Design of Clinical Trials in Radiology for Improved Assessment of Diagnostic Tests.

Stavroula J. Theodorou, M.D., Daphne J. Theodorou, M.D. Department of Radiology; University of California, San Diego, Veterans Affairs Medical Center, San Diego, CA

ABSTRACT

Proper design of clinical trials has become a prerequisite for both the performance and interpretation of research in radiology as well as in other disciplines in medicine. This paper reviews the process of designing clinical research, the research question, the review of existing data, the study design, the study population, and the study measurements. *Theodorou SJ, Theodorou DJ. Design of clinical trials in radiology for improved assessment of diagnostic tests. Internet Medical Journal, 2005, 5:1. http://www.medjournal.com/forum/showthread.php?p=1450*

In recent years, several studies have demonstrated the inadequate level of knowledge and understanding of study design and statistics held by physicians including radiologists (Weiss and Samet, 1980, Berwick et al, 1981, Goldin and Sayre, 1996). Rigorous research begins with the precise definition of the research question, which describes the uncertainty of something observed in the population that the researcher believes can be resolved by making measurements on study subjects (Hulley and Cummings, 1988).

THE RESEARCH PROTOCOL

The research protocol is the written plan of the investigation that defines the research question, the rationale for the study, the materials and methods to be used in detail, and the anticipated problems that may occur in the execution of the study. Optimally, the steps taken to avoid bias and thus, eliminate the most common reason responsible for the lack of validity should also be described in detail in the study protocol. Ideally, the protocol is evolving and improving during the execution of the study, as the researcher reevaluates and revises its components. A timeline of the expected dates of completion of each part of the study and factors which warrant the feasibility of the study, including financial and other logistical constraints encountered in the research process, should be precisely

stated in the protocol, as well. Indeed, research funding is one of the major parameters that ensures the feasibility of the research study. In fact, a rigorous and detailed protocol should always be a prerequisite for all clinical trials and in terms of academic honesty, should never aim at the request of research funds, alone.

THE RESEARCH QUESTION

The research question is developed in a form termed the hypothesis. Hypothesis defines the study design, the sample, and the measurements that need to be made. Good research questions should qualify the following criteria: I) importance, II) novelty, III) feasibility to answer, IV) ethics, and V) relevance. The following research question serves as an example for illustrative purposes. Suppose a new imaging modality (IM) has been developed that provides the radiologist with data concerning the disease (D). The research question inquires whether IM provides information associated with the diagnosis of D. IM would be of any significant value and the research question would prove a good one, if only D is a major health problem resulting in great morbidity, if left undiagnosed. If the diagnosis of D can already be made utilizing other existing imaging modalities or even can be attained on the basis of clinical and/or laboratory findings, then the research question would be of a limited importance.

Traditionally, this question of importance is also known as the "so what? " criterion, in terms of why someone should spend his time and money to answer it. With respect to the answer to the research question itself, ideally it has to be novel, so that it can provide additional information to the already existing scientific knowledge or optimally constitute new information. Given the necessity of a thorough investigation of previous studies associated with the present research question, a review of the literature is considered extremely important. However, this certain procedure may lead to the modification of the initial research question, since questions which remain unanswered may be identified.

For the purposes of the study, one of the main features of the research question is the answerability, as the research question should be stating the study's major clinical variables and also be explicit. In the aforementioned example however, the research question is rather vague and as a result it remains unanswerable, necessitating a modification. Indeed, questioning whether IM is sensitive and specific for the diagnosis of D can be challenging and thus, sufficient for facilitating answerability. Nevertheless, it must be noted that definition of the sensitivity and specificity of IM is a prerequisite for a positive or a negative answer. Accordingly, the research question

arising is whether the sensitivity and specificity of modality IM is higher than those of the modality A for diagnosing D.

With regard to ethics, a properly posed research question should be characterized by three substantial principles: I) respect for people, II) beneficence, and III) justice (Hulley and Cummings, 1988). With respect to the first principle, all study subjects after being informed about the potential risks and benefits of the study sign a consent form. Cooperation between the researcher and the subjects is only attained when the latter do not feel like objects to be studied, but rather like collaborators and contributors who volunteer for the humanity well-being. The research question is beneficent when the benefits of the research are proportionate to the potential risks of the subjects. The even distribution of the benefits and burdens of research to all people with no discrimination regarding physical disability, races, and imprisonment status characterizes an ethical research question. In accordance to the ethics of the study, it is also the responsibility of the researcher to conceal the identity of the people willing to be study subjects and thus, preserve patients' confidentiality. In terms of academic honesty however, deviations from good research practices may place intellectual integrity at increased risk for misconduct and violation. Fabrication of fictitious data or results, falsification and selective, non-scientific presentation of information, plagiarism and

forgery of the ideas of other investigators represent deliberate and serious forms of academic deceit.

THE STUDY DESIGNS

Given that the study design is the methodology for answering the research question, an understanding of the study design is important in analyzing the concept of a research protocol. Prospective studies provide more explicit data as all the subjects undergo a determined protocol and the data are collected uniformly. Conversely, retrospective studies are characterized by a lower cost and are definitely less time-consuming processes.

In a cohort study, which is well-accepted as the most scientifically sound type of study, the subjects are followed over a certain period of time in order to describe the incidence or natural history of a condition, and to analyze the risk factors as well as the confounding factors (Hulley and Cummings, 1988). It is not surprising that the cohort study requires a prolonged study time, enough to allow the condition to develop, as well as numerous study subjects. The measurement of the risk of an outcome relative to whether a predisposing factor is present, known as the relative risk (rr) and the odds ratio is of particular interest in medical research. The relative risk of the disease or the risk ratio can only be calculated from cohort

studies. It represents the ratio of incidence in persons that have been exposed to the predisposing factor (risk group) to incidence in those who have not been exposed to the predisposing factor (control group). The relative risk can be estimated by the following formula:

$$rr = \frac{\frac{\text{total number in risk group with outcome}}{\text{total number in a risk group}}}{\frac{\text{total number with outcome in control group}}{\text{total number in control group}}}$$

The odds ratio is the odds that a patient is exposed to the risk factor divided by the odds that a control is exposed (Prentice et al, 1975).

In the cross-sectional study, both the dependent and the independent variables are measured at a single point in time. By comparison to the cohort study, a cross-sectional study may be less time-consuming and may even cost less, but unfortunately, the chance for error is increased. In a case-control study however, a retrospective analysis of the prevalence of risk factors in a sample of patients defined as the cases, is compared to the prevalence of risk factors in a sample of people free of disease, defined as the control group (Hulley and Cummings, 1988). Not uncommonly, in radiology research involves the evaluation of imaging diagnostic methods. In

this setting, diagnostic methods are usually appreciated with randomized blinded trials.

THE STUDY POPULATION

The next step to be taken is the determination of the study population, that is to specify the group of patients in whom the research question has to be answered. In specific, in our theoretical example the research question arising is whether sensitivity and specificity of IM are higher than sensitivity and specificity of modality A for diagnosing D in the patient population P. For the purposes of the study, the target population needs to be defined by stating the clinical and demographic features of the subjects included. Not infrequently, initial definition of the target population ends up in a large number of patients that needs to be remarkably further eliminated. With the definition of inclusion or selection criteria such as the age range, gender, geographic location and time, clinical symptoms, laboratory findings, type, stage, and severity of the disease, the number of patients eligible for the study, constituting the accessible study population is determined.

Although the accessible population by and large meets the inclusion criteria set for the study, it may still be too large to be studied. For this reason, a sampling method resulting in the intended

sample needs to be applied. Indeed, numerous sampling methods ranging from random sampling to judgemental sampling exist. However, a randomized controlled study design should always be considered when the research question concerns the evaluation of a new diagnostic method (Hulley and Cummings, 1988). Therefore, the study design selected, as well as the reason for its choice should also be clearly stated.

TYPE OF SAMPLING

It becomes obvious however, that each time the selection of the most appropriate type of sampling depends on the research question. It is not surprising that the judgemental sampling introduces selection bias to the research. It is noteworthy however, that consecutive sampling approximates random sampling, in a fashion that the next series of patients to be included in the study is already predefined. Indeed, random sampling may deal with confounding factors such as the existence of a second illness, patient's uncooperativeness which in fact, may influence the other variables being studied. Finally, a recruitment strategy maximizing the number of patients eligible to be actually participating in the study, is also required. Overall, the study population should be representative in terms of that the findings in the actual sample should be applicable to the target population, a feature

of the study well-known as generalizability. When the actual sample does not reflect the target population itself, a serious error known as sampling error, occurs. Indeed, sampling error maybe systematic and thus, may alter the study findings.

In this setting, the intended sample may be improperly defined or there may be recruitment bias in the actual sample. Likewise, sampling error may be random owing to sample selection. For this reason, large study populations are considered especially well suited for the elimination of random error. An obvious limitation to a retrospective study however, is the development of an appropriate study population, as the investigator is practically unable to control the inclusion criteria set for the study. When writing up an article including observations about a certain entity in a series of patients, it is fundamentally important for the materials and methods section to be written early, while the study is being designed. The benefit is two-fold as inclusion criteria to the study population have to be carefully outlined from the very beginning and the researcher is prompted to think thoroughly through the methods.

THE STUDY VARIABLES

In a thoughtfully designed study, it is useful to identify which variables are dependent, independent, and interfering (confounding). For

example, included among dependent variables are the clinical outcome, while included among independent variables are the age, gender, and clinical symptoms of the patient, which indeed represent factors being under researcher's control.

A variable may be dichotomous, categorical or continuous. A dichotomous variable represents one of two choices e.g gender (male or female). A categorical variable is one of several choices e.g race (white, black, asian etc.). A continuous variable is one of infinite choices e.g age. The gold standard variable is usually one of the dependent variables, against which other dependent variables are compared. An imperfect gold standard variable may introduce biases into the study results. Specifically, bias occurs when the dependent variable being evaluated is used to define the gold standard variable (McNeil, 1991). The measurements of the study are enforced when precision and accuracy are maximized. Precision describes the degree to which a variable has nearly the same value each time it is measured. When there is a high precision, increased statistical power is produced. Accuracy describes the degree to which the variable actually represents what it is intended to represent, expressing how close the measurement is to actual reality. Precision, an indicator of measurement reproducibility, increases when the variability of the observer, the subject, or the instrument decrease. In particular, the

presence of a standard observer trained in the study methods and the use of an operation manual may significantly reduce the observer variability.

VARIABILITY IN CLINICAL STUDIES

The specification of observer/reader population (the radiologists interpreting the diagnostic test) indicates a whole new dimension in the problem of assessing diagnostic modalities (Hanley, 1988, Hanley, 1989). It is conceivable that a study involving a select group of highly-specialized readers would provide accuracy assessment classified in the upper end in the distribution of accuracies between the readers' population nationwide. When the researcher interpretes the results he should always bear in mind the presence of intra- and inter-observer variability. Both interobserver variability testing the degree of agreement among different observers under the same conditions, and intraobserver variability testing the degree of consistency of a single observer at different times, are of particular importance.

Subject variability measured by subjects undergoing more than one test, may affect study result interpretation. Instrument variability is quantified by repeat measurements. In addition, refinement and automation are considered helpful in reducing instrument variability. The specification of the technical features of the diagnostic imaging

modality constitutes a standard part of the research protocol. Optimally, when participating in a certain study, all patients should have to be studied using the same imaging technique, unless a comparison between alternative techniques (e.g use or not of contrast medium) is to be made.

To recapitulate, the thorough design of a research study requires experience, close supervision by an attentive and a dedicated investigator who will control the methodology protocol, maintenance of accurate records of research procedures and results, and above all, abundant time to be invested in the on-going effort. Research findings and conclusions drawn from the research become available to the scientific community through open communication and publications. Particularly, the ideal diagnostic test has to present certain features. In fact, the diagnostic test has to be quick, not complicated, painless for the patient, reliable, inexpensive, and sufficient to provide the right answer (Hulley and Cummings, 1988). In accordance to the aforementioned criteria, a diagnostic test that already complies with the majority of them, may prove that the introduction of a new diagnostic test finally, is not noteworthy.

KEY POINTS

- The research protocol is developed to answer the question.
- The research question should qualify the following criteria: I) importance, II) novelty, III) feasibility to answer, IV) ethics, and V) relevance.
- An ethical research question should be characterized by respect for people, beneficence and justice.
- Diagnostic methods are most commonly appreciated with randomized blinded trials.
- The study population is the group of people from whom data is collected for the study.

REFERENCES

1. Berwick DM, Feinberg HB, Weinstein MC (1981) When doctors need numbers. Am J Med **71:**991-8
2. Goldin J, Sayre JW (1996) A guide to clinical epidemiology for radiologists: Part II statistical analysis. Clin Radiol **51:**317-24
3. Hanley JA (1988) Alternative approaches to receiver operating characteristic analyses. Radiology **168:**568-70
4. Hanley JA (1989) Receiver operating characteristic methodology: the state of the art. CRC Crit Rev Diagn Imaging **29:**307-35

5. Hulley SB, Cummings SR (1988) Designing clinical research: an epidemiologic approach. Williams & Wilkins, Baltimore, MD: 1-235

6. McNeil B (1991) Experimental design for clinical studies and the importance of statistics. Presented at the Conjoint Committee on Diagnostic Radiology's 5th
Grantmanship Workshop, November 8-9, Washington DC

7. Prentice C, Forbes C, Morrice S, et al (1975) Calculation of predictive odds for possible carriers of haemophilia. Thromb **34:**740-7

8. Weiss ST, Samet JM (1980) An assessment of physician knowledge of epidemiology and biostatistics. Am J Med Education **55:**692-703