UK. The death rate from carcinoma of the lung in the UK increased progressively, with a 15 - 20-year lag period behind the rising tobacco consumption. Many studies, including nine prospective surveys, have shown that the possibility of death from lung cancer increases as the number of cigarettes smoked increases. The consistency of the epidemiological evidence, together with the dose-response relationship and the fact that numerous carcinogens are known to be present in cigarette smoke, provide the basis for the now unequivocally accepted relationship between smoking and lung cancer.[8] There is also evidence to suggest that the risk of developing lung cancer increases with earlier age of onset of smoking and with the dose of carcinogens present in cigarettes. The early onset of smoking in South African schoolchildren and the high tar and nicotine content of South African cigarettes are thus cause for great concern.[9] The combination of cigarette smoking and failure to prevent death from pulmonary infections, in particular tuberculosis, highlights the deficiencies in the preventive and curative health services in the RSA. Major changes need to be made to our health care system to rectify these deficiencies.[7]

REFERENCES

1. Wyndham CH. Trends in the mortality rates for the ten leading causes of death among white, coloured and Asian children under 5 years of age in the RSA, 1968-1977. *S Afr Med J* 1984; **66:** 719-725.
2. Wyndham CH. Mortality trends in the populations of the RSA from causes commonly observed in developing communities, 1968-1977. *S Afr Med J* 1984; **66:** 945-950.
3. Benatar SR. The changing pattern of respiratory medicine. In: Shapiro HA, ed. *Medicine and Health in Developing Southern Africa: The Fourth Interdisciplinary Symposium of the College of Medicine of South Africa* (Durban, 5-6 July 1984). Cape Town: College of Medicine of SA, 1984: 122-128.
4. Wyndham CH. The loss from premature death of economically active manpower in the various populations in the RSA. *S Afr Med J* 1981; **60:** 411-419, 458-459.
5. Central Statistical Services. *Reports on Deaths of whites, 'coloureds' and Asians* (Reports 07-03-17 - 07-03-21). Pretoria: Government Printer, 1982-1985.
6. Central Statistical Services. *Population Census 1980: Age by Statistical Regions and Districts* (Report 02-80.04). Pretoria: Government Printer, 1985.
7. Benatar SR. Failure of tuberculosis control in South Africa — need for a unitary national health service (Editorial). *S Afr Med J* 1986; **70:** 247-248.
8. Horsefield K. Smoking and the lung. In: Scadding JG, Cummings G, eds. *Scientific Foundations of Respiratory Medicine.* London: Heineman, 1982: 517-527.
9. Prout S, Benatar SR. Smoking in white high-school children in Cape Town. *S Afr Med J* 1983; **63:** 483-486.

Epidemiological research methods

Part VI. Planning a research project

J. L. BOTHA, D. YACH

'There is something extraordinarily satisfying in designing an RCT of "place of therapy", writing the protocol in such a way as to avoid all the ethical pitfalls, persuading all the necessary people to participate, and checking to see that no one cheats.'

Thus wrote Cochrane in 1971.[1] The need for writing a good protocol is still paramount and applies not only to randomised controlled trials (RCTs), but to all research projects. Writing a good protocol reflects good planning, and in this paper we address some vital aspects of planning a study and writing a protocol. The information given here should be augmented by referring to earlier articles in this series,[2-6] as well as guidelines published elsewhere.[7-15]

'Protocol', as used here, refers to a written record of the rationale for the study, and of the activities and procedures planned for use in the study, including a standardised form for recording the observations. In writing a protocol, the researcher has to state his/her ideas explicitly, and has to anticipate and pre-empt problems that may occur in the execution of the study, and/or in the analysis and interpretation of the results. Writing the protocol provides an opportunity for the researcher to sharpen or re-examine his/her own concepts. The written protocol also serves as a means to obtain constructive criticism from peers at a stage when suggested changes can still be accommodated. During the execution of the study the protocol serves as a reference document of decisions made and how and why they were made. This helps to avoid inconsistency in *ad hoc* decision-making, particularly with changes in research staff.

We write here of 'a protocol' as if it is a single document produced once only, and for all time. We do not mean to create such an impression: a protocol is a dynamic document, developing with planning, peer review and pilot studies, and improving all the time under the rigours of the research methods to be used. On the other hand, a draft of the protocol may have to be used for a specific purpose at a given time as if it is a final document, e.g. to meet a deadline for fund applications.

We suggest that when writing or reading a protocol, the writer or reader should be in a position to answer the six questions discussed in detail below.

1. Is the research question relevant?

This question implies knowledge of the content area of the proposed research. Such knowledge should be demonstrated in the protocol by explaining why the research idea is a good one. A literature review should form part of this motivation, to demonstrate that the researcher is aware of previous and current attempts to answer the same question. In surveying published work, both the content and the research methods should be reviewed. The implications and practical relevance of study results should be clearly spelt out in the protocol.

When using a protocol to apply for funds, researchers should be aware of the criteria used to decide on funding,

Institute for Biostatistics of the South African Medical Research Council and Departments of Community Health, University of Stellenbosch, Parowvallei, CP, and University of Cape Town
J. L. BOTHA, M.B. CH.B., M.SC. (CLIN. EPID.)
D. YACH, M.B. CH.B., B.SC. HONS (EPID.), M.P.H.

especially if many funding agencies exist. Cancer researchers should know that a funding agency for cancer research exists,[5] and diabetes researchers should know about funding for diabetes research. In countries such as the RSA, where the bulk of research funding comes from a few agencies, which may have as their source public (taxpayers') money, care should be taken to use clear, content-based and methodological criteria. The other five questions discussed in this paper incorporate clear methodological criteria.

Content-related criteria are difficult to propose where many researchers from different content areas compete for funding from the same source. We wish to suggest some common-sense guidelines often used in deciding on the relevance of epidemiological research. We believe they have wider applicability.

What is the impact of the problem?

The greater the number of people affected, the greater the number of people who can benefit from a solution. Examples are: mortality in black children, and cardiovascular disease in other population groups.

Is the problem amenable to intervention?

Problems for which interventions of proven benefit exist deserve priority, particularly in health care research, or in areas where prevention is possible. Examples are: child mortality, tuberculosis and hypertension.

Will action follow the results?

Problems which health care authorities (official or community-based) are willing and able (financially and in terms of manpower) to tackle should have higher priority than those they are loath or unable to address.

Does the project present a unique opportunity for research?

If the research cannot be done elsewhere, either because the problem is unique to the RSA, or because the technique is only available locally, the project should be given higher priority.

2. Is the project feasible?

The feasibility of a project can be seen to have three components: epidemiological, logistical and ethical.

Epidemiological feasibility is present if three questions can be answered in the affirmative:

Can an answer to the research question be obtained by using the chosen study design? The chosen design should be clearly stated with the reasons for that choice, remembering that descriptive, analytical and intervention studies are done to obtain answers to different types of questions.[2-4] Similarly, the reasons for choosing a particular type of analytical study[5,6] should be stated.

Is the proposed sample size large enough so that the chance of missing a clinically meaningful difference will be small? The chance referred to here is also known as the type II error and is often denoted by beta. Its complement (1-beta) is often referred to as the power of the test in not missing real differences, and should be specified in the protocol. A probability of 0,1 or 10% of missing a real difference is acceptable to most researchers.

Is the proposed sample size so large that the chance of picking up a clinically meaningless difference as statistically significant will be large? The chance referred to here is also known as the type I error and is often denoted by alpha. It is seen in the analysis as the *P* value, and should be specified in the protocol. A probability of 0,05 or 5% of this occurring is acceptable to most researchers.

In practice, these last two questions about sample size can only be answered if the sample size estimation procedure is clearly shown in the protocol.

Logistical feasibility is present if the proposed work can reasonably be expected to be completed by the existing manpower resources (or those budgeted for) in the time allocated. In the protocol the budget should be clearly spelt out, detailing salaries of staff, costs of travelling (including the cost of attending conferences), and amounts set aside for equipment, computing, publication, stationery, and other running expenses (e.g. stamps and phone calls). The scheduling of activities (who does what, when and where), and the duties of each team member should be indicated.

Ethical feasibility is present if the proposed project conforms to the ethical principles generally agreed on, either internationally[16] or locally.[17] In community-based studies it is particularly important to obtain community consent and participation, in the RSA often from formal as well as informal community leaders. When the proposed project is a randomised trial of one kind or another, ethical considerations relating to allocation of patients with informed consent have high priority.[18-20]

3. Will the results be valid?

The validity (or accuracy) of a study's results is defined as their closeness to the true or real situation. Lack of validity is described as a bias, or systematic deviation from the truth. Such a bias can arise from many sources,[21] and the steps which will be taken to avoid applicable biases should be clearly spelt out in the protocol. We will concentrate on biases from two major sources: sampling and measurement.

Sampling bias

Sampling bias, or systematic sampling error, occurs if the selected sample is unrepresentative of the population of interest, which means that sampled individuals differ systematically from those not sampled. For valid inferences to be made about a population effect from an estimate of that effect in a sample, the sample must be representative of that population. The likelihood of obtaining a representative sample is high if some form of random sampling is used.[3] It is therefore important that the researcher specifies clearly in the protocol what target population he/she has in mind, whether a sample will be used and how such a sample will be drawn. In addition, the researcher should indicate how the representativeness of the sample can be validated, e.g. against census information. This is necessary for all types of study design in which samples are drawn.

If the study design is analytical, i.e. two or more groups are being compared with regard to a previously specified hypothesis, sampling bias can arise from lack of comparability of the groups.[21] For the researcher to be able to ascribe an observed difference in outcome between two groups to a particular putative determinant of that outcome, the two groups should differ with respect to that putative determinant alone. They should be comparable with respect to all other variables (or research procedures) which may be associated with that determinant and which may also determine the outcome.

Steps taken to ensure or check the comparability of the study groups should also be clearly specified in the protocol. These steps differ according to the study design used: if it is an intervention study,[4] some form of random allocation is desirable; if a case-control[5] or follow-up study,[6] other options (e.g. matching or prognostic stratification) should be considered.

Measurement bias

Measurement bias, or systematic measurement error, occurs when, for example, an observer or instrument measures an attribute in the same individual repeatedly higher, or repeatedly lower, than the true value. To find out whether measurement bias occurs, measurements taken as they would be in the study should be compared against measurements taken on the same subjects under the same conditions but using the 'gold standard' of measurement. Examples of such criterion-related validity are: 'How closely does the patient's recall of health care utilisation agree with documented utilisation for the same time period?', or 'How closely does the indirect measurement of blood pressure (with cuff and stethoscope) agree with the simultaneous intra-arterial blood pressure?' This implies that the true value must be measurable to assess validity; it certainly is the only way to find out how close to the truth the measurements really are. In a protocol, validation procedures, preferably involving a 'gold standard', should be clearly specified.

Sometimes no 'gold standard' exists, or the attribute to be measured is abstract, as is often the case with questionnaire items. Criterion-related validity cannot be estimated in such cases, and a second-best validation procedure called construct validity must be used. Here a manifestation of the attribute is substituted for the truth in an hypothesis constructed about the relationship between the attribute and its manifestation. For example, we may wish to measure the attribute 'satisfaction with health care', for which the truth is not measurable. A manifestation of this attribute may be use of the same service on a future occasion, and we could include a question on the respondent's intent to do just that. The measurement of satisfaction can then be validated against a manifestation of satisfaction, by examining the hypothesis that satisfaction with health care is associated with the intent to use the same service in future.

Sometimes neither criterion-related nor construct validity is feasible, and then even weaker validation procedures need to be used. Content validity refers to the apparent comprehensiveness of a measure, e.g. does a measure of the use of physician's services tap referrals, hospitalisations and casualty visits as well as primary care visits? Face validity is even weaker and refers to the credibility of the measure: does a measure of social function include questions about interactions with other people? These validation procedures should be specified in a protocol only if neither criterion-related nor construct validation is feasible.

Other important design-specific sources of bias, e.g. non-response, are discussed in detail elsewhere.[4-6,21]

4. Will the results be reliable?

Reliability (precision, reproducibility or repeatability) is defined as the closeness of repeated measurements of the same attribute to one another. The true situation has no bearing on the reliability of results. Lack of reliability is termed random error or variation and can arise from at least two major sources, sampling and measurement. The steps taken to minimise random error due to these factors should be clearly spelt out in the protocol.

Random sampling error

This reflects the variation around the true situation that would occur if many different random samples were taken from the target population and the results estimated from each. Generally speaking, precision can be improved by increasing the sample size, keeping in mind the earlier comments on epidemiological feasibility. In the protocol, the reasoning behind the suggested sample size should be made clear. Sample size can be estimated during planning by a statistician or epidemiologist, or looked up in tables[22,23] if the researcher provides four pieces of information: (*i*) the size of the expected effect or that which would be meaningful; (*ii*) the size of the random error or variation to be expected; (*iii*) the chance he/she is prepared to take in falsely saying there is an effect (alpha); and (*iv*) the chance he/she is prepared to take in missing a real effect (beta).

The general idea in using this information to estimate an optimum sample size can be seen from the following formula:

$$n = k \text{ (noise/signal)},$$

where n is the sample size; k is a coefficient incorporating alpha and beta, as well as the characteristics of the statistical test to be used; noise is an estimate of the random variation (e.g. standard deviation); and signal is the expected (meaningful) difference between groups.

From this it is clear that the larger the random variation, the larger the sample size required for statistical significance for a given difference will be, and vice versa. Similarly, the larger the sample size required for statistical significance for a given variation, the smaller the difference will be, and vice versa. Not so clear from this formula is the fact that the sample size required will be larger the smaller either the alpha or the beta value (or both) specified by the researcher. In practice, a single sample size is not estimated, but rather a range of sample sizes, using realistic variation in the abovementioned four pieces of information. The research can then decide how large a sample to use, given his/her resources and given the limitations imposed by the stipulated values for alpha, beta, the signal and the noise. The first two pieces of information often have to be estimated in a pilot study.

Random measurement error

This refers to the concordance reached in repeated measurements (in this pure sense, irrespective of how close to the true value they are). There are usually three parties involved in a measurement, giving rise to four possible sources of error. Inter-observer reliability refers to the concordance between two (or more) observers if they use the same instrument under the same conditions to measure an attribute in the same individual. Intra-observer variation refers to the concordance between two (or more) measurements of the same attribute in the same individual by the same observer using the same instrument under the same conditions. The same types of variation can occur with regard to the instrument, the subject, or the interpretation of the readings or responses. In the protocol, steps planned or taken to quantify this variation in a pilot study and/or in the final study should be specified. In addition the way in which this information will be or has been used should be specified, e.g. to standardise instruments, train and/or select interviewers, or adjust results. Examples of such sources of variation and techniques to quantify them are described elsewhere.[24-27]

5. Will appropriate analysis and interpretation be possible?

The answer to this question is inextricably bound with some of the issues mentioned under the first four questions, notably,

epidemiological feasibility (particularly sample size estimation), validity and reliability of results. However, there are some additional issues to consider here even if the earlier questions can be answered satisfactorily.

Details of data capture and management must be given. If data are to be analysed by computer, a pro forma should be enclosed with the protocol, and the computer and analysis package named, as well as the people who will do the analysis (if not the researcher). It is particularly important for the researcher to draw up some dummy tables of expected results. This may alert the researcher to tables that cannot be drawn up because the necessary data have not been collected, or because the necessary data have been collected, but in the wrong format. Unnecessary data may be identified in this way, and dropped, thereby increasing the feasibility of the study. In addition, specific objectives are often clarified by the act of drawing up such tables. In this respect it is useful to specify clearly in terms of the stated hypothesis which are to be regarded as response (dependent) variables, which as explanatory (independent) variables and which as interfering (confounding) variables. This tends to make researchers aware of how many different tables (combinations of variables) are possible, and may force them to consider only important issues for investigation or to consider obtaining statistical advice in the planning stage (perhaps even about multivariate analysis).

6. Are pilot studies being planned or have some been done?

This is an important issue and is mentioned separately to emphasise that enough time should be set aside to allow for planning the study, pretesting procedures and modifying the protocol before the final study. There should be enough time to do more than one pilot study if necessary.

Earlier, mention was made of pilot studies when considering the logistical feasibility, the validity and reliability of results, and the analysis and interpretation of results. All these objectives cannot be met in a single pilot study: (*i*) to test the logistical feasibility of the study, the exact protocol proposed for the final study needs to be executed; (*ii*) to validate results against a 'gold standard', subjects known to have a range of values when measured by this standard need to be measured by the test measurement as well to see how well they compare over the full range of possible values; (*iii*) to obtain information on the size of random error and of the expected effect for estimation of the required size, a random sample of the target population to be used in the final study needs to be drawn and measured; (*iv*) to quantify inter-observer error the same individuals need to be observed by different observers — these individuals need not be representative of the target population; and (*v*) to assess the feasibility of data capture, management and analysis, data from a pilot study should be put through the proposed analysis.

These pilot studies clearly cannot all be done on the same group of patients, and only after the other pilot studies have been executed and the final protocol decided on can the feasibility studies be done.

How many subjects should be used in a pilot study? Clearly, fewer subjects should be used than in the final study, otherwise the budget for pilot studies would exceed that for the final study. It is best to consult with an epidemiologist or statistician in every case since requirements may vary depending on the availability of subjects, type of outcome and the aim of the pilot study.

Can the data from a pilot study be added to those of the final study? This can be done only if the subjects or methods used in the pilot study do not differ from those in the final study. In practice these conditions are rarely, if ever, met because pilot studies are run expressly to find out how to change methods for the better.

Concluding remarks

When writing a protocol it is useful to try to ascertain how readers will evaluate the protocol. The protocol should therefore be reviewed and during such review all six questions, with sub-questions, should receive answers in the affirmative before the protocol can be labelled ideal or perfect. In practice if the protocol is read with funding in mind, the first four questions at least, have a hierarchy built into them: (*i*) if question 1 is not relevant, the project should not be funded; (*ii*) if question 2 is relevant, but not feasible, the project should not be funded; (*iii*) if question 3 is relevant and feasible, but results are invalid, it should not be funded until validity has been improved; and (*iv*) if question 4 is relevant, feasible and valid, but the results are unreliable, there should be no funding until reliability has been improved.

REFERENCES

1. Cochrane AL. Effectiveness and efficiency. In: *Random Reflections on Health Services.* London: The Nuffield Provincial Hospitals Trust, 1971.
2. Botha JL, Yach D. Epidemiological research methods: Part I. Why epidemiology? *S Afr Med J* 1986; **70:** 267-270.
3. Botha JL, Yach D. Epidemiological research methods: Part II. Descriptive studies. *S Afr Med J* 1986; **70:** 766-767.
4. Botha JL, Yach D. Epidemiological research methods: Part III. Randomised controlled trials (for interventions). *S Afr Med J* 1987; **71:** 657-660.
5. Yach D, Botha JL. Epidemiological research methods: Part IV. Case-control studies. *S Afr Med J* 1987; **71:** 763-766.
6. Yach D, Botha JL. Epidemiological research methods: Part V. Follow-up studies. *S Afr Med J* 1987; **72:** 266-269.
7. Spitzer WO. Ten tips on preparing research protocols. *Can Nurse* 1973; **69:** 1-4.
8. Yach D, Hessel PA. The research protocol in epidemiological studies: lessons from the review of 10 examples. *S Afr J Epidemiol* 1987 (in press).
9. Stolley PD, Schlesselman JJ. Planning and conducting a study. In: Schlesselman JJ, ed. *Case-Control Studies: Design, Conduct, Analysis.* New York: Oxford University Press, 1982: 69-104.
10. Burr ML. Criteria for planning a research project. *Community Med* 1979; **1:** 157-159.
11. Liechtenstein MJ, Mulrow CD, Elwood PC. Guidelines for reading case control studies. *J Chron Dis* 1987; **40:** 893-903.
12. Gore SM. Statistics in question: assessing clinical trials — protocol and monitoring. *Br Med J* 1981; **282:** 369-371.
13. Zelen M. Guidelines for publishing papers on cancer clinical trials: responsibilities of editors and authors. *J Clin Oncol* 1983; **1:** 164-169.
14. Botha JL. How not to do clinical trials. *S Afr J Hosp Med* 1982; **8:** 122-126.
15. Gardner MJ, Machin D, Campbell MJ. Statistics in medicine: use of check lists in assessing the statistical content of medical studies. *Br Med J* 1986; **292:** 810-812.
16. Bankowski Z, Howard-Jones N. *Human Experimentation and Medical Ethics.* Geneva: Council for International Organisations of Medical Sciences, 1982.
17. Lochner JdeV. *A Guide to Ethical Considerations in Medical Research.* Parowvallei, CP: South African Medical Research Council, 1979.
18. Zelen M. A new design for randomized clinical trials. *N Engl J Med* 1979; **300:** 1242-1245.
19. Smith W, Zelen M, Smith W. Randomization and optimal design (Correspondence). *J Chron Dis* 1983; **36:** 609-612.
20. Hessel PA. Ethical issues in epidemiological research. *S Afr Med J* 1987 (in press).
21. Sackett DLS. Bias in analytic studies. *J Chron Dis* 1979; **32:** 51-53.
22. Fleiss JL. *Statistical Methods for Rates and Proportions.* New York: John Wiley, 1981: 260-280.
23. Sackett DLS, Haynes RB, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine.* Boston: Little, Brown, 1985: 186-187.
24. Department of Clinical Epidemiology and Biostatistics, McMaster University. Clinical disagreement: I. How often it occurs and why. *Can Med Assoc J* 1980; **123:** 499-504.
25. Department of Clinical Epidemiology and Biostatistics, McMaster University. Clinical disagreement. II. How to avoid it and how to learn from one's mistakes. *Can Med Assoc J* 1980; **123:** 613-617.
26. Koran LM. The reliability of clinical methods, data and judgements: part I. *N Engl J Med* 1975; **293:** 642-646.
27. Koran LM. The reliability of clinical methods, data and judgements: part II. *N Engl J Med* 1975; **293:** 695-701.