This set of guidelines was written to provide a systematic explanation of the process of evaluation applied to regional Medical Programs, as required by Public Law 89-239. Goals of the programs are the improvement of health care of patients suffering from heart disease, cancer, stroke and related diseases and improvement in the practice of health professionals. The first step in evaluation is the development of objectives--both immediate and long range. The second phase is the selection or design of measuring instruments or the design of other procedures to collect data that will lead to evidence for evaluation. Next comes the collection of data--from the health professional as a participant in a learning experience and as practitioner, and also from society. The fourth phase is analysis of the data; then judgment is made of how well objectives have been met. (A checklist is included; an appendix gives examples of decisions and modification.) (PT)
REGIONAL MEDICAL PROGRAM

GUIDELINES FOR EVALUATION

Gary S. Dean, Ed.D.
Sandra A. Robinson, Ed.D.
Bruce E. Strem, Ph.D.
Jack E. Thomson, Ed.D.

University of Southern California
School of Medicine
Division of Research in Medical Education

March, 1968

This report was written by Post-doctoral Fellows in the Training Program in Evaluation in Health Care Education supported by National Institutes of Health, Contract PH-43-68-62, Division of Regional Medical Programs, United States Department of Health, Education, and Welfare.
THE PURPOSE OF THIS DOCUMENT IS TO DEFINE THE STEPS OF EVALUATION.

THE AUTHORS' INTENT IS TO HELP PLANNERS OF R M P PROJECTS CONTROL THE STRATEGY OF THE EVALUATION FOR WHICH THEY WILL ULTIMATELY BE RESPONSIBLE.
2. Skills 11

3. Attitudes 11

C. The Health Professional as Practitioner 12

D. Society 12

V. Analysis of the Data 12

VI. Judgments and Decisions 13

A. Judging How Well Objectives Were Met 13

B. Decisions 13

C. Participants' Reactions 14

VII. Evaluation Checklist 15

Appendix - Examples of Decisions and Modification 17
REGIONAL MEDICAL PROGRAM
GUIDELINES FOR EVALUATION

I. Regional Medical Programs and Evaluation

A. Public Law 89-239

In October, 1965, Public Law 89-239 was enacted. The purpose of this law is to provide the medical profession with the means to make available to their patients the latest advances in the diagnosis and treatment of heart disease, cancer, stroke and related diseases. To this end, Regional Medical Programs were established to provide research, training and demonstrations of advances in patient care with respect to these diseases.

B. Origin of the Evaluation Requirement

Early in 1966 the Division of Regional Medical Programs was established, and its Planning and Evaluation Branch began to develop procedures for reviewing proposals for the planning and operational phases of the several Regional Medical Programs. These review procedures called for (1) an initial assessment of the type of activities to be carried out under the proposed planning grant and (2) continuing assessment to determine how well the regional goals were being achieved.

C. Procedure of the Division of Regional Medical Programs

When a proposal is sent to the Division of Regional Medical Programs multiple copies are made. Various members of the staff look over that part of the proposal concerned with their specialties (cardiology, evaluation, etc.). The proposal and the recommendations of the specialists are sent to a Review Committee. In the meantime two members of the Review Committee review the entire state, region or planning grant (if one is available). The Review Committee then offers one of five recommendations, from accept to reject, and the proposal is sent to the National Advisory Council, which makes the final decision.
D. Regional Advisory Group and Evaluation

Prior to submission to the Division of Regional Medical Programs, the initial planning grant application and subsequent operational grant applications must be reviewed and approved by the Regional Advisory Group, the guiding body of each Regional Medical Program. The Regional Advisory Group must then demonstrate to the National Review Committee and to the National Advisory Committee the effectiveness of its coordinated activities to advance the attack on heart disease, cancer, stroke and related diseases. The Regional Advisory Group, following the Official Guidelines, therefore requires that a local group, in submitting suggestions for operational projects, must demonstrate how the goals defined in the proposals will be met and how evidence of success will be obtained. The local group must be able to demonstrate that its effort has produced significant change in information, skills or attitudes of physicians and of other professionals in the health community.

E. Evaluation Defined

Evaluation is a five-phase process. The planner of a teaching-learning experience:

1. Defines objectives, i.e., makes statements about the learner each with an action, a content and an evidence component.

2. Selects or designs measuring instruments or other procedures to collect data that will lead to the evidence specified in phase one.

3. Collects data.

4. Analyzes and summarizes the data he has collected so they can be readily comprehended and so they constitute pertinent evidence.

5. Uses the summary evidence to judge how well objectives have been met.

The five phases of evaluation will be discussed in succeeding sections and guides will be presented to help the planner accomplish each phase.
II. Goals and Objectives

Stating goals sets the stage for the first phase of the evaluation process, definition of objectives.

A. Goals

The long-range goal of RMP is improved health care for patients suffering from heart disease, cancer, stroke and related diseases. This long-range goal can be achieved through coordinated intermediate and immediate goals. The intermediate goal of RMP is improvement in the practice of health professionals as a result of increased knowledge or skill or as a result of changed attitudes. The immediate goal of RMP is, then, to increase the fund of health professionals' knowledge and skill and to change their attitudes such that they will use the new knowledge and skills.

As the initial step in the planning process, the planner of a training program defines some general goals consistent with the overall goals of RMP. He can describe these goals in a general way, e.g., "I wish to increase physicians' knowledge of new research on the effect of the use of Pap smears* on cure or survival rates in cervical and uterine cancer." From the general goal, he can develop more specific immediate, intermediate and long-range goals.

Using, for example, a symposium on Pap smears

1. The **immediate goals** are to provide physicians with:
   a. knowledge regarding the value of Pap smears.
   b. skill in the procedure.
   c. attitudes conducive to maximal utilization of the procedure.

* Cervical smears using the Papanicolaou technique, hereafter referred by the common abbreviation "Pap smears."
2. The **intermediate goal** is that the information and attitude developed will lead to an actual change in behavior by the physician, i.e., increased frequency of taking Pap smears in routine physical examinations and of mailing reminders to patients to come in for routine Pap smears.

3. The **long-range goal** is that the increased frequency of taking Pap smears in routine physical examinations will result in the detection of more cases of early cancer of the cervix thereby making early treatment possible and improving health care.

B. Objectives

The first step in evaluation is the development of objectives. Objectives are derived from goals. Goals are statements about what the planner hopes to achieve; objectives are statements about the learner or participants in a program. They differ from goals in two ways:

1. Objectives state the desired learner behavior explicitly.

2. Objectives specify evidence that the behavior has occurred.

Immediate objectives are derived from immediate goals; these are likely to be concerned with learner behavior during or immediately after the learning experience. The focus here is **can** he perform the new behavior.

Intermediate objectives are derived from intermediate goals; these are likely to be concerned with learner behavior at some later time, six months or a year after the learning experience. The focus here is **does** he perform the new behavior.

Long-range objectives are derived from long-range goals; these are likely to be concerned with changes over longer periods of time and may include the effect the changed behavior has on the society.

The definition of objectives and the specification of pertinent evidence comprise an iterative process. After an objective has been written, the planner may find that pertinent evidence is
not attainable within his budget, and he will then have to revise his objectives. An objective for which pertinent evidence cannot be collected should be deleted from the list of stated objectives for any effort.

C. How to Write Objectives

Objectives tell what is expected to happen to the learner as a result of participation in a program, i.e., what he will be able to do when he has completed the learning or training experience that he was not able to do before. Objectives have three parts: an action part, a content part, and an evidence part.

1. **Action.** This part of the objective states what the learner will be able to do after the learning experience that he could not do before.

   Examples: "to know," "to identify," "to administer"

2. **Content.** This part of the objective defines the area or subject matter or content in which the learner shall act.

   Examples: "to know latest research associating routine Pap smears with percentage cure in cervical and uterine cancer."

   "to identify squamous cells in prepared microscope slides."

3. **Evidence.** This part of the objective specifies (1) how learner behavior is to be observed or tested and (2) states what constitutes acceptable performance or evidence that the learner has achieved the behavior.

   Examples: "to know latest research associating routine Pap smears with percentage cure in cervical and uterine cancer such that he will (1) give correct answers on a paper and pencil post-test to (2) 40 or more of the following 50 questions."
"to identify squamous cells in prepared microscope slides (1) in a one-hour laboratory session (2) such that he will select all slides that contain one percent or more of squamous cells when these cells are one percent or more in a mixture of cells."

D. Examples of Objectives

As examples of development of objectives from goals, the following are some objectives which could be developed from the stated goals of the Pap smear symposium referred to on page 3.

**IMMEDIATE GOALS:** To provide physicians with: (a) knowledge regarding the value of Pap smears; (b) skill in the procedure; (c) attitudes conducive to maximal utilization of the procedure.

These goals permit many objectives relating to knowledge, skills and attitudes. An example of an objective relating to the learner's skill is:

**Immediate objective (a statement about the learner's skill):**

**Action:** To interpret

**Content:** lab reports of abnormal Pap smears

**Evidence:** Average of interpretations in agreement with instructor is 15 out of 20 cases. (Note: the evidence specifies average performance by a group, not individual performance.)

**Completed Objective:** The physicians shall learn to interpret lab reports of abnormal Pap smears so that the average of their interpretations in agreement with the instructor is 15 out of 20 cases.

**INTERMEDIATE GOALS:** To increase the frequency of taking Pap smears in routine physical examinations and of mailing reminders to patients to come in for routine Pap smears.
These goals permit objectives relating to (1) increased frequency of taking Pap smears and (2) sending reminders to patients. An example of an objective regarding increased frequency of taking Pap smears is:

**Intermediate objective (a statement about the learner's behavior):**

**Action:** To increase

**Content:** frequency of taking Pap smears in initial physical examinations for women over age thirty

**Evidence:** During the time period from 6 to 12 months after the symposium, the physicians' initial case histories will show a higher percentage of orders for Pap smears than do initial case histories in the time period from the 12th to the 6th month before the symposium.

**Completed objective:** The physician shall increase the frequency with which he orders Pap smears for new female patients over age 30 as shown by higher percentage of Pap smears ordered in the 6 to 12 months after the symposium than in the time period from the 12th to 6th month before the symposium.

**LONG-RANGE GOAL:** The increased frequency of taking Pap smears in routine physical examinations will result in detecting more cases of early cancer of the cervix thereby making early treatment possible.

**Long-range objective (a statement about the effect of the learners' changed behavior on society):**

**Action:** To diagnose

**Content:** cervical or uterine cancer at an earlier stage of development.

**Evidence:** The ratio of early to late detection of cervical or uterine cancer in the geographic region of the RMP shall increase over a 10-year time span.
Completed objective: The ratio of early to late detection of cervical or uterine cancer as shown by a survey of lab reports from the geographic region of the RMP shall show an increase over a 10-year period following the symposium.

The development of a long-range objective from the stated long-range goal is shown as an example of the process. It seems unlikely that the planners of any single effort will be able to show the achievement of long-range goals; it will likely be the responsibility of an RMP to show that its coordinated efforts have achieved long-range goals.

III. Data and Evidence for Evaluation

A. Evidence of Change

The second phase of evaluation is the selection or design of measuring instruments or the design of other procedures to collect data that will lead to evidence for evaluation. Data are the raw numbers; evidence is a summary, or analysis, of the data. The evidence that can be developed depends on the data and the way the data are collected. Objectives of a teaching-learning experience call for change in learner behavior; evidence for the evaluation of objectives is, therefore, evidence of change. The usual way of showing change is to demonstrate by count or by measurement that the learner has more knowledge, more skill or a more desirable attitude at the end of a teaching-learning experience than he had at the beginning. This is the familiar pre-test and post-test procedure.

B. Measurement versus Counting

Two general ways to collect data for evaluation are (1) counting of observable behavior or outcome events and (2) measurement of abstract quantities, e.g., general skill in diagnosis, attitude towards low socio-economic patients. "Counting" has its usual meaning here, but "measurement" has a narrow meaning.
A measure is a score or number assigned to an individual by means of a measuring instrument. In choosing between counting or measurement as data collection methods, the planner commits himself to different procedures, different problems and different possibilities for derived evidence.

C. Floor-and-Ceiling Effect

Whether he counts observable behavior or outcome events or measures abstract quantities, the planner encounters a problem called floor-and-ceiling effect when he attempts to show change. If a learner has a high pre-test score or count, he is near the "ceiling;" the instrument (observation method) cannot show much increase in his score (count). If a learner has a low pre-test score (count), he is near the "floor;" he may by chance increase his score (count) when in fact he learned nothing. Whether he truly learned or increased his score by chance, the learner near the "floor" will appear to have increased his performance more than will the learner near the "ceiling." For example, learner A may have a pre-test score of 95 and a post-test score of 100; he appears to have increased his knowledge 5 percent. Learner B may have a pre-test score of 1 and a post-test score of 6; he appears to have increased his knowledge 600 percent. Both learners increased their scores 5 points, but the learner with the lower pre-test score appears to have increased his knowledge far more. Floor-and-ceiling effect is most obvious in these extreme cases, but it operates for mid-range scores as well.

D. Reliability, Validity and the Measurement of Change

In selecting or designing a measuring instrument, the planner faces two questions:

1. Does the instrument give the same score or number if the same person is measured twice, that is, are the scores reproducible (reliable)?

2. Does the instrument in fact measure the quantity or condition it purports to measure, that is, are the scores valid measures?
If an instrument does not yield reproducible scores, the planner is not sure if different scores for a learner before and after a teaching-learning experience are due to change in the learner or due to error (lack of reliability) in the instrument. Further, if an instrument yields unreproducible (unreliable) scores, there is some doubt about what the instrument actually measures, that is, doubt about its validity. For these reasons, test constructors usually develop tests to be reliable, that is, the tests tend to yield the same score to the same learner at two different times. Because reliable instruments tend to give the same score "before" and "after," it is difficult to demonstrate change in a learner with tests developed to be highly reliable.

IV. Sources of Data for Evaluation of Objectives

A. Introduction

The third phase of evaluation is data collection. Techniques of data collection may be left to experts, but the planner should select sources of data appropriate to his objectives. There are three major sources of data for the evaluation of objectives. First, there is the health professional as participant in a teaching-learning experience. Second, there is the health professional as practitioner. Third, there is society.

The following sections comment on opportunities and problems in drawing data from each of these three sources.

B. The Health Professional as Participant

1. Knowledge

Increase in the health professional's knowledge during the time of a teaching-learning experience is appropriately determined by a paper-and-pencil achievement test administered before (pre-test) and just after (post-test) the experience. If the teaching-learning experience is extended through time via once a month meetings or weekly TV presentations for example, then the planner should provide
evidence that a gain in knowledge from pre-test to post-test is primarily a result of the experience and not due to other sources of information such as journals, the press, or conversation with colleagues.

2. Skills

Increase in the health professional's skill during the time of a teaching-learning experience can be determined in two general ways: (1) sample tasks and (2) expert observation and judgment.

a. sample tasks

For testing simple skills, the planner may be able to design tasks so pass-fail judgment is easy and different judges of the same performance will agree that the physician has "passed" or "failed." The planner cannot be sure that a health professional who does not perform the simple skills in a test situation cannot, in fact, perform them in his routine practice. Also, the planner cannot be sure that a health professional who can perform a number of prerequisite simple skills can assemble them to perform a complex skill.

b. expert observation and judgment

For testing complex skills the planner may decide to rely on expert judgment of the level of the health professional's performance. Two or more judges of any performance may not agree. Agreement among judges is promoted by providing them with observation schedules or itemized checklists or by otherwise organizing and standardizing the observation procedure. Agreement among judges is also promoted by asking them to judge the same performances after which they discuss their judgments.

3. Attitudes

The measurement of attitudes presents unique problems. If the planner asks transparent and naive questions about attitudes, the health professional can detect the socially
desirable response and may give it. If the planner asks covert and subtly worded questions, he cannot be sure that responses relate to the attitude being measured. Further, what the health professional says on a written questionnaire that he feels or would do may not be what he actually feels or does.

C. The Health Professional as Practitioner

Changes in health professional practice approximate the intermediate goals of RMP. Data for evidence of change in practice come from that practice. To know from an achievement test that a health professional has the required skills or knowledge is not to know that he has changed his practice. Collection of evidence of change in practice may continue through time; the planner wants to know if the health professional continues a change in practice after the enthusiasm of the teaching-learning experience wears off.

D. Society

The society, or that part of it within any RMP area, will prove a difficult source of data. The sampling problems will require the attention of expert statisticians. The collection of data on even one of the many aspects of national health in the area of heart disease, cancer, stroke and related diseases over a period of years will require planning by information systems designers. The comparability of data over time and from different reporting units will require attention from epidemiologists and public health officers. Planners of a teaching-learning experience who wish to use data from the society face problems of sampling, data collection and comparability of data over a period of many years before they can complete their evaluation.

V. Analysis of the Data

The fourth phase of evaluation is analysis of the data to produce understandable and pertinent evidence with which to evaluate how well objectives have been met. If objectives were stated completely,
including specification of evidence, then analysis of the data should be straightforward. Experts or statistical clerks may be trusted with performing the analysis, or summary, of the data; but they should not be responsible for interpreting that evidence.

VI. Judgments and Decisions

A. Judging How Well Objectives Were Met

In the fifth phase of evaluation, the planner uses summary evidence to judge how well objectives have been met. The reporting of evidence is not evaluation. The planner may judge from the evidence that objectives have been exceeded, i.e., the change in behavior was more than expected. He may judge that objectives may have been met exactly, or he may judge that objectives may not have been met, i.e., the change in behavior was less than expected. Under any of these three conditions, the planner faces the same kinds of decisions.

B. Decisions

First, the planner faces the decision, "Shall the program be repeated?" This decision will be made jointly with the RMP. If the decision is to offer the experience again, the planner will then face alone the decision,"Shall the program be changed?" If the program is to be changed, the following decisions will have to be made:

1. Should the presentation be modified?

   Decisions have to be made as to who will present the program, when and where to present it, and by what method.

2. Should objectives be modified?

   The planner may judge that the objectives established for the program were not appropriate and need modification. Perhaps with the passage of time new objectives will be
more appropriate due to generation of new knowledge and changing health needs or problems within the Region. Objectives can be modified in any or all of their parts: action, content, or evidence.

3. **Should data collection and analysis procedures be modified?**

   The planner may judge that data collection procedures or the analysis were inadequate or inappropriate for determining how well objectives were met. Planning for the future program may require the collection of different data or a different analysis. These modifications are indicated when, in the judgment of the planner, the data or its analysis did not answer satisfactorily how well the objectives were met.

4. **Should data sources be modified?**

   The planner may judge that the sources of data were not appropriate for answering the question of how well the objectives were met. Decisions have to be made whether to gather data from the health professional as participant, from the health professional as practitioner, or from the society. Again, the decision is based on a judgment of how appropriate the data source is to the objective.

Appendix A provides examples of decisions a planner might make after judging how well objectives were met.

C. **Participants' Reactions**

To help the planner make decisions, he may want participants' reactions to the teaching-learning experience. It may be valuable in planning for next time to know that participants found one part of the presentation "over their heads," another part insulting to their intelligence and still another part of great interest but inadequately covered. Gathering information on participants' reactions is not, however, evaluation of the effect of the teaching-learning experience on the behavior of the participant. Although it may help the planner decide how and where to modify the experience, information about participants' reactions does not provide evidence that objectives have been met, i.e., that specified changes have occurred in the behavior of the learner.
VII. Evaluation Checklist

Public Law 89-239 provides for the establishment of Regional Medical Programs and requires from the Programs reports of activities and justification for continuation. This requirement was interpreted by the Division of Regional Medical Programs to mean evaluation; and a Planning and Evaluation Branch was established to review plans for evaluation contained in each proposal submitted.

This set of guidelines was written to provide a systematic explanation of the process of evaluation to aid in the development of evaluation procedures. At the time a proposal is being prepared for RMP, the general strategy and most of the details of the required evaluation can be planned. An accurate budget for evaluation is impossible until objectives are written, data sources are specified, and measuring instruments or other methods of data collection are selected. The following checklist is provided as a summary and further aid in the planning of evaluation.
CHECKLIST FOR PLANNING EVALUATION

Goals

☐ 1. Are the goals consistent with the goals of RMP? (See page 3)

☐ 2. Are the goals identified as immediate, intermediate or long-range? (See pages 3-4)

Objectives

☐ 3. Are the objectives identified as immediate, intermediate or long-range? (See pages 4, 6-8)

☐ 4. Are the objectives statements about the learner or participants in the program? (See page 5)

☐ 5. Does each objective include action, content and evidence? (See pages 5-6)

☐ 6. Does the action part of each objective state what the learner will be able to do after the learning experience? (See pages 5-7)

☐ 7. Is the content part of each objective stated specifically? (See pages 5-7)

☐ 8. Does the evidence part of each objective specify how the behavior is to be observed or measured? (See pages 5-7)

☐ 9. Does the evidence part of each objective specify what level of performance will be considered evidence that the learner has achieved the behavior? (See pages 5-7)

Data Collection

☐ 10. Does the data collection include assessment of what the learner knows prior to the learning experience so that change can be demonstrated? (See page 8)

☐ 11. Will the data lead to evidence for determining how well the objective has been met? (See pages 8-10)

☐ 12. Has the source of data been defined? (See page 10)

☐ 13. Is the source of data a realistic one; will there be serious difficulties in gathering the desired data? (See pages 10-12)

☐ 14. Have statistical procedures been selected for the analysis of the data to be gathered? (See page 12)
APPENDIX

Examples of Decisions and Modification

I. Immediate Objectives

The first objective as stated on page 6 is: The physician shall learn to interpret lab reports of Pap smears so that the average of their interpretations in agreement with the instructor is 15 out of 20 case histories.

A. Failing to Meet Objective:

The average score of the participating physicians was agreement with the instructor in 10 out of 20 case histories.

1. Possible Decisions

   a. The future program needs better instructional methods such as projection of slides during lecture, more variety of sample lab reports, less ambiguous terminology, instructors who make more interesting presentations.

   b. Modify the action part of the objective from "interpret" to "categorize."

   c. Modify the evidence part of the objective from "agreement with instructor in 15 out of 20 case histories" to "agreement with a previously established panel of pathologists as to which of three categories the lab report best belongs."

B. Meeting the Objective:

The average score of the participating physicians was agreement with the instructor in 15 out of 20 case histories.
1. Possible Decisions.

Similar decisions for modification may be made by the planners who by their judgment believe that modification of the program, objective, data collection and analysis, procedures or data sources will serve better meet the goals of the RMP.

C. Exceeding the Objective:

The average score of the participating physician was agreement with the instructor in 20 out of 20 cases (i.e., everyone had a perfect score).

1. Possible Decisions

a. The judgment of the planners may be that the test materials, i.e., the case histories, was too easy.

b. It is likely that the planners will modify a program in which everyone has attained a perfect score on the test material.

II. Intermediate Objectives

The intermediate objective as stated on page 7 is: The physician shall increase the frequency with which he orders Pap smears for new female patients over age 30 as shown by higher percentage of Pap smears ordered in the 6–12 months after the symposium than in the time period from the 12th to 6th month before the symposium.

A. Failing to Meet Objective

The frequency with which participating physicians order Pap smears decreases or remains the same in the time period specified.

1. Possible Decisions

a. The planner may decide that a more dramatic presentation of the efficacy of Pap smears in diagnosing cancer needs to be offered.
b. The planner may decide that the action part of the objective "to increase" needs to be modified.

c. The content part of the objective may need to be modified from "frequency of taking Pap smears in initial physical examinations for women over age thirty" to "frequency of routine Pap smears for new and present women patients over age thirty."

d. The evidence part of the objective may need to be modified by changing the time period stated.

e. It may be decided that data collection procedures were too difficult to carry out and therefore need to be modified. The analysis of data may need to be modified to include the total number of female patients seen during each time period as the base for comparison of an increase in frequency in taking Pap smears.

f. The data source may not have been adequate and needs to be modified to include physicians who did not participate in the program.

B. Meeting the Objective

The frequency with which the participating physicians ordered Pap smears increased during the time period specified.

1. Possible Decisions

a. The planner may decide that the action part of the objective "to increase" should be stated more specifically such as "increase by at least 25 percent."

b. The planner may decide that the content part of the objective should be modified from "frequency of taking Pap smears in initial physical examinations for women over thirty" to "frequency of routine Pap smears for new and present women patients over age thirty."
c. The planner may decide to modify the evidence part of the objective by changing the time period stated.

d. The planner may decide that the data collection procedures should be modified.

e. The planner may decide that the data source needs to be modified (possibly because of the discovery of more appropriate data available).

C. Exceeding the Objective

The way in which this objective was written it cannot be exceeded; the amount of increase in frequency was not specified.