia Robles²

ABSTRACT

Every year dozens of cross-sectional studies are carried out that estimate the prevalence of risk factors for chronic noncommunicable diseases. Given that, there is potentially a large amount of information that could be extremely useful for risk factor surveillance. However, there are good reasons to question the methodological rigor and the reliability of the results coming from many of these studies. The potential benefits of the data are curtailed by the studies' shortcomings, in part often because there is no clear and explicit methodological information providing the details needed to assess the procedures that were actually used, as well as a failure to apply a uniform methodology that would allow comparisons over time or across studies.

This paper is intended to be both a didactic and practical tool. It describes the construction of an instrument for assessing the usefulness of prevalence studies done within the context of surveillance activities, using the example of hypertension. The paper discusses and illustrates the most common pitfalls found with prevalence studies, and it also offers methodological standards that can guide future prevalence studies as well as stimulate future research efforts in this area.

Key words

Surveillance, prevalence, hypertension, risk factors, research methodology.

Over the past several years, hundreds of cross-sectional studies have been carried out to estimate the prevalence of various risk factors for chronic

noncommunicable diseases. Despite the fact that this task is rather simple in conceptual terms, there are a wide array of methodological approaches. There is also great variability in the information that authors report when they describe the results of their research efforts. Further, there can be great variability in such areas as sampling design, age groups comprised, geographical scope of the study (national, regional, selected sites, etc.), diagnostic criteria and rigor in gathering primary data, type of data collected, and ways of arriving at estimates. Such

a wide variety of situations makes one seriously question the value of the data being reported, and the variety also greatly undermines the possibility of successfully comparing study results (1). There is a need to promote a methodological approach that standardizes some of the steps involved in such efforts, both in terms of methodology and ways of reporting information.

If a tool were available to help things move in that direction, researchers would have guidelines pointing to the need to take into account certain basic methodological steps that often get

¹ Instituto Superior de Ciencias Médicas de La Habana, Vicerrectoría de Investigaciones, La Habana, Cuba. Send correspondence to: Luis Carlos Silva, Instituto Superior de Ciencias Médicas de La Habana, Vicerrectoría de Investigaciones, Calle G y 25, 6^o piso, Plaza, La Habana, Cuba; e-mail: lcsilva@infomed.sld.cu

² Pan American Health Organization, Division of Disease Prevention and Control, Program on Non-Communicable Diseases, Washington, D.C., United States of America.

overlooked. In addition, persons interested in summarizing, comparing, interpreting, and assessing the studies performed could do so more productively, while paying special attention to what is needed for disease surveillance.

The Centers for Disease Control and Prevention of the United States of America has defined a health surveillance system in this way (2): "... the systematic collection, analysis, and interpretation of health data that are essential for planning, implementing, and evaluating public health actions, in close connection with the timely dissemination of such data to those who need them. The final link in the surveillance chain is the application of such data in the area of disease prevention and control."

It should be noted that this definition begins by emphasizing the *systematic* nature of the data collection, which necessarily involves looking at reality from a temporal perspective. On the other hand, the aim is not to carry out a number of disjointed efforts, or to simply "take a look" at what is happening, but rather to gain an increased understanding ("analysis and interpretation"), in order to take action based on the knowledge provided to the appropriate individuals ("timely dissemination").

In the case of chronic noncommunicable diseases (NCDs), one aims to learn how factors determining disease processes with a long latency period and that are strongly dependent on human behavior are expressed, given the fact that NCDs differ markedly from infectious diseases in the way they change over time. Infectious-disease surveillance requires that attention be focused on measuring incidence and not prevalence, and thus it relies on the continuous observation of reality. But in the case of NCDs, their risk factors, and their determinants, surveillance requires looking at successive cross-sections at intervals that can change depending on the disease and the factors that are under surveillance.

It must be acknowledged, however, that prevalence studies are of interest not only as part of that process of systematically assessing the reality under surveillance, but because such studies

are relevant in and of themselves. In the first place, prevalence studies provide information that, in spite of being static, involves past data and points toward areas that may require further attention. In the second place, prevalence studies may be the source of baseline information on which to base future assessments of changing patterns by means of measurements performed over time. Thirdly, they are useful for quantitatively and qualitatively assessing the changes that take place, a feature that makes them potential instruments for evaluation purposes.

Obviously, if a study is to be considered "useful" within the context of disease surveillance, it must satisfy certain minimum quality requirements. Furthermore, even among studies that satisfy such requirements, there are subtle variations in the rigor with which they're carried out. Such variations will determine, to a greater or lesser degree, the quality of the data and also the amount of trust that can be placed in the results.

In order to obtain such a measure of "usefulness," an instrument is needed for collecting the basic elements on which to base the measurement, and the points being considered in the analysis should stem from an explicit rationale. Note that this is not a checklist for evaluating generic aspects, such as structure or the quality of the references, which are typically included in guidelines such as those of the so-called Vancouver group, named after a meeting held in that Canadian city by the International Committee of Medical Journal Editors (3). Instead, the instrument is a guide for assessing the scope and reliability of the *contents*. Such an instrument can also be used as a guide for planning and carrying out future studies.

This paper was commissioned by the Program on Non-Communicable Diseases of the Division of Disease Prevention and Control of the Pan American Health Organization (PAHO). It was intended as a response to the technical needs that PAHO Member States had in the area of surveillance. The paper also follows from a consultation process initiated by that PAHO Program in 1996 in order to define the main components

that should be included in a surveillance system for the Region of the Americas (4), one critical element of which was to make the available information usable. In this version of the paper, we have focused on one of the most important risk factors, which is in itself a disease: hypertension.

The construct leading to our proposed instrument stems from an effort to achieve synthesis while aiming for the greatest simplicity, so that the process can concentrate on the essential aspects of assessment.

THE ASSESSMENT MODEL

Our instrument for assessing a scientific report or article contains 19 questions covering six technical aspects:

- declared objectives
- the population under study
- sampling design
- methods for gathering information
- the processing of the information
- communicating the results

The instrument's 19 questions should be answered in light of what has been specifically stated in the report or article. We must emphasize that providing all relevant information is a part of the technical discipline of reporting on research, and making certain this is done is part of the responsibility of editors. Not reporting on something is as good as not having done it in the first place.

Our evaluation strategy is based on the notion that a paper can cross a particular minimal threshold or fall short of it in terms of its usefulness for surveillance purposes. Four conditions, which are assessed through the instrument's first four questions, must be met by a report or article if it is to reach that threshold: 1) it must be a population-based study, 2) the sampling design must be described, 3) the sampling design must be probabilistic, and 4) estimates must be broken down by sex and well-defined age groups. Papers that first meet these four conditions are then assessed using the instrument's remaining 15 questions, and assigned a point score.

Our evaluation model works with the questionnaire shown in Figure 1.

As indicated in the figure, the instrument's 19 questions are divided into three groups: 4 "basic" questions, 11 "assessment" questions, and 4 "risk-factor-specific" questions. The first 4, "basic" questions, A through D, should be answered affirmatively in order for the paper to reach the minimum quality threshold. Some of the remaining 15 questions allow for three possible answers: YES, which indicates that the task was satisfactorily performed; NO, which indicates that the task was not performed, was not communicated, or was not performed satisfactorily; or PART, which indicates that the task was only partially performed. The third option, PART,

makes no sense when applied to questions 3, 4, 5, 12, 13, 14, and 15, so the only possible answer to these 7 questions is either "YES" or "NO." The points given for the respective answers to the scored questions are also shown in Figure 1.

The scoring weights assigned to the questions are based on a consensus reached by the authors of this piece. We gave high weights to efforts to obtain high-quality primary data and to work meeting high standards in the use of sampling techniques.

The maximum total score that can be obtained is 100, when all 15 of the scored questions are answered "YES." The minimum score is 0, when the answer to all those questions is "NO." The paper receives a final grade as follows:

- not useful: It does not meet the minimum threshold of satisfying the 4 basic questions.
- minimally useful: It reaches the threshold but receives fewer than 35 points on the 15 scored questions.
- useful: It receives 35–69 points.
- very useful: It gets 70 points or more.

It should be noted that a paper can be methodologically sound or even provide answers to interesting and relevant research questions yet still be graded as "minimally useful," since the assessment involves not only the quality of the scientific work, but also its usefulness from the standpoint of surveillance.

The authors of this piece as well as several external specialists in research

FIGURE 1. Form used to assess the usefulness of prevalence studies done for surveillance purposes: hypertension example

| Basic questions | | YES | NO | |
|---|--|------|-----|------|
| A | Is the problem being studied in a general population (rather than one that is captive or institutionalized)? | | | |
| B | Is the study's sampling design fully described? | | | |
| C | Was a probabilistic sample used? | | | |
| D | Are prevalences given by age groups and sex? | | | |
| NOTE: Only studies for which a YES answer has been given to all four of the Basic Questions should be scored with the following 15 questions. | | | | |
| Assessment questions | | YES | NO | PART |
| 1 | Is the problem under study described in both quantitative and qualitative terms? | (5) | (0) | (3) |
| 2 | Were standardized techniques used to measure arterial blood pressure? | (10) | (0) | (5) |
| 3 | Were universally accepted cut-offs used in diagnosing the ailment? | (5) | (0) | |
| 4 | Did the data collectors receive training? | (5) | (0) | |
| 5 | Were certified instruments and observers used? | (5) | (0) | |
| 6 | Was there quality control of the primary data? | (5) | (0) | (3) |
| 7 | Were estimates calculated according to the sampling design? | (10) | (0) | (5) |
| 8 | Were estimates made by place of residence, occupation, or educational level? | (5) | (0) | (3) |
| 9 | Are the errors of the estimates reported according to the sampling design? | (20) | (0) | (10) |
| 10 | Are extrapolations explained or discussed? | (5) | (0) | (3) |
| 11 | Are any qualitative judgments made that can serve as the basis for action? | (5) | (0) | (3) |
| Risk-factor-specific questions | | YES | NO | PART |
| 12 | In addition to prevalence, was mean blood pressure estimated? | (5) | (0) | |
| 13 | Is the percentage of hypertensive individuals who know of their condition indicated? | (5) | (0) | |
| 14 | Is the percentage of hypertensive individuals under treatment indicated? | (5) | (0) | |
| 15 | Is the percentage of hypertensive individuals whose disease is under control indicated? | (5) | (0) | |

methodology whom we consulted made numerous revisions to the procedures described in this paper. These revisions led to successive corrections after the procedure was applied to several dozen studies (that analysis will be presented in a separate, forthcoming paper (5)). Other revisions were made after the procedure was evaluated by experts during what was essentially an assessment of criterion validity and face validity of this instrument. It is possible, furthermore, that, using an approach that one researcher has recommended (6), the construct validity of our instrument will be assessed. That will be done by considering, as an independent construct, the impact factor of the journals where the studies that we analyzed (5) had been published.

RATIONALE FOR AND COMMENTS ON THE QUESTIONS

This section lists the 19 questions asked. In each case, there is an explanation of the question's theoretical framework, that is, the context that makes the question relevant or necessary. There are also comments on the meaning of each of the questions.

Basic questions

A. Is the problem being studied in a general population (rather than one that is captive or institutionalized)? Knowing the prevalence of the risk factor in the whole, "general" population of a given country, geographical jurisdiction, etc. is essential, since the actions of health service providers should be population-oriented. While they may be useful for other purposes, studies that attempt to describe the prevalence of hypertension among such specific groups as patients who attend a particular hospital or who work in a particular field of employment are of little or no usefulness in terms of population surveillance. This should not be confused with the practice—which is essentially legitimate—

of using as a sample the subjects being cared for by a sample of physicians or health care facilities. This is done, for example, in so-called "physician-based surveillance."

B. Is the study's sampling design fully described? In prevalence studies the quality of the sample plays a decisive role. In such cases, the purpose of the study is to estimate the prevalence of a given risk factor in a population. This is different, in theory and practice, from studies that are designed to answer questions explaining or identifying the association between a risk factor and a disease condition. Such association studies require comparisons between or among groups, and the importance is placed on the comparability of the selected groups. With prevalence studies, however, the emphasis is on the representativeness of the population under study—and thus on the sample.

With all prevalence studies, it is crucial that the sampling procedures employed be sound. This also applies to the infrequent situation of studies that have presumably included "the entire population," since this population under study is, strictly speaking, a sample that seeks to be temporally representative. This is the everyday reality. For example, the steps that are taken in response to a patient satisfaction study performed among hospitalized patients will apply to a population that is essentially different from the one that was studied (i.e., the surveyed population vs. a population of persons hospitalized at a later point in time). It is likely that people who are currently hospitalized do not differ essentially from those who were studied at an earlier date. Therefore, the real inference being made will be legitimate, even if the formal rule of extrapolating the results only to the sampled population is not followed.

C. Was a probabilistic sample used? It is generally believed that a sample has been drawn with statistical rigor only if it meets the following two conditions: 1) the procedure assigns each element in the population a previously

known probability of being included in the sample, and 2) such a probability is not zero for any of the elements. In cases satisfying these two conditions, a so-called probabilistic sampling method has been followed.

For example, if one wishes to study the prevalence of hypertension in a city having 50 census districts, and 20 of the districts are randomly chosen, followed by taking 1 out of every 4 square blocks within each district, and finally 1 out of 20 adults living in those 20 square blocks that were chosen, a probabilistic sample will have been drawn, in which everyone in the city has 1 probability in 200 of being selected for the study. This figure is derived by multiplying the probability of being selected at each of the three sampling stages: $20/50 \times 1/4 \times 1/20 = 1/200$.

This is a requirement of the highest importance, since probabilistic procedures satisfy the intuitive requirement of eliminating, or at least minimizing, the burden of the subjectivity that might influence the selection of the elements to be examined and therefore the resulting conclusions. Chance ensures against systematic distortions, whether they be deliberate or not, and its role in that respect is generally irreplaceable. Secondly—and this is absolutely essential—only a probabilistic method allows for measuring the degree of precision with which estimates are made.

One way in which the probabilistic nature of the sample is quite frequently destroyed is to draw the sample from patients who seek care in a particular facility or from a particular physician. Another way of destroying the probabilistic nature is when a self-selection bias is introduced, such as when a call is made to the general public and the people who respond are those who want to or who find it convenient to do so. Also destructive to the probabilistic nature is setting a certain number of subjects and then just choosing them from a particular place or facility until that number is reached.

The representative nature of the sample is of crucial importance in prevalence studies, and the biases introduced through flaws like the ones

just described cannot be corrected during the analytical phase of the study, given that the mistakes leading to such biases were made during the study design phase.

D. Are prevalences given by age groups and sex? It is not a requirement that the study include all age groups within the population; the study can focus on a particular age bracket such as persons over 30, or the elderly. It is recommended, however, that ages be divided into 5- or 10-year intervals after the age of 15. If the study involves a particular age range, that range should be divided into 5- or 10-year groupings. This is extremely important because it makes it possible to compute age- and sex-adjusted rates that can be used later to make comparisons. Furthermore, knowing the prevalences for different age groups and for each of the sexes separately is of interest in terms of refining surveillance. Both physiologically and behaviorally, the reality of one group can differ substantially from that of other groups. In fact, this does occur in the case of hypertension. In general, it would be advantageous to work with at least the following seven age groups: 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, and 75 years or older.

Assessment questions

1. Is the problem under study described in both quantitative and qualitative terms? Obviously, any method used to resolve a research problem can only be decided on once the problem has been clearly formulated. This question is intended to reveal whether this requirement has been met.

Unfortunately, many so-called "descriptive studies" are little more than an act of observation, rather than a tool for evaluating reality based on a description of that reality's characteristics. A study that is genuinely descriptive should go beyond a neutral assessment of reality. Studies that describe a situation are not genuine scientific research when they are indistinguishable from administrative reports

that offer little more than facts that could be generated by a computer and that fail precisely at the point where our knowledge and ability to interpret information is most indispensable. Genuine research takes place only when it stems from a critical approach that allows numerical results to be translated into value judgments. It is imperative that the results be precise and that it be possible to extrapolate them. However, translating results into conceptual terms and converting them into value judgments requires that researchers have a scientific approach, which begins with a clear formulation of a problem that merits research. This approach goes beyond just sampling methods or other statistical tools.

On the other hand, the objectives of many studies are quite nebulous, and the body of the papers is filled with statistical details that are often unnecessary or that bear no relation to the studies' true objectives. An example would be the inclusion of hypothesis tests among groups, such as by sex or occupation, that contribute little to the analysis of the problem under study.

It is often wrongly believed that for a descriptive study to be truly "scientific," it must assess causal factors. Other authors think that, at least, they have to assess to what extent putative risk factors should be corroborated as such. Causal analyses are thus provided that fruitlessly distract from the true objectives of the study. One of the problems linked to these practices lies in the fact that the methodological requirements of explanatory studies and of descriptive studies are virtually incompatible. A study that seeks to make causal associations (instead of assessments), such as a case-control study, and that is well designed, can be performed with a small sample, perhaps as few as several hundred subjects. This would be unthinkable in the case of a descriptive study, since such a small sample size would be clearly inadequate.

In summary, the goal is for the researcher to express his or her purpose within a conceptual framework and to state objectives that go beyond a mere knowledge of a set of numbers.

2. Were standardized techniques used to measure arterial blood pressure?

The use of standardized techniques for measuring any risk factor is a must. In the case of arterial blood pressure, these techniques were recently described in detail (7). Following such standards is especially important in order to be absolutely certain that the measurements are accurate. Further, using standardized techniques is critical for results to be considered valid and particularly for genuinely fruitful comparisons with other studies.

3. Were universally accepted cut-offs used in diagnosing the ailment?

Over the years, a number of different criteria have been used to determine if an individual is hypertensive. However, the criterion that at present is considered almost universally valid is a reading of 140 mm Hg for the systolic pressure and 90 mm Hg for the diastolic pressure (8, 9). In practice, any subject who is being treated for hypertension, either with drugs or another approach, or who has a reading that is above his or her usual pressure, should be considered hypertensive. One should avoid the practice, which is seen in some reports, of reporting only one of the two measures, either diastolic or systolic.

4. Did the data collectors receive training?

Questionable blood pressure readings can often be attributed to measurements being taken without using standard methods and instruments. Therefore, training of nurses, technicians, or other personnel is of crucial importance. Training can be done through classes or by using special tapes or films.

5. Were certified instruments and observers used?

Certified instruments and observers are another basic methodological ingredient in obtaining valid measurement data. Since 1927, when the first sphygmomanometry standard was published, there have been many reports about available devices. For example, the Association for the Advancement of Medical Instrumentation (AAMI) has approved a sphygmomanometer equipment stan-

standard that covers such elements as safety, accuracy, and instructions; the AAMI also has various certification programs for equipment specialists and technicians.

6. Was there quality control of the primary data?

As with training and certification, quality control is crucially important in guaranteeing the quality of the primary data. Quality control is also vital to reducing intraobserver and interobserver bias. The two fundamental aspects of quality control are: 1) performing a repeat survey (ideally, another observer should perform it on a subsample) and 2) making certain that when recording a measurement, such as weight or blood pressure, there is no bias in favor of particular final digits, especially zero. If, for instance, the frequency of observations ending in zero is greater than 30%, the data are not considered to be of sufficient quality to be valid. A properly performed study should include the detailed results of this analysis. (For further details on the subject of quality control in epidemiologic studies, see Chapter 11 in Nieto and Szklo (10).)

7. Were estimates calculated according to the sampling design?

Quite often, study designs are not equally probabilistic; that is, every subject in the population does not have an equal chance of being included in the general sample. In such cases, it is necessary to resort to weighing in order to adjust the estimates. This is done by correcting estimates for which not all subjects had the same chance of being in the sample, as tends to be the case when a more or less uniform sample size is used despite the fact that the age pyramid is not uniform for all ages. Attributing the unweighted sample prevalence to the entire population is an error in such cases.

For example, let's suppose that in a study of elderly persons, there are only two age groups, 65–79 years and 80–95 years, and that 200 individuals are selected from each group. Let us also suppose that 80 are found to be hypertensive in the first group, and 100 in the second group (rates of 40%

and 50%, respectively). Without weighing, the estimate would be a rate of 180/400, or 45%. However, if the two groups are not the same size in the general population—as would be expected—it becomes necessary to use weighing. If, for instance, in the population there are 15 000 subjects in the first age group and only 5 000 in the second, an accurate estimate would involve applying weights (0.75 and 0.25) to the prevalences of both groups, that is: $(0.75 \times 40\%) + (0.25 \times 50\%) = 42.5\%$.

8. Were estimates made by place of residence, occupation, or educational level?

For surveillance purposes, making estimates by place of residence, occupation, or educational level is desirable since it provides the basis for planning actions. Nevertheless, satisfying this requirement is not an absolute condition for a study to be valid. For purposes of prevention, having such data could contribute to a more refined intervention, to more accurate planning and impact evaluation, and to better-oriented educational efforts (11, 12).

9. Are the errors of the estimates reported according to the sampling design?

Precision can be indicated either by giving confidence intervals or by separately showing the maximum error that could be affecting the estimates. This step should not be skipped, since it allows one to judge to what degree he or she can be certain to really know the parameter being estimated. The International Committee of Medical Journal Editors is categorical on this point in its universally accepted document known as the *Uniform Requirements for Manuscripts Submitted to Biomedical Journals*, in the section devoted to presenting statistical data (3). That section asks authors to: "... quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals). Avoid relying solely on statistical hypothesis testing, such as the use of *P* values, which fails to convey important quantitative information."

For example, let's suppose that two countries are found to have a 25%

prevalence rate of hypertension. Let's also assume that, if we take into account the sample size used in each study—which has a decisive influence on the magnitude of the error—as well as the sampling design employed—which is also a crucial determinant of the way the errors are calculated—the error estimates for the two countries are 2% and 23%, respectively. In the first instance, one can be quite certain that the rate lies somewhere between 23% and 27%. With the second country, however, one can only be certain that the rate falls between 2% and 48%. The difference between what is known about each of the two countries is immense. The first study can be considered very informative, whereas the second one could be virtually useless, since it was almost certain beforehand that the prevalence, despite being unknown, was not less than 2% or greater than 48%.

Care should be taken with sometimes complex probabilistic studies that are performed and where later the sampling errors are not estimated, or only vague statements are made about them. That is, the design phase is rigorously conducted, with a specialist consulted in an effort to ensure that the sample is probabilistic, but in the analytical phase the calculation of errors is omitted. This undermines the initial efforts at performing serious research. It is likely that those who work in this fashion mistakenly believe that if they develop a formal design in which chance plays a role, they are making the sample representative. Instead, the only thing they are doing is ensuring the objectivity of the person designing the sample and the possibility of estimating the degree to which the estimates are accurate.

Very frequently, errors or confidence intervals are computed using the formula normally applied for simple random sampling (SRS). But in nearly all cases, the sampling design is not a simple one, but rather multistage or cluster sampling, which is used at least 90 times out of 100. The sampling error usually made when estimating a proportion—particularly a prevalence rate—is thus greater than the one that is being taken as valid when SRS formulas are used.

The assessment tool that we describe in this paper, it is worth noting, does not ask about the sample size employed. The reasons for this have to do with something that is not adequately understood: *no matter what the sample size, sampling errors can always be calculated* a posteriori. The formulas for computing errors explicitly take into account the sample sizes used. The structure of these formulas is such that the researcher will be unable to draw firm conclusions from the data if the sample size is inadequate; therefore, sufficiently large samples should be taken. At the same time, the formulas' structure ensures caution when the sample size is too small.

It is quite often better to rely on common sense and to take into account the sample sizes used in similar studies, rather than to base decisions on the presumed objectivity of the formulas. On the other hand, it is senseless to decide on the sample size without taking efficiency into account. That is, if resources were unlimited, many times there would not even be a need to use a sample, and the entire population could come under study. Considerations such as time, availability of personnel, and budget should, and in fact always do, play a decisive role in decision-making, even though the influence of such limiting factors is implicit or is masked.

Most books overlook these facts. Fortunately, there are exceptions; one of them is the classic text by Kenneth Rothman (13), published in 1986, which clearly acknowledges the impossibility of a theoretical solution, stating: "In short, the problem of determining the best sample size is not a technical one; it cannot be resolved through computations, but instead must be approached with judgment, experience, and intuition."

Notwithstanding, what has become standard belief among students and researchers is that for every problem there is a single number that can be "discovered" by specialists, aside from personal considerations, and that this one number can be found through technical means by a few "chosen" ones who are capable of understand-

ing complex formulas. Many experts in such methodology can and do put modest researchers in a bind by asking them to justify their sample size in light of what they have seen in the literature, or in light of the resources they have available. In such cases, the researchers may resort to using formulas that, as has been explained and illustrated in detail elsewhere (14), are far more subjective than they would be for an individual who relies on his or her own common sense to choose the sample size.

In the case of hypertension, prevalence rates typically range from 15% or 20% up to 40%. Given that, studies in this field should have no fewer than 200 subjects in each age and sex category. Sample sizes can be more or less the same in each group, and this is advisable so that similar precision can be attained in all estimates. However, this strategy will almost certainly be incompatible with equal probabilities. That is, there are fewer individuals in the older age groups, so each one of those persons will have a greater chance of being included in the sample than will younger individuals. In such cases, weighing must be used to estimate general prevalence rates, whether or not those rates have been adjusted in accordance with a reference population for comparison purposes.

10. Are extrapolations explained or discussed? The legitimacy of extrapolations is one of the thorniest topics in the field of prevalence studies. Students of sampling design are quite familiar with the rule that states that "inferences made on the basis of a sample should be applied only to the population from which the sample was drawn." From a strict point of view, this is unquestionably true. However, firm adherence to this rule would have such paralyzing effects that, in practice, it tends to be overlooked.

It must be acknowledged that frequently, for one reason or another, with the particular method used to select subjects, not all individuals in the population have a chance to be included in the sample, yet the infer-

ences objectively drawn from the results apply to the entire population and not only to the portion from which the sample was drawn. The degree to which such a "transgression" of the sampling rule can be "pardoned" is not, generally speaking, a statistical matter. Rather, it is inherent to the problem being studied, and it depends on the researchers' judgment, based on their common sense and their understanding of the problem, which must serve as the basis for the final word.

Let us assume that, in a study like the ones we have described, subjects are selected from the three largest cities within a country, and that on that basis it is estimated that 20% of those cities' residents have abnormally high blood pressure levels, with a 95% confidence interval of 16% to 24%. What does this mean? The classic answer is, "We can be reasonably certain that the percentage of hypertensive individuals in those three cities when the survey was conducted was between 16% and 24%"—and no more and no less than that. What must be emphasized is that, formally speaking, the inference is limited to the prevalence *in those three cities at that point in time*. However, no one would perform the study if the results were to be no more than a historical anecdote. If such results are published, it is frequently because there is a tacit belief that they are indicative of something that occurs beyond the three cities involved and beyond the time the survey is conducted. What happens is that the decision regarding the geographical and temporal scope of the extrapolation is often left in limbo, which relieves the researchers of having to openly commit themselves. However, it does not resolve the fact that the inference being objectively drawn transcends the sample on which it is based.

Obviously, the reality unveiled by the study will not be pertinent 20 years later, nor will it be possible to extrapolate it to the rural population. However, it is likely that the findings do reflect what occurs in other cities, and that the findings are essentially valid as long as the determining factors do not change.

Now let us assume that we wish to estimate the prevalence of hypertension in a given city, but that, for practical reasons, the sample will be drawn from an incomplete list containing only 90% of the current population. Let us imagine, for instance, that the sample includes only those individuals who have been seen at least once in a health facility, given that the sampling frame will be the records kept in such facilities. In such cases, the resulting sample will only include subjects who have had that experience.

The key question, which is clearly not statistical in nature, would be as follows: Are there reasons for suspecting that having been seen at a health facility at some point is somehow related, either directly or indirectly, with having a given risk factor? If the answer is yes, there is no "pardon" possible. But if, despite any theoretical speculations on the part of public health experts and physicians, there is no such link, either direct or indirect, between the two conditions, the natural inclination would be to give "methodological absolutism" and allow extrapolation. Let us recall that, in any event, any knowledge acquired through a sample is temporary and subject to improvement. On the other hand, if we are flexible in making temporal extrapolations, why can we not be flexible when making spatial extrapolations? Ultimately, it is a matter of being flexible within a rigorous framework, which is far better than being rigid while adhering to a conceptually nebulous foundation, as happens so often and in so many situations.

In summary, the proper degree of extrapolation is usually a problem equally involving formal sampling formulas and common sense. What is not permissible is a failure to specifically comment on the particular population to which results can be extrapolated.

11. Are any qualitative judgments made that can serve as the basis for action? Genuine descriptive research is typically an effort to assess reality, particularly when performed within the context of surveillance. However,

it is essential not to engage in conducting observational exercises based on an uncritical, superficial use of descriptive statistics, which renders any study meaningless.

Understanding the research process as a complex and integrated activity should lead us to avoid formal categorizations, such as separating descriptive from explanatory research, even though making that distinction may be useful in certain circumstances, such as for teaching purposes. It would not be possible to conduct genuinely fruitful analytical research without the knowledge on which to base the hypotheses being tested; in general, such an empirical foundation either stems from descriptive studies or is consistent with them. Thus, whereas descriptive studies are not explanatory procedures in themselves, they are a form of biomedical research that is not only legitimate, but necessary beyond doubt in order to design practical tasks and proper surveillance in the future. It is so much so that, according to Greenland (15), "the first duty of the epidemiologist is descriptive." Conclusions derived from prevalence studies should ideally go beyond quantitative measures and produce useful judgments that the various health actors can apply in taking corrective measures.

Risk-factor-specific questions

12. In addition to prevalence, was mean blood pressure estimated? There is a broad recognition of the importance of estimating prevalence based on a dichotomization, that is, whether a person does or does not have the condition under study. And while that dichotomization may be useful for guiding action or for legal purposes, it does not communicate all the relevant information about the disease. Therefore, it is useful to present ordinal or continuous values when possible, such as means and dispersion. In the case of hypertension, that would be both systolic and diastolic blood pressure, measured in mm Hg. Population means are important for monitoring the changes in

the population at large, thus its public health importance. The classification of having or not having the condition has mostly clinical significance, since it is the basis for treatment decisions.

Clearly, arterial blood pressure as such is highly important, especially among persons who are ill. Therefore, some intervention programs include among their goals the reduction of blood pressure levels in and of themselves, independent of whether or not the programs can reduce the prevalence of the disease.

13. Is the percentage of hypertensive individuals who know of their condition indicated?; 14. Is the percentage of hypertensive individuals under treatment indicated?; 15. Is the percentage of hypertensive individuals whose disease is under control indicated? These three questions respond to the need to have data that are of crucial importance to taking action in areas that can be directly changed through health interventions and that allow more accurate estimates of the prevalence of high blood pressure. This applies to changes that can be developed by health service providers as well as to ones in which the patient and the community participate actively and consciously.

All three questions are results indicators, which are essential to any secondary-prevention program. In and of themselves, such data are crucial for surveillance and are useful for designing interventions. Needless to say, a population in which 25% of those who are ill know of their condition is potentially much less protected than one in which 50% of those with the illness are aware of their problem.

FINAL CONSIDERATIONS

Descriptive studies have been given the "bad reputation" of being useless for studying causality, and of thus being useless in general. Nothing could be more wrong. It is imperative not only to reestablish the legitimacy of descriptive studies as an intrinsic

cally valuable public health instrument, but also to underscore the urgent need to revitalize their presence in current research.

This is particularly true for the type of research being performed to actively transform health conditions within a community, which is the ultimate aim of research.

Keeping in mind the value that prevalence studies have as tools within health surveillance systems, there is an

obvious need to increase their methodological rigor and thus allow us to place more faith in their results.

As indicated in the subtitle of this paper, we chose as an example studies of the prevalence of hypertension. While we believe that the overall philosophy of this paper is applicable to any risk factor, such as smoking, hypercholesterolemia, or obesity, there is no doubt that in other cases it will be necessary to make proper adjustments

as required by the specific risk factor under study.

We hope that the sizable limitations in terms of clear and explicit methodological information and of homogeneity that are present in current research and scientific literature can be overcome, or at least reduced, by using a standard tool such as the one described in this paper. We recommend its use from both a practical and a didactic perspective.

REFERENCES

1. Yunis C, Krob HA. Status of health and prevalence of hypertension in Brazil. *Ethn Dis* 1998; 8(3):406-412.
2. United States of America, Centers for Disease Control and Prevention. CDC surveillance update. Atlanta, Georgia: CDC; 1988.
3. Comité Internacional de Editores de Revistas Médicas. Requisitos uniformes para los manuscritos enviados a revistas biomédicas. *Medicina Clínica (Barc)* 1997;109:756-763.
4. Pan American Health Organization, Program on Non-Communicable Diseases. Networking for the surveillance of risk factors for non-communicable diseases in Latin America and the Caribbean. Washington, D.C.: PAHO; 1999. (Publication PAHO/HCP/HCN/99.08).
5. Ordúñez P, Silva LC, Rodríguez MP, Robles S, Belis D. Prevalence estimates for hypertension in Latin America and the Caribbean: are they useful for surveillance? *Rev Panam Salud Publica*. Forthcoming 2001.
6. Silva LC. Cultura estadística e investigación científica en el campo de la salud. Una mirada crítica. Madrid: Díaz de Santos; 1997.
7. Beevers G, Lip GYH, O'Brien E. ABC of hypertension: blood pressure measurement. Part II-conventional sphygmomanometry: technique of auscultatory blood pressure measurement. *BMJ* 2001;322(7293):1043-1047.
8. Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. The sixth report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Arch Intern Med* 1997; 157(21):2413-2446.
9. 1999 World Health Organization-International Society of Hypertension guidelines for the Management of Hypertension. *J Hypertens* 1999;17(2):151-183.
10. Nieto J, Szklo M. *Epidemiology: beyond the basics*. Gaithersburg, Maryland, United States: Aspen Publishers; 1999.
11. Diez Roux AV, Merkin SS, Arnett D, Chambless L, Massing M, Nieto FJ, et al. Neighborhood of residence and incidence of coronary heart disease. *N Engl J Med* 2001;345(2): 99-106.
12. Marmot M. Inequalities in health [editorial]. *N Engl J Med* 2001;345(2):134-136.
13. Rothman JK. *Modern epidemiology*. Boston: Little, Brown and Company; 1986.
14. Silva LC. *Diseño razonado de muestras y captación de datos en la investigación sanitaria*. Madrid: Díaz de Santos; 2000.
15. Greenland S. Randomization, statistics, and causal inference. *Epidemiology* 1990;1(6): 421-429.

Manuscript received on 20 August 2001. Accepted for publication on 27 August 2001.

RESUMEN

Metodología para valorar la utilidad de estudios de prevalencia realizados con miras a la vigilancia: el ejemplo de la hipertensión

Puesto que cada año se realizan docenas de estudios transversales que estiman la prevalencia de factores de riesgo de las enfermedades no transmisibles, es posible que exista mucha información de gran utilidad desde el punto de vista de la vigilancia de dichos factores. No obstante, hay fuertes motivos para poner en tela de juicio el rigor metodológico y la fe que se puede depositar en los resultados de muchos de estos estudios. Los posibles beneficios que aportan estos datos se ven limitados por las deficiencias de los estudios, en parte porque no hay información metodológica clara y explícita con los detalles necesarios para evaluar los procedimientos empleados y porque no se aplica una metodología uniforme que permita hacer comparaciones a lo largo del tiempo y entre distintos estudios.

Este trabajo tiene por objetivo servir como herramienta didáctica y práctica. En él se describe la construcción de un instrumento para evaluar la utilidad de los estudios de prevalencia efectuados en el contexto de las actividades de vigilancia, utilizando a manera de ejemplo el caso de la hipertensión. El trabajo también examina e ilustra las deficiencias más comunes de los estudios de prevalencia y ofrece pautas metodológicas que servirán para orientar cualquier estudio futuro y estimular investigaciones adicionales en este campo.