ED 098 728 EA 006 569

AUTHOR Cogan, Eugene A.
TITLE Decisions about Data Collection Strategies.
INSTITUTION Human Resources Research Organization, Alexandria, Va.
REPORT NO HumRRO-PP-23-69
PUB DATE Jun 69
NOTE 17p.; Paper presented at U.S. Army Operations Research (8th, Durham, North Carolina, 1969)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Data Collection; *Decision Making Skills; *Information Seeking; Mathematical Applications; *Operations Research; Research; Research Methodology; Research Tools; *Systems Analysis; Systems Approach
IDENTIFIERS *Strategies

ABSTRACT

"Pure" academic research rules on data collection do not apply directly to operations research (OR). OR data collection should be viewed in terms of objective, cost, and effectiveness. For the model formulation objective, proper data strategies emphasize multiple views of the operating system to identify the "relatednesses" to be depicted. For the objective of estimating parameters or testing predictions, bias, precision, and level of confidence of results are effectiveness concepts to be balanced against cost. Decision and utility theory, sensitivity analysis, and sequential analysis apply or OR data collection strategies and employ operational parameters to define data needed and, hence, minimize costs. (Author)

Professional Paper 23-69
June 1969

Decisions About Data Collection Strategies

by

Eugene A. Cogan

Presentation at
U.S. Army Operations Research Symposium
Durham, North Carolina May 1969

HumRRO

The George Washington University
HUMAN RESOURCES RESEARCH OFFICE

The Human Resources Research Office is a nongovernmental agency of The George Washington University. HumRRO research for the Department of the Army is conducted under Contract DAHC 19-69-C-0018. HumRRO's mission for the Department of the Army is to conduct research in the fields of training, motivation, and leadership.

Published
June 1969
by
The George Washington University
HUMAN RESOURCES RESEARCH OFFICE
300 North Washington Street
Alexandria, Virginia 22314

Prefatory Note

This paper was prepared for the Eighth Annual U.S. Army Operations Research Symposium by Dr. Eugene A. Cogan of the Director's Office, Human Resources Research Office. The symposium was sponsored by the Office of the Chief of Research and Development, Department of the Army, and planned and managed by the Institute of Systems Analysis, Combat Developments Command and the U.S. Army Research Office-Durham. A report of the Proceedings is being published by the Army Research Office-Durham.

DECISIONS ABOUT DATA COLLECTION STRATEGIES

Eugene A. Cogan

The strong endorsement operations research has received from decision makers is shown by the typical requirement for results of a study to be available yesterday. As a consequence of short fuse, limited resources, and other constraints in the real life existence of the operations researcher, the question of how one devises an efficient data collection strategy is often moot--there is frequently no time for *any* data collection.

To meet a quick response requirement, operations research studies must often be based on data collected incidental to some other study--and such data are rarely, if ever, exactly what is needed. Or, it may be that the requirement can be met only by analyzing data from administrative or other operational sources that happen to include seemingly relevant information--and such data are of unknown (or unknowable) accuracy. In other instances, analysis must be performed with none but the scratchiest data base or, perhaps, no data base.

The unhappy circumstances that preclude developing hard data for a study are those in which time and cost constraints are so severe that it is simply not sensible to give thought to the parameters of decisions about data collection strategies. Since the framework for such highly constrained studies is normally one in which the decision maker has a humble request--something a bit better than he can do without operations research input--such studies do serve a useful purpose. But, instead of considering data collection strategies for these, one might be more inclined to consider tranquilizers to ward off the debilitating effects of frustration in the researcher.

Leaving aside the very highly constrained studies, there are studies for which requirements are more permissive, where it *is* feasible to collect data bearing on a problem and still provide a solution within time and cost constraints. For these studies, it is meaningful to consider data collection strategies.

The traditional view of the scientific method, as seen from the university citadels, emphasizes the pursuit of truth, perhaps spelled with full capital letters. This pursuit is a goal separate from time frame or cost or any other parameter measured by practical, operational units. To the "pure" scientist, concepts such as cost, or time, or even the precise purpose of a study receive scant formal attention. At best, operations researchers deal with very specific objectives, externally imposed time frames, limitations on facilities available, and limited funding; therefore, additional factors must be superimposed on the scientific method for it to be applicable for the operations researcher--or for anyone else engaged in mission-oriented research.

A cue to adapting academic rules of research to the practical world comes from the rationale of operations research itself; the fundamental concepts the operations researcher applies to an operational problem are applicable to the practical aspects of conducting OR. First, in approaching an operational problem, the researcher will ask the question, "What is the objective, or the mission, or the purpose of the operational activity?" This question fits for his research method as well. Second, alternate strategies for accomplishing the objective will be considered in terms of parameters of cost in dollars and time--and these apply too for research methods. Third, the degree to which the objective is achieved is considered in terms of the concept of effectiveness or benefit--questions appropriate to research method.

Since every study has its own unique characteristics, treating of the "typical" study or "typical range" of studies risks treating of nothing that exists in the real world. Nevertheless, some general implications of an operations research approach to operations research data collection strategies can provide useful guidelines with a wide range of relevance. This paper was prepared toward deriving such guidelines.

Objective of Data Collection

Data collection for operations research and, in fact, any scientific study is simply a mechanism for developing information pertinent to a problem. For OR--just as in the general scientific domain--it is useful to distinguish two kinds of information gathering purposes. Hans Reichenbach[1] proposed terminology for these in traditional scientific enterprise. One he called the "Context of Discovery" and the other, the "Context of Justification". These provide a useful beginning point for considering OR data collection but require some modification and interpretation for OR purposes.

Reichenbach spoke of the "Context of Discovery" to formalize a vital and often overlooked feature in scientific research. Simple textbooks on scientific method begin with a theory or a model or hypothesis and then treat the research mechanisms used to "justify" or verify hypotheses as if this were the total research process. In real life, however, the researcher does not find ready-made propositions to be tested and, in fact, the most vital, creative, and human portion of science deals with the process of developing hypotheses to be "justified." Such hypotheses may come intuitively or inspirationally, as with myths surrounding the source of Archimedes' principle of displacement of mass in a fluid medium, but more often the hypothesis is a product of carefully digested, rearranged, massaged, and well worked data.

Translating "discovery" to the context and terminology of operations research, transforms "discovery" to formulating a model of an operational phenomenon so that specific study can be undertaken; model formulation represents the creative heart of OR. The rules for

[1]Reichenbach, Hans, *Experience and Prediction*, University of Chicago, Chicago, 1938.

developing a model are based on immersion in the facts, mechanisms, and workings of operations to develop what psychologists call the *gestalt* or form of arrangements. From this form, an explicit mathematical model is formulated. Model formulation can receive support from well-conceived data collection strategies whose purposes are to provide a broad and flexible view of operations.

For model formulation, the major ingredients of relevant data are oriented toward developing a *picture* of operations.

1. What are the crucial and measurable characteristics of input and output for operations?

2. What are the measurable mediating variables relevant to input-output relationships?

3. What are the measurable relationships between input, mediating, and output variables?

For "discovery" or model formulation, data collection strategies should be geared to maximize the number of alternate views of "what's going on" that are available to the researcher so that he can formulate a meaningful and useful model. "What's going on" for OR must be represented in a single picture; ideally, this picture will provide the framework for working with the operating system towards solving the practical problem.

A host of data collection strategies should be employed to provide the meat for problem formulation. Emphasis should be placed on the term "host" because the model eventually emerging will inevitably be from a particular perspective and the OR worker should have exposure to as many perspectives as possible before fixing on one to be used in solving the problem. For "discovery", "hard" and formal data are not critical nor is it usually economical to invest heavily in data collection because the OR problem has not yet been defined. Gaining views of operations from many perspectives is the critical element of data collection strategy.

What does the operation seem to be like from the point of view of the manager? Data collection for this picture can be accumulated by reading formal documents on the mission and characteristics of the operation, exploring rules and regulations on the flow of activities in the operation, and extensive interviewing of those having immediate management responsibilities for the operation, with those at an echelon higher than the immediate manager and with those in lower-ranking supervisory jobs.

What is the operation like from the perspective of those performing the system functions? To gain a picture of this perspective, it is desirable to observe and interview at length people representing each of the different kinds of jobs or functions in the operation system. Where time allows, and the activity is dispersed, questionnaire or extensive field observation and interview approaches may be most useful.

What are the input-output factors? Here, data can come as byproducts of interviews and observations discussed earlier, but also a picture can emerge through careful study of "raw materials" entering the operational system (be they information or be they materiel for manufacturing processes), and finished work leaving the operational system (be they decisions or machined fittings). Data in the form of requisitions or invoices and inventory or shipping lists or messages received and messages emitted can be invaluable in helping to formulate a picture of input and output.

What are the mediating mechanisms and their arrangements? Explorations on the inner workings can come from study of the jobs that people perform. These can be rendered into flow charts or other descriptive vehicles and can draw upon SOPs, organization charts, job descriptions, and personal observation. The key aspect of the system study at this point concerns the details of inner workings whereby input is transformed into output.

With which other operational activities must the system being studied interact? Here, it is necessary to pay attention to how a single operating system at issue fits together with a larger system to accomplish an overall mission. While the chain from system to system or the interface of one system with others can become an almost infinite regress, some attention must be paid to crucial couplings so that an OR solution can take these into account. Attention must be paid to how an operating system being studied articulates with other systems it works with.

Taking data from many sources, including whatever may be available in the formal operations research literature or earlier studies, the operations researcher must put everything together to formulate a mathematical model. No mathematical model can ever represent all aspects of anything nor is there a measuring instrument for goodness of model other than the judgment of experienced people--at least until later stages, when predictions about the effects of modifications in a system can be tested against actual occurrence.

The two key points on data for "discovery" or model formulation are: first, it is *patterns*, relatedness, structure, or form that is being sought. Second, attention is directed at what is actually going on in the real operating system from as many perspectives as possible. Out of such study and its product--the model--will emerge a precise definition of system output from which measures of effectiveness can be derived as well as a picture of "what's really going on." Frequently, these aspects of the system are better understood by operating personnel after operations research formulation even though the basis of the model is the operating personnel's ideas and experiences.

The second objective served by scientific information--"Context of Justification"--differs from "discovery" in that its reference is not directly to an operational system but rather to the model formulated as a result of the first phase of operations research study. The classical paradigm for "justification" research is that of formal hypothesis testing, the hypothesis being derived from a theory or model and representing a test or justification of the validity of the theory. Translating this concept to operations research requires broadening the notion to include estimating parameters as well as testing predictions from the model. It is crucial to recognize that while the overall objective and purpose of an OR study is neither more nor less than improvement in an operating system, test and parameter estimation are of the mathematical model rather than operations directly. This must be true because the operations research or any other model is an idealized presentation of phenomena and cannot be a mirror of the chaotic and complex real world.

Models exist in many degrees and many qualities. At one extreme are those that are far from rigorous mathematical representations but rather are in the form of an informal schematic representation of "what's going on." At the other extreme are sophisticated and highly formalized mathematical representations with clearly identified parameters and ways to assess these parameters. The nature of data needed for "justifying" a model is, of course, highly dependent upon the nature of the model. Perhaps the best guideline for kind and quality of data vis-a-vis model rests in Gilbert and Sullivan's *Mikado*--let the characteristics of data fit the kind of model. With a model that simply provides for an inverse relationship of *some* sort between two variables, measurements to 5 decimal places are unlikely to be needed; with a model whose quantification aspects are thoroughly developed, measurements of acceptable precision and reliability must be used.

As a first step in planning to use a model for operations research, some activities need to be programed towards validating the representation or assessing the model for relevance to the phenomena being studied. That is, regardless of how exotic the model may be and regardless of how sophisticated its mathematical characteristics, unless it represents the operations, the results derived from the model are worse than useless because they can mislead decision makers. Data to evaluate the model at this stage are best derived from operational experts--rather than from operations research experts. There *is* a danger that a researcher will formulate a model for operations that does *not* represent the phenomena of interest to the decision maker. An instance of this sort comes from my personal experience. An algorithm was being derived to treat "suppressive fire." Unfortunately, it came out in the mathematics not as suppressive fire but as an exchange of fire. That is, the notion that bursts of fire can inhibit enemy fire and, thus, assure at least momentary protection for exposed friendly troops was ignored and line of sight visibility and hit probability parameters were applied under the assumption that normal fire would occur whenever a target is available, without regard for suppression of fire. The operations research model producer mistranslated suppressive fire into the dynamics of a simple duel. A more appropriate algorithm came from formulating an algorithm to represent total suppression followed by gradual recovery.

A safeguard for translation rests in detailed and thorough communication between the operations researcher and operational personnel. The model must be characterized by its operations research proponent in language the operational expert can understand, and the latter must learn to bridge the gulf between the phenomenon as he understands it and the phenomena as described by the operations research model; the operational expert can then provide data for the operations research specialist on whether the model makes sense.

Cost of Data Collection

The concept of cost, including time, as applied to data collection comprises a set of budget categories, schedules, and other well known facts of research life. There are, however, some factors warranting special attention since they may be overlooked.

First, costs of research are normally viewed in terms of costs to the research budget. While for some purposes this is appropriate, the operations researcher planning a study should not ignore costs of the research to the operating system. Occasionally, costs of that sort (e.g., time of subjects interviewed) are actually charged to a research budget. More often, they enter the picture as a feature in the negotiations between the client system and the OR team to establish the amount and kind of facilities and support the system can provide for the research. How these costs should be taken into account will vary from situation to situation.

Second, and more pertinent to the research process itself, are concepts of over- and under-"kill" that can have major impact on cost. For any given problem, and as a consequence mainly of the nature of the mathematical model, research information requirements are for a particular degree of precision. Over-"kill", or precision beyond needs, is wasteful because costs of research data collection depend on the amount of data collected and the care taken for each measurement--which translate to costs. On the other hand, a data collection enterprise producing data too imprecise for the specifications of the model may represent a waste of resources. Often, choice should be between using existing data or gathering data that will serve what is needed for the model. If a plan for gathering new data is subject to too much attenuation from what is needed, cost consciousness may be better served by cancelling data collection plans and "making do" with what is available.

Effectiveness of Data Collection

Assuming there has been expert opinion validation of the model and further assuming that dimensions of data have been conceived to match the kind of model available, next steps have to do with the technical characteristics of data. In general, the notion of accuracy fits as a way to think of data, but that concept is subject to analysis into several components--bias, precision, and level of confidence.

Bias. Collecting data either for estimating a parameter or for testing a prediction from a model requires careful attention to defining the population domain of the system being studied. The researcher must recognize he can never sample exactly the operational domain that is relevant, but explicit and knowledgeable decisions rather than unmoderated expediency should guide what is--or must be--ignored and what is--or must be--represented.

For the absolute purist, the population or domain of interest in estimating parameters is "tomorrow's" system since his recommendations or input can have no effect on what has already occurred. Since research or other technology has not yet developed to a point where we can sample "tomorrow" today, a perfect solution is, on the face of it, impossible. If time-related change is evident in a system, one can extrapolate from a set of time-related measures to the future, but this procedure is imperfect, resting as it must on projecting the form of relationship.

Less metaphysical than stopping time to allow study to take place, are a complex of factors or points of view that concern defining a system and then a sampling strategy. For each point of view, a decision must be made about whether that aspect can be ignored without risk of serious distortion or whether that aspect is so important it must be represented.

First, if a system exists in multiple copies, each with its own unit manager, each with its own variations in operations, each with its own constellation of personnel, each with a given size, and each with its own character, data collection strategies must consider how or whether individual units are to be represented. Here, judgment should be based on "taken altogether, are the variations from unit to unit within tolerances for the data needed for the model or are they out of tolerance?" While a critical decision of this sort can sometimes be aided by pilot data collection, elements of subjective judgment will necessarily be crucial.

Second, an operating unit is likely--at one time or another--to be serving its function in a range of the missions of its present organization. Here, one might select one or more key mission contexts as the "pay-off" kind and ignore others, or, one might schedule data collection carefully in order to sample a range of mission contexts, perhaps by extending data collection over a long period of time.

Third, ecological or neighboring system states or other environmental factors may have strong effects on the operation of a system. These factors must be considered and either ignored by decision or taken into account by design.

It is tempting to try to plan a study so as to randomly represent any or everything and do so with a large number of observations. In that fashion, one can rest easily with confidence in the results being based on the Gaussian distribution and the laws pertaining to that distribution. However, in the real world, facilities are not available for such studies. Instead, the researcher must make a series of decisions about many potentially biasing factors and rest on many "negligible effects" assumptions. In practice, all too often the decisions are made--and a decision *is* made whenever a data collection plan is formulated--but the researcher acts as if he is unaware he is making them. Worse, he may *be* unaware what decisions he is making. Bias from sample selection in data collection is a very serious matter since its direction and magnitude are rarely known. That it will exist must be accepted, but the researcher can manage the bias most effectively if he carefully considers a broad range of dimensions for the domain he is studying and, after careful consideration, makes a series of knowledgeable guestimates of which ones he will consider to be negligible. Ideally, the rationale section in the report on a study should include clear statement and discussion of "considered to be negligible" assumptions.

With the rationale of sampling decisions as an overt and explicit part of study design, it is possible to explore the meaning and ramifications of alternate data collection strategies with the decision-maker. It is also possible to employ efficient stratified sampling techniques and *it may* even be possible to convince an operating system manager and a research budget manager that more resources should be expended for a study than they originally intended.

Precision. The factors underlying precision in a set of data are shown quite simply in the formula for the standard error of a mean. Precision is inversely related to the standard deviation of the observations and directly related to the size of the sample. Some contributors to the size of a standard deviation are variations in the "real world" while others are a consequence of errors of measurement that can be reduced by more refined measurement techniques.

For the researcher, there is the prime question of what precision is needed in order to avoid either over- or under-"kill". After that decision is made, his data collection strategy can be adjusted in one of two ways: first, he can increase or decrease the number of observations to yield a particular level of precision with a given measuring technique, *or*, second, he can refine or coarsen his data instrument so that for a given number of observations a particular precision emerges. This trade-off is a fundamental factor in the cost and effectiveness of data collection.

There is no general rule regarding whether it is best to adjust number or quality of observations in an operations research study. The economics of trade-off are too finely interwoven with the specifics of a particular study to allow simple generalizations.

It is, however, pertinent to point out that costs related to observations and costs related to refinement of measurements show sharp discontinuities and a "best balance" as to cost-effectiveness solution may require careful analysis. That is, tripling refinement of measurement is unlikely to cost three-halves as much as doubling refinement--it may cost more, or it may cost less. Tripling observation is unlikely to cost three-halves as much as doubling them--it usually costs less than that, but it may cost more. Similar discontinuities exist for reducing observations and refinement of observations.

It must be kept in mind that concepts of refinement or coarsening of measurements are based on the assumption that no systematic biases creep in to coarser measurement. The least level of coarseness in measurement assumed includes lack of unknown systematic distortion in the data.

There are two techniques familiar to operations researchers that can be applied to guide precision aspects of data collection strategies by helping define and then applying specifications for precision in the experiment. Either can be used alone, and maximum gain may be expected by combining them.

The first of these--sensitivity analysis--is best known as a technique in war gaming or other simulation study and has been used typically after the fact of data collection or after some other mechanism for estimating a parameter has been employed. In sensitivity analysis, one, essentially, plays devil's advocate to ascertain the effects on the outcome that an alternative value of a parameter would have. The purpose served by sensitivity analysis translates quite directly to determining how much precision is needed.

While a full-blown series of simulations to ascertain needed precision in data collection would only rarely--if ever--be economically possible or cost-effective, some partial runs or analytic explorations toward estimating sensitivity and, hence, defining needed precision can be invaluable as well as practical.

Once a sensitivity analysis has helped define needed precision, these specifications can, in turn, be profitably applied to data collection by allowing one to use a sequential analysis experimental design.[1]

Normally, sequential analysis as a technique for the operations researcher's use is reserved to application in quality control systems. However, with specifications for needed precision available, the same kinds of possible economies in amount of data collected can be achieved in an experiment to estimate values for an operations research study having no special relevance to a quality control problem.

Level of Confidence. Closely related to precision--and, in fact, the inverse of precision for a given set of data--is level of confidence or probability level for fiducial limits or, in its commonest inverted form, significance level.

Most of us have been introduced to these concepts in the very conventional mold that has evolved for "pure science" in which almost mystical qualities are attributed to $p<.05$. It would be very interesting to trace how $p<.05$ evolved as a minimum standard for "pure science" and to review the justification for that convention, but that matter goes beyond the scope of this paper.

With regard to level of confidence, the operations researcher is caught in a dilemma and typically solves it by either joining or ignoring the purist's $p<.05$, but there is an intermediate solution. Whatever the history or rationale for the mystic $p<.05$ in the community of pure science, it does not sensibly translate uncritically to OR. For "pure science" there are two factors: first, a methodological conservatism to accept a fact as true, leading to a criterion of $p<.05$ for acceptance--short of that criterion it is ignored; second, a presumption that a datum will be a timeless building block for the edifice of science--here, too, admission requirements may properly be severe. However, neither factor fits for operations research because the risks and values relevant for OR come from an operating system and the needs of a decision maker.

Rather than using $p<.05$ or ignoring the whole question, an operations researcher should give very careful attention to comparative risks or costs of overestimating or underestimating a parameter or accepting or rejecting a proposition. In addition, the circumstances and needs of the decision maker yield input to set a level of confidence. Taken altogether, it would be the unlikeliest event that $p<.05$ is the proper level for an OR study. A frequent alternative--ignoring the whole question--is no more satisfactory since circumstances should dictate what confidence level is needed.

While discussion has tended to suggest a value higher than $p<.05$ for level of significance may be appropriate in OR, this is not necessarily true, and some OR studies may properly be designed for an almost infinitesimal range of uncertainty.

Developments in statistical theory past those of classical Gaussian or Fisherian formulations provide a basis for something better than

---

[1]Wald, Abraham, *Sequential Analysis*, J. Wiley & Sons, Inc., New York, 1947.

adherence to or rejection of usual conventional levels of significance. These fit under the general designation of Decision Theory or Utility Theory.[1] By careful analysis of the operational consequences of alternative and uncertain decisions, and considering "gains" and "losses" associated with the ramifications of each alternative (including being right and being wrong), one can deduce an appropriate level of significance for an operations research study. It is in this way that operations researchers have clear advantage over the academician researcher who is pursuing the abstract goal of adding to basic knowledge. The operations researcher works in a practical context and consequences of being right or wrong can be assessed; such assessment for purely academic pursuits is probably impossible.

CONCLUSIONS

Guidelines for Data Collection Strategies

1. Applying basic concepts of the operations research approach to problems in data collection for OR is fruitful; costs, constraints, effectiveness and analysis of the objectives in data collection provide a useful frame of reference to guide choices in data collection strategies.

2. Two phases of operations research study should be carefully distinguished: model formulation and parameter estimating or validation of the model. For model formulation, focus is on the operating system itself. Emphasis should be placed on data directly from the operating system and from many perspectives on that system. The system should be studied in order to identify patterns, forms and relationships to provide an experience base for formulating the model. Data assessing the model from the perspective of the system manager are crucial for "validating" the model before formal test is undertaken.

Parameter estimation and test of the model are best viewed as focusing on the model although data are taken from the operating system. These data must be carefully considered with respect to needed accuracy. Avoiding data overly accurate for the nature of the model represents a saving; gathering data with insufficient accuracy can be catastrophic and risks wasting the entire data gathering investment. Accuracy should be considered from its several aspects: bias, precision, and level of confidence.

3. For a given level of precision, there is a trade-off between number of observations and refinement of measurements. The rules of cost-gain for the two factors in trade-off are specific to a given research situation. In general, relating costs to either number of observations or to degree of refinement is likely to show complex patterns with discontinuities. Hence, alternate dollar investment for data collection of a given quality is likely to be defined by alternate and separate patterns, rather than simple intersections of two cost-gain curves.

---

[1] Chernoff, Herman, and Moses, Lincoln E., *Elementary Decision Theory*, John Wiley & Sons, Inc., New York, 1959.

4. Modern decision theory offers alternatives to the conventional significance level of $p<.05$. Each OR problem has characteristics that can be used to infer an appropriate level of significance or confidence for that study; the appropriate level may be much lower or much higher than conventional usage in "pure science." We can only be certain that it is most unlikely that $p<.05$ (or 95% confidence level) is appropriate for an OR study--the rationale for the convention in pure science bears little if any relevance for what is needed for a solution to an operational problem.

5. Sensitivity analysis--a well known technique in the OR repertoire--should be used *before* data collection. This technique can help define needed accuracy, as input in designing data collection strategies.

6. Where the nature of data collection allows, sequential analysis can be used as an automatic cost-minimizing approach. Applying sequential analysis to problems in "pure science" is often frustrated because accuracy requirements cannot be set. However, for an operational problem in an actual system, tolerances *are* amenable to definition.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Human Resources Research Office The George Washington University Alexandria, Virginia 22314	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

DECISIONS ABOUT DATA COLLECTION STRATEGIES

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Professional Paper

5. AUTHOR(S) (First name, middle initial, last name)

Eugene A. Cogan

6. REPORT DATE June 1969	7a. TOTAL NO. OF PAGES 14	7b. NO. OF REFS 3
8a. CONTRACT OR GRANT NO. DAHC 19-69-C-0018 b. PROJECT NO. 2Q062107A712 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Professional Paper 23-69 9b. OTHER REPORT NO.(S) (Any other numbers that may be assigned this report)	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES Presentation at U.S. Army Operations Research Symposium, Durham, No. Caro., May 1969	12. SPONSORING MILITARY ACTIVITY Office, Chief of Research and Development Department of the Army Washington, D.C. 20310

13. ABSTRACT

"Pure" academic research rules on data collection do not apply directly to operations research. OR data collection should be viewed in terms of objective, cost, and effectiveness. For the model formulation objective, proper data strategies emphasize multiple views of the operating system to identify the "relatednesses" to be depicted. For the objective of estimating parameters or testing predictions, bias, precision, and level of confidence of results are effectiveness concepts to be balanced against cost. Decision and utility theory, sensitivity analysis, and sequential analysis apply to OR data collection strategies and employ operational parameters to define data needed and, hence, minimize costs.

DD FORM 1 NOV 65 1473

Unclassified

Security Classification

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Data Collection						
Data Strategies						
Operating Systems						
Operations Research						

Unclassified

Security Classification