Tips for Making a Good Research Proposal

Siti Setiati

INTRODUCTION
An important activity in the process of conducting a scientific study is the writing the research proposal prior to the commencement of the study. A research proposal is created for two aims, which are: to be used as guidelines in conducting the study and to explain the items used as guidelines in conducting the study to aiding parties. A good research proposal should generate a good research. Nevertheless, many people reverse the study process by creating a research report based on available data, without initially formulating a research outline or proposal. The research hypothesis is instead formulated after the data is obtained. Such data dredging or data fishing is greatly inadvisable. This paper will discuss several important points that demand attention in writing a research proposal.

RESEARCH TITLE
A good research title should be able to reveal a common problem that will be studied as well as the main independent and dependent variables. The title should consist of a short, simple, phrase (approximately 8-10 word long).

THE BACKGROUND OF THE PROBLEM
The first question that appears when a person is reading a research report or proposal is the novelty or “pearl” of the study. Why is it important to conduct this study? Is it something new that has not been studied before? Will its method or substance add to or improve previous studies, or is there controversy in previous studies? The novelty of the research should be clearly stated in the introduction. Often, there is not much information to be obtained from a study, making what is going to be studied unclear or unfocused. On the other hand, sometimes there is inadequate sample size for the number of hypotheses.

PROBLEM FORMULATION
The research problem could be formulated by means of the differentiation of an existing theory. A good research problem must be specific (discussing only a single aspect), could be empirically tested, reflects a correlation between variables, has a theoretical basis or framework, has practical use, is expected to contribute to science, and be up-to-date.

Be Careful with Multiple Hypotheses
Researchers often make multiple hypotheses without realizing that the sample size is inadequate. The research power becomes very small, making it unable to detect differences or the relationship among variables.

LITERATURE REVIEW AND THEORETICAL FRAMEWORK
Literature review and theoretical framework are important aspects, and a great deal of effort needs to be placed in its making. A good literature review should provide a complete picture of the theoretical framework of the proposed study. All of the factors that will be studied or that are associated with the factor that will be studied should be critically discussed in depth in reference to the literature or previous studies. Care should be taken in citing previous studies to ensure that the studies are of value and worth citing, or whether they belong to the group of “garbage in, garbage out” (GIGO) studies.

The Hill criteria can be used as guideline for citation or critique previous studies. The Hill criteria mentions several things that we need to pay attention to in order to determine whether the study results truly demonstrate a causal relationship, which are: strength, temporality, consistency, plausibility, dose response, coherence, and

Division of Geriatric, Department of Internal Medicine, Faculty of Medicine, University of Indonesia, Dr. Cipto Mangunkusumo Hospital, Jakarta
in concordance with other studies. Strength pertains to the correlation between two variables, as demonstrated by the p value, where a small p value (and therefore narrow confidence interval) is stronger than a larger p value (or wider confidence interval). If the odds ratio (OR), risk ratio (RR), or prevalence ratio are being calculated, the value farthest from the number 1 is stronger than values approximating 1. Temporality depicts how events could only have a causal relationship if it is clearly believed that the cause precedes the effect (the independent variable preceding the dependent variable). This aspect could be illustrated through clinical trials, cohort studies, and case control studies, with the strength of correlation in descending order. Dose-response demonstrates the degree of change association with a change in dose of exposure or risk factor, making a cause-effect association more possible. If there is a consistent result, both among subject groups within the study or with other studies, a causal association is more possible. In order for a relationship to be causal, the relationship between dependent and independent variables should be explainable using current theories (plausibility). A relationship is considered coherent if it is in line with the general distribution of risk factors and effects in a certain population. If the study results support what is found in other studies, a causal relationship is more possible.

After the literature review has been completed, a theoretical framework could be formulated. A theoretical framework is the framework of relationships between all of the concepts, constructs, or variables, which are systematically organized to explain a particular phenomenon. Thus, all of the variables associated with the phenomenon that is to be studied and their organization are written in the theoretical framework.

In developing the thinking framework, all of the important variables are decided, and their correlation with a central hypothesis identified. A causal model is useful to identify all relevant variables that influence the hypothesis, and to determine the causal relationship between variables.

THE FACT-HYPOTHESIS MATRIX

The hypothesis-facts matrix is a systematic way of organizing the causal relationship suggested in each causal model. Each relationship is identified as a fact or hypothesis. The more a causal relationship is not proven or hypothetical, the more difficult it is to satisfactorily predict the outcome of the study. In order to facilitate assessment of various literature or previous studies related to the phenomenon that will be studied, a fact-hypothesis matrix could be formulated. Table 1 illustrates an example of a fact-hypothesis matrix.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Cause</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orthostatic hypotension</td>
<td>Hypertension</td>
<td>Mader SL. JAMA (1987)</td>
</tr>
<tr>
<td>Orthostatic hypotension</td>
<td>Age</td>
<td>Myers MG. CMA Journal</td>
</tr>
<tr>
<td>Orthostatic hypotension</td>
<td>Sex</td>
<td>(1978)</td>
</tr>
</tbody>
</table>

CONCEPTUAL FRAMEWORK

The conceptual framework is a miniature of the theoretical framework, drawn up in the form of a diagram containing only the variables that will be studied. Other variables that will not be studied but is associated with the phenomenon in question (dependent variables) are placed in the theoretical framework. Thus, the theoretical framework should be differentiated from the conceptual framework. The diagram in the conceptual framework should be able to demonstrate the correlation between variables. A good conceptual framework should provide clear information to facilitate the researcher in choosing the study design. A common mistake is to make the study design framework as the conceptual framework.

The Variable Indicator Matrix

The variable indicator matrix is a systematic way of organizing relationships between the studied variables and the potential indicator of the variable in question. Variables are characteristics of the study subject (example: nutritional status), while indicators pertain to the measurements to be taken, which are assumed to portray the variable in question. The indicator should be the most valid form of measurement towards the variable that will be measured (e.g. the indicator for the variable nutritional state is body mass index (BMI) or body weight/age). The variable indicator matrix has to correlate each variable (cause) of the 1 causal model in at least one indicator, demonstrating the method of measuring the indicator, noting the reference of the chosen measurement method. Table 2 demonstrates an example of a variable indicator matrix.

Domain

The study sample selected from the population ideally should represent the population, so that characteristics of the study subject represent those of
the whole population. Thus, it is very important to know the research domain (target population). The understanding of population in a study refers to all subjects (which can be people, laboratory animals, etc.) that meet certain desired characteristics. The domain or target population is the final target population for the application of the study findings. The target population is general, and in clinical studies are usually limited by demographic characteristics (age, sex), and clinical characteristics. Examples of target populations are healthy children, elderly patients. The accessible population is the portion of the target population that is accessible to the researcher. This population is a part of the target population limited by time and space. The sample will be selected from this accessed population, consisting of the subjects that will be directly studied. The selection of the accessible population is solely based by reality, and not based on a systematic selection process. On the other hand, sampling from the accessible population has to be conducting using a particular procedure, in order to obtain a sample that is representative of the accessible population. The sample selection will determine the external validity of a study.

**METHOD**

**Design**

The study should be designed according to the motive of the study, and it is extremely important to consider various things, such as previous studies, the frequency of the problem that will be studied, and cost. The study design should also be in line with the study aim and hypothesis. In order for the collected data to remain objective, reliable, and valid, we should choose the most appropriate, inexpensive, effective, and efficient study design. Study designs are classified based on the presence or absence of intervention, which of observational studies (cross-sectional, case-control, and cohort studies), and experimental studies (pre-post test, and clinical trials). There are four things that deserve our attention prior to selecting a study design, which are time frame, program or intervention, observation or measurements, and individual or group research subjects. If only an observation would be made, it needs to be determined whether the observation would be made at one point in time or a follow-up will be conducted after a certain period of time. It also needs to be determined whether the researcher plans to study an event that has taken place (retrospective study) or to follow a subject in order to study an event that has not taken place (prospective).

**Sampling**

We need to consider whether the samples or research subjects are chosen randomly, either by means of random selection of observational studies or by means of random allocation in clinical trials. Random selection pertains to random drawing or selection of study samples from the population. Sampling could be classified in two, probability sampling, and non-probability sampling. If the samples are randomly chosen, it is expected that we could generalize the study findings to a larger population to a certain extent. Thus, random selection is highly associated with the external validity of research findings. It is better to choose a sample using probability sampling. If this is not possible, due to a limited number of patients, for clinical studies, consecutive sampling is recommended. The use of convenience sampling and judgmental sampling should be avoided. Random allocation or randomization pertains to determine which study subject will receive a different treatment and which ones will be control subjects (in clinical trials). Randomization creates two groups, the test group and control group, with identical characteristics, to avoid confounding and selection biases. Thus, random allocation is highly correlated to the study design, and further guarantees the internal validity of the study.

**Inclusion and Exclusion Criteria**

In determining the inclusion criteria, the researcher should be careful to ensure that the criteria is relevant to the study problem. The study inclusion criteria pertains to the general characteristics of the study subjects within a target or accessed population. Some subjects that meet the inclusion criteria need to be excluded from the study for several causes, such as: the presence of other conditions/diseases that could confound measurements or interpretation of results, conditions that make it difficult to perform the study, such as patients with no permanent residence, which make it difficult to follow up, ethical considerations, and subjects who refusing to participate. Inclusion and exclusion criteria should be carefully formulated, since they also determine sample quality and of course the internal validity of the study.
Sample Size

A good research should state the estimated minimally required sample size. The sample size is often ignored or disregarded, and is thus not calculated according to the correct sample size formulation, when in fact there is a phrase that says “Too many samples prove everything, too few samples proves nothing”. The calculation of sample size is often confused with sample size to estimate prevalence and to test a hypothesis.

The size of the sample required for a study depends on the aim of the study, for example to determine the prevalence (percentage) of a certain matter, or to determine the significance or the lack of difference in significance of two conditions. The sample size also depends on the type of study, whether it is a survey, case-control, cohort, or experimental study. In addition, the sample size also depends on sampling, whether it pertains to a paired sample, or an independent sample, the expected statistical significance, whether it is at the level of p=0.05 or other, and the available facilities (human resources, instruments, and funds).9

Another thing that needs to be considered in determining the sample size is the type I level of error (α) and research power (1-β) that will be used. The α value needs to be determined by the researcher, and is usually either 0.05 or 0.01. This value influences the sample size, where a lower α will necessitate a larger sample size. The research power (1-β) pertains to the capacity of the study to obtain differences that are statistically significant, and whether such difference exists in the population. The power is determined by the researcher, common ones being 80% or 90%. The higher the expected power, the smaller the b, and the larger the sample size needed.

It is not uncommon for a single study to consist of several designs. For this, each sample size should be calculated according to the formula for each design used. From the calculation, the greatest number of subjects should be selected to be used for the study. In estimating sample size, researchers should try to be economical, logical, and creative10.

The Need for Control and Random Allocation

For studies designed as case-control, cohort studies as well as clinical trials, the control plays an important role. The control subjects should come from the same population as that of the case subjects, so that the case subjects and the control subjects have the same opportunity to be exposed to the risk factors. There are several ways to choose a good control, which is to choose the case and control subjects from the same population. For example, if the cases come from all patients of a certain population, the control should be randomly selected from the rest of the population. The case or control subjects could also be selected from the predetermined population, which is usually smaller (cohort) and matched, where control subjects are chosen according to matching characteristics with the cases for each variable that may play a role as a risk factor but is not being studied, and by selected more than one control group.

An important aspect associated with the selection of control subjects is random allocation, particularly for clinical trials. The main aim of random allocation is to limit selection and confounding biases, due to unequal distribution of variables that are not being studied among the groups. Several randomization methods that are commonly used are simple randomization, block randomization, and stratified randomization11.

Be Alert for Bias (for Each Type of Design)

Biases that could occur in a study are classified as selection bias, measurement or observational bias, or confounding bias. Selection bias could occur if the study subjects are not representative of the population they are suppose to represent. Bias associated with measurement bias could be procedural bias, recall bias, bias due to insensitive measurement, detection bias, and cooperation bias. Procedural bias occurs if the measurement, procedure, treatment, or other things that the groups are subject to are not the same. A way of avoiding this bias is by masking the procedure (double-blind study), so that the researcher does not know which subject belongs to which group. In the detection bias, there is a change in the ability of the measurement instrument to detect the disease. Cooperation bias occurs if the cooperation in following the procedure is different from one group to the next. Confounding bias occurs if there is a confounding variable in the study, and the researcher does not attempt to eliminate or control the variable. Confounding biases are variables associated with the independent and dependent variables, but are not variables in-between, which could distort the correlation between the independent and dependent variables that are being studied. Identification of the confounding variable is very important, because if not, it would result in an incorrect interpretation. Confounding variables could be identified through adequate literature review, besides experience and logic. Restriction, matching, and randomization could be performed in order to eliminate confounders in the design. Confounders could be eliminated from the analysis by means of stratification and multi-variate analysis12.
Internal and External Validity

The internal validity of a study demonstrates whether the study results are free of randomization errors, bias, and confounding factors. Studies with a high internal validity have zero or minimal bias, randomization error, and confounding factors. On the other hand, a study with low internal validity reflect the presence of bias, randomization error(s), and/or confounding factors, which means that the association(s) found may be due to something other than the studied variable. A research is required to have a high internal validity, which means that it is free from various kinds of bias.

External validity demonstrates how much the findings of a study could be applied for a larger population. A new study could have a high external validity if it has a high internal validity. Studies with a poor internal validity could not possibly have a high external validity, rendering the question of external validity to be irrelevant.

CONCLUSION

A research proposal is the written plan for the study, written with the aim of guiding the researcher throughout the research process. When writing a research proposal, the researcher should have the ability to come up with the idea, creativity, methodological capacities, mastery of the material, an understanding of statistical application, language abilities, and logical consistency. The different parts of the research proposal are not separate, but is presented in a logical order. The whole research proposal should be arranged cohesively.

REFERENCES