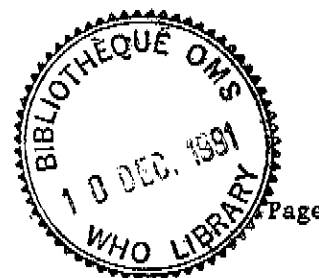




New Haven, USA, 26-30 August 1991

REPORT ON AN INFORMAL CONSULTATION ON ADVANCED  
METHODS FOR SOCIOECONOMIC RESEARCH ON TROPICAL DISEASES

Table of Contents



I.	OBJECTIVES.....	2
II.	SUMMARY OF OBJECTIVES.....	2
III.	EXPLANATORY MODEL INTERVIEW CATALOGUE (EMIC).....	2
IV.	ANTHROPAC.....	6
V.	MICROSIMULATION MODELS.....	9
VI.	GRADE OF MEMBERSHIP (GOM).....	11
VII.	GENERALIZED AUTOMATIC TEXT ORGANIZATION AND RETRIEVAL SYSTEM (GATOR).....	13
VIII.	HEALTH RESOURCE ALLOCATION GAME/MODEL (HRAM).....	16
	LIST OF PARTICIPANTS.....	19

This report contains the collective views of an international group of experts convened by the UNDP/WORLD BANK/WHO SPECIAL PROGRAMME FOR RESEARCH AND TRAINING IN TROPICAL DISEASES (TDR). It does not necessarily reflect the views of TDR/WHO. In the interests of rapid communication it has been submitted to only minimal editorial revision. Moreover, any geographical designations used in the report do not imply the expression of any opinion whatsoever on the part of TDR or WHO concerning the legal status of any country, territory, city or area or of its authorities concerning the delimitation of its frontiers or boundaries.

Ce rapport exprime les vues collectives d'un groupe international d'experts réuni par le PROGRAMME SPECIAL PNUD/BANQUE MONDIALE/OMS DE RECHERCHE ET DE FORMATION CONCERNANT LES MALADIES TROPICALES (TDR). Il ne représente pas nécessairement les vues du TDR/OMS et, en vue d'une diffusion accélérée, il n'a pas été l'objet d'une mise en forme particulièrement soignée. En outre, les noms géographiques utilisés dans le présent rapport n'impliquent, de la part du TDR ou de l'OMS, aucune prise de position quant au statut juridique de tel ou tel pays, territoire, ville ou zone, ou de ses autorités, ni quant au tracé de ses frontières.

## ADVANCED METHODS FOR SOCIOECONOMIC RESEARCH ON TROPICAL DISEASES

### OBJECTIVES

The objectives of the informal consultation on advanced methods for socioeconomic research on tropical diseases were as follows:

- (1) To demonstrate several new social science methods, as well as methods which had never been applied to the TDR target diseases, and to evaluate their application within tropical disease research;
- (2) to initiate plans for the use of the relevant methods in SER research; and
- (3) to facilitate the initiation of collaborative research between developed and developing country participants.

### SUMMARY OF ACTIVITIES

The methods discussed included both qualitative and quantitative methods for data collection and analysis: the Explanatory Model Interview Catalogue (EMIC) for Cultural Research, a method for both collecting and analysing data; ANTHROPAC for the analysis of anthropological data; Microsimulation for the design of field studies; Grade of Membership Models and their applications; and GATOR, a text retrieval and organizational database system. In addition, the group was introduced to the Health Resource Allocation Game (HRAG), a model of particular use to health planners, and still in need of further development and refinement. Finally, presentations were heard on rapid assessment methods which are under development for use in tropical disease research, as well as a critique of these methods from the perspective of traditional anthropology. A visit was also included to the Hill Health Centre, which provides community health and development services using an integrated approach to a poor and underprivileged population of New Haven.

In the following sections of the report, each method is briefly presented and evaluated, corresponding to discussions at the meeting. The brief description of the methods were provided by the presenters of each method (see agenda) and the evaluations were prepared in small group sessions. References are given for the use of those wishing further details on relevant bibliography and computer packages, and where they can be obtained. Some key papers are available from TDR/SER, but the packages must be ordered directly from the distributors.

### EXPLANATORY MODEL INTERVIEW CATALOGUE (EMIC) FOR CULTURAL RESEARCH

#### DESCRIPTION

Ethnographic methods of anthropology characterize the culture of communities, and quantitative methods of epidemiology clarify features of the disease burden of populations. Efforts to develop clinical ethnographic research integrating anthropological and epidemiological aims and methods have produced the Explanatory Model Interview Catalogue (EMIC), making the study of explanatory models operational, both for quantitative and qualitative research to study illness-related perceptions, beliefs and practices.

The EMIC, an acronym that indicates that the aim of the method is to clarify local experience in its own terms, was developed to clarify the impact of culture on illness and to determine how culture relates to clinical and public health problems. By identifying patterns of distress, perceived causes, preferences for help seeking and treatment, and the context of general illness beliefs, the EMIC characterizes normative explanatory models of a community and their range of variation within that community. By comparing components of these explanatory models with illness behaviour and clinically significant outcomes, research not only clarifies local cultural features of illness but also identifies their practical implications.

Experience studying psychiatric and medical patients in India motivated development of techniques to identify operational components of explanatory models for complementary quantitative and qualitative study. Use of EMIC in a cultural study of leprosy, mental health and treatment compliance contributed to the development of the method and demonstrated its value for studying the gold standard of stigmatized conditions. Leprosy is an especially appropriate health problem for considering the inter-relationship of culture, mental health and medical illness because of deeply rooted cultural meanings, a significant emotional burden and underutilization of effective therapy. In our study fifty percent of recently diagnosed leprosy outpatients, 37% of controls with another stigmatized dermatological condition (vitiligo), but only 8% of controls with a comparable non-stigmatized skin disorder (tinea versicolor) met psychiatric diagnostic criteria (DSM-III-R) for emotional disorders. Patients participating in the study appeared to benefit from psychosocial supports the study afforded, and they had significantly better clinic attendance than comparable patients who were not invited to participate in the study.

Data collected with an appropriate version of the EMIC, adapted for use in a particular culture to study a particular health problem, include both quantifiable data coded during the interview according to precoded categories, as well as qualitative prose elaboration of these categorical data and clinical ethnographic notes. A programme in a relational database system (DataEase) has been linked by designated categories to a prose-oriented database (Notebook). Categorical and numeric data from DataEase (or other database systems, such as dBase, Paradox, etc.) can be directly transported for statistical analysis by means of a conversion utility (DBMS/COPY or DBMS/COPY Plus) to standard statistical software programmes (e.g. SAS, SPSS, Systat, BMDP, etc.). Quantitative analysis identifies relationships between components of explanatory models, constituting independent variables and outcomes. For example, humor beliefs about cause, consistent with the Ayurvedic medical tradition, predict better clinic attendance for treatment.

Quantitative data may also be used to set conditions that identify subsets of the sample (e.g. women over 30 years-of-age who think less of themselves and attribute their problems to karma). Searching specific prose fields or the entire record in Notebook can also identify subsets of the sample based on searches of one or more text strings linked by Boolean operators. Topical prose fields (or entire records) for this subset (or the entire sample) can be printed to disk or hard copy for qualitative analysis that clarifies the meaning of quantitative findings or provides information independent of quantitative findings. Prose elaboration may identify subcategories embedded in the codes. By clarifying the meaning of statistical associations,

qualitative data also indicate how to formulate appropriate strategies for health education, interventions, health programmes, and clinical management.

Analytical techniques and a research agenda discussed in subsequent sessions of the report complement methods embodied in the EMIC. ANTHROPAC enables one to identify recurring patterns among component groups of variables constituting explanatory models (viz., patterns of distress, perceived causes, and preferences for help seeking and treatment) by employing multi-dimensional scaling and cluster analysis. Patterns among these groups of variables specify more complex explanatory models, which may be analyzed to identify relationships between them (not just explanatory model components) and outcomes of practical significance. GATOR makes it possible to organize free-form prose data, including clinical ethnographic notes, according to categories the interview studies and themes the researcher may identify in the course of analyzing data. Developers of the Health Resources Allocation Game (HRAG) recognize the importance of identifying cultural variables that influence the utilization of health services. It may be possible to incorporate variables from research with the EMIC to enhance the sophistication of the model that underlies the HRAG and to guide research that demonstrates the relevance of culture for health planning.

#### References

1. Weiss, M.G., Doongaji, D.R., Siddharth, S., Wypij, D., Pathare S., Bhatawdekar, M., Bhave, A., Sheth, A., Fernandes, R. The explanatory model interview for classification (EMIC): method for cross-cultural research developed in a study of leprosy and mental health. Br. J. Psychiatry, accepted for publication (available from C. Vlassoff).
2. Weiss, M.G., Leprosy and Mental Health. Paper presented at Third International Conference on Traditional Asian Medicine in a panel entitled, Cultural Beliefs and Mental Health. Bombay, 6 January 1990 (available from C. Vlassoff).
3. Choprapawon, C., Chunsutiwat, S., Kachondham Y., Weiss, M.G. Cultural study of diarrhoeal illness in two villages of Central Thailand. J. Diarrhoeal Disease Research, September 1991.

Weiss, M.G. Cultural models of diarrhoeal illness: conceptual framework and review. Soc. Sci. Med., 1988; 27(1): 5-16

### EVALUATION OF EMIC

#### INTRODUCTION:

The purpose of EMIC is to obtain an understanding of illness and health-seeking behaviour from the point of view of a person within his/her own culture. The basis of method is clinical ethnography. Its development began by considering Kleinman's eight major questions. The idea of the explanatory models was then operationalized, identifying four main areas of concern: patterns of distress, perceived causes, health seeking and treatment and general illness beliefs. The method needs to be embedded within a wider research project and not used in isolation.

STEPS IN METHOD:

- (1) Series of in-depth (unstructured) pilot interviews with patients to investigate the four main areas.
- (2) Development of a structured interview schedule using questions and pre-coded responses derived from in-depth interviews (including standard questions where appropriate) so that the interview is coding driven.
- (3) In addition to the coded responses for certain questions in the structured interviews, open-ended responses are recorded verbatim, plus comments or reactions. A second person (rater/recorder) also records open-ended responses (or a tape recorder is used).
- (4) Two data files are used for data recording and management:
  - (i) Data base file, e.g. "Data Ease", for numerical data from pre-coded questions.
  - (ii) Prose file, using word processing package, accessible by a retrieval system, e.g. Notebook.

QUESTIONS

1. In what contexts is EMIC useful for TDR/SER projects?

EMIC is useful for exploring many questions related to human behaviour, why it occurs and for understanding behavioural differences. The specific ways in which it is useful for SER/TDR are:

- (a) It is applicable to all of TDR diseases (e.g. in schistosomiasis, questions concerning perceptions of importance/severity of infection and health seeking beliefs and practices). This information can be used to design interventions.
- (b) To explain why an intervention (e.g. "bednets" in a community) has been successful/unsuccessful. It can be used as part of an evaluation, or monitoring scheme.

2. What are the strengths and limitations of the method from TDR/SER's perspective?

Strengths

- (a) Combines advantages of both quantitative and qualitative methods, i.e. better quality data plus generalizability.
- (b) Makes handling of qualitative data easier as it is more organized (in smaller pieces) than unstructured interviews.
- (c) Understandable and more attractive than in-depth interviews to researchers not well acquainted with qualitative research.
- (d) Helps interpretations of the results (by ability to go back into qualitative data and by moving between qualitative and quantitative data).

Limitations

- (a) More time consuming and costly than structured interviews or focus groups.

- (b) Needs more highly skilled interviewers (i.e. skilled in qualitative data collection) than normal structured interviews.
- (c) Need for long period of training of interviewers.
- (d) Not fully validated in comparison with in-depth interviews.
- (e) Pre-coded questions may be suggestive of knowledge whereas unstructured interviews are not.

3. What is needed to make it usable by new researchers?

- (a) Clear description of the method and its different components, methods and procedures.
- (b) Methodological development related to specific SER questions.
- (c) A multi-disciplinary research team involving at least a medical professional and a social scientist from a qualitative discipline.

4. What action should be taken by SER?

- 1. Facilitate 3 (a).
- 2. SER/SC to consider promoting use of the method in some of the following ways:
  - (a) Encourage a few studies using EMIC to answer questions of strong interest to SER. Studies need not be on the same subject.
  - (b) As in (a) but all studies on same subject with comparison of results as one of the objectives. In (a) + (b) the researchers do the methodological development and document it in conjunction with someone with experience in EMIC, who can help in the formulation of questions and adaptation to the specific context.
  - (c) SER commissions/encourages methodological development and later sets up studies using developed methodology.
  - (d) In general, SER/SC should encourage more research using a mix between qualitative and quantitative methods.

ANTHROPAC\*

DESCRIPTION

The ANTHROPAC software programme is especially suited for exploratory pattern analysis concerning cultural and behavioural systems. This type of exploratory data analysis is appropriate in the first stage of planning of an intervention programme or in a pilot study for a more extensive (quantified) research project. More specifically, the software is designed to help in generating systematic information concerning the cultural vocabularies of specific topical domains such as "types of illnesses," "treatments for illness", "symptoms", "foods suitable for infants", and all the other topical areas about which health researchers need concrete information in order to understand local cultural beliefs and behaviours.

---

\* ANTHROPAC is a trademark copyrighted by Dr Stephen P. Borgatti, University of South Carolina, Columbia, S.C., USA (tel: 803 777 3123).

Pattern analysis often begins with the exploration of cultural categories - how people group concepts such as illnesses or symptoms into "things that are similar". The ANTHROPAC programme is designed to analyse data from pile sorting, triad sorting, paired comparisons, and other techniques used for finding the systematic relationships among items of "culture traits". Exploration of "categories", "groupings", and/or "dimensions" is carried out by means of multidimensional scaling, cluster analysis and other techniques.

Some of the main aspects of the ANTHROPAC "philosophy" (and utility) include the following:

(1) The programme is intended for explorations involving the inter-relations of qualitative and quantitative data.

(2) Most frequently the data gathering (various sorting tasks, rating, rank ordering, etc.) will involve small samples - no more than 50 persons (or other units), although larger data sets can be accommodated without problems.

(3) The exploratory data-gathering methods are intended to get emic data - data that reflect the vocabulary and categories of the local people among whom programmes and projects are to be carried out.

(4) In the ideal case the computerised data analysis will be in the field situation, with rapid feedback between data-gathering, analysis, and further data collection.

(5) The software is menu-driven, so that persons with relatively little experience with computers can learn essentials of the programme without extensive periods of training.

(6) The software includes a number of "utilities" or helpful tools for manipulating data, including dropping variables or cases; transforming rows and columns of data; display of data sets; merging datasets; and others.

(7) CONSENSUS MODELLING - a unique and important feature in ANTHROPAC - allows the researcher to determine the extent to which there is "cultural consensus" among any group of respondents. This feature also permits the researcher to identify the informants who best fit (correlate with) the cultural consensus in a given set of items (answers to questions, sorting or grouping of items, rating of attributes of items).

(8) The programme operates with ordinary ASCII files, which can be imported from a variety of different databases.

Additional features (selected list)

Some additional features in the current version of ANTHROPAC include the following:

- a. univariate (summary) statistics (e.g. mean, SD, range).
- b. a scale-construction programme (with factor analysis and Cronbach's alpha) for quickly testing the statistical composition of proposed quantitative multi-item scales. A Guttman scaling procedure is also included.
- c. PROFIT (property-fitting). To examine the fit of a hypothesized dimension by regression of data points on a multi-dimensional scale plot.
- d. loglinear modelling.
- e. feature for randomizing items (e.g. questions) in a rating or ranking questionnaire; also "unrandomize" resulting data set.
- f. "collapse" and dichotomize variables.

- g. insert missing values (e.g. means of columns or rows).
- h. QAP - to compare two proximities matrices to test hypothesis of "significant difference".
- i. The simple-to-use Editor function in ANTHROPAC which uses basic WORDSTAR commands, provides an easy means for entering datasets directly into the programme.

#### References

4. Hudelson, P.M., Kendall, C., Leontsini E., Pelto, P.J. (1991) Using formal anthropological methods and ANTHROPAC to design dengue fever health education materials. Dengue Newsletter Vol. 16: 48-57. World Health Organization (South-east Asia and Western Pacific Regions). In Collaboration with the South-Pacific Commission.

### EVALUATION OF ANTHROPAC

#### INTRODUCTION

The presentation covered only a small part of the ANTHROPAC package - in particular, pile sorting, multi-dimensional scaling and consensus testing.

#### QUESTIONS:

1. In what contexts is ANTHROPAC useful for TDR/SER projects?

The package, and the methods whose use it facilitates, could be very useful for exploratory research for TDR/SER in the following ways:

- (a) for determining associations of symptoms with diseases,
- (b) for determining choice of treatment for conditions,
- (c) for identifying useful key informants for research and other purposes,
- (d) for helping researchers organize information collected in early stages of research,
- (e) for helping researchers draw out the vocabulary of any domain.

2. What are the strengths and limitations for the methods from TDR/SER's perspective?

#### Strengths

- (a) It can be used with small samples.
- (b) It is menu-driven.
- (c) It provides easy access to analytic techniques for classification analysis (although most are available, e.g. in SAS or SPSS).

#### Limitations

- (a) There is no good manual for how to use the programme.
- (b) The field techniques it uses (e.g. pile sorts, triad sorts) are not widely known.

(c) Labels for some of the output (e.g. clustering display) could be improved.

3. What is needed to make it usable by new researchers?

(a) A manual that includes full descriptions of the field techniques (selection of and interaction with respondents and recording of data) and use of the programme (entering data, appropriateness of different analyses, interpretation of output).

(b) Further training of researchers either in a workshop or through individual collaboration.

4. What action should be taken by SER?

(a) Encourage development of better manual (currently underway).

(b) Hold a workshop for two main reasons:

- (1) to train a small number of people to use these methods, and
- (2) to generate a set of examples that could be incorporated into a manual.

### MICROSIMULATION MODELS

#### DESCRIPTION

Guided by numerical data from previous field and laboratory studies together with one or more theories of population dynamics and disease transmission, a microsimulation model is a computer implementation of the proposed dynamical processes constrained by the empirical evidence. From the perspective of many TDR field studies, a microsimulation model would consist of a "computer village", defined on a spread sheet, testing persons, houses and other village structures, vegetation, etc., together with a set of algorithms describing human population mobility (and, hence, exposure) to a vector borne disease), interventions, and their potential effects on people and vectors. The model could be used to generate artificial data sets, consistent with the prior empirical knowledge base; and then the data could be subject to the same statistical analysis that would be carried in an actual field study. This exercise allows the investigator to assess, in advance of a study, precisely those questions that are or are not answerable with data sets of various sizes. Thus microsimulation of different sizes of populations serves as the instrument for sample size determination in complex studies where the investigator is interested in addressing a multiplicity of questions.

#### EVALUATION OF MICROSUMULATION MODELS

##### QUESTIONS

1. In what context is it useful within TDR/SER?

(a) The model could be used in any project concerning interrelationships of factors and to test concepts and options for predictions. The areas of application range from biomedical questions to behavioural and economic issues.

(b) Nearly every project could benefit from this kind of analysis, even if only as a checklist of specific information available and required to answer a question. When a researcher has a hypothesis he/she should write down what relationships are expected.

- (c) The method helps to identify the major components in which data is needed, e.g. to define "population at risk" or also a "vector population".
- (d) It is useful even where a great deal of data exist for a more comprehensive assessment of a situation; particularly (i) how much risk behaviour is attributable to particular groups and (ii) what elements of a given system are/will be more sensitive to an intervention?

## 2. Strengths

- (a) It is systematic, allowing gaps in knowledge to be identified. At the same time it facilitates the discovery of new relationships (e.g. that use of a drug leads to less infective vectors). It generates new research hypotheses.
- (b) It allows testing for and improvement of the sensitivity of the different components of a particular model.
- (c) It can be used for any disease, condition or situation. Although it originated from the disease side (morbidity and mortality) it can be used for other outcomes as well (behavioural, economic).
- (d) It does not fit into conventional concepts of modelling - rather than creating more compartments it aims to reduce them to the fewest and simplest, most applicable form possible.

## 3. Limitations

- (a) Special training or collaboration with researchers experienced in use of model is necessary. The model requires a considerable amount of "hands on" help.
  - (b) Requires logic, basic mathematics, understanding of disease and related factors. These skills may not always be combined in one individual so multi-disciplinary groups may be necessary to develop and use the models.
  - (c) It is not useful in itself to answer exploratory anthropological questions but answers to such questions should be amenable to quantification for use in the model, as relevant.
3. What is needed to make it usable by new researchers in terms of training and methodological development? How costly is it?

## Methodology

The basic methodology is clear but it needs more development for different applications. For this someone who understands modelling as well as a biomedical scientist is required.

A guide rather than a manual is needed in order to explain the conceptual framework as well as practical use of the model. The conceptual framework must be made very clear.

## Training

It is not amenable to a workshop training exercise but rather requires on the job training and one-to-one dialogue. Such training is very intensive and would be very difficult to achieve within a Ministry of Health framework, for example.

#### Costs

Very difficult to estimate as they depend on the training process which will be different for each setting (see above).

#### 4. What action should be taken by SER?

- (a) SER could pick projects which are in the planning stage at present and provide support to the research team to learn and use this approach in projects.
- (b) Show usefulness of this modelling approach in the planning stage of a project (e.g. in relevant workshop of protocol development, consultancies for project development) which will depend on capability and interest in the SER group and associated colleagues.
- (c) SER could also encourage interaction with those working with the model and skilled researchers in developing countries.

#### GRADE OF MEMBERSHIP

Grade of Membership (GOM) models represent multi-dimensional discrete data in terms of profile of characteristics (e.g. ideal symptomatology for diseases, indications of exposure to acquisition of vector-borne diseases, clusters of indicators of economic status, etc.) and, for each individual, a set of scores (GOM scores) interpreted as the person's degree of similarity to the profiles. These models are particularly useful for characterizing very heterogeneous populations. They acknowledge the fact that people are often not classifiable into crisply defined and readily interpretable categories. Just as only a few people with a specific medical disorder are "classic cases" with all symptoms, so only occasionally do people fall crisply into one category of health status. More commonly, people show diversity in their symptomatology, especially those drawn from community samples. Characterizing them by their degree of similarity to two or more profiles of symptoms and disabilities, rather than along a single dimension, is more appropriate for very heterogeneous populations. Furthermore, many individuals have multiple disabilities or conditions, no combination of which occurs with high frequency in a community-based population. Thus crisply defined categories cannot be meaningfully constructed.

Grade of Membership models are of a form of latent variable models and may be interpreted as a form of discrete-variable factor analytic specifications.

In the absence of well developed theories about the structure of meaningful profiles, GOM models may be used in an exploratory mode to detect possible profile structures compatible with given data. Alternatively, if in the context of, for example, malaria transmission in a region of substantial population mobility we have a clear idea of the indicators of exposure, then risk profiles in a GOM model can be imposed a priori and GOM scores estimated for each individual. The a priori constrained-profile-GOM model can then be tested for compatibility with the data set at hand.

Outcome responses - e.g. malaria illness within the past 2 months - can then be defined conditional on risk/exposure characteristics (i.e. GOM scores) of different sub-populations. In addition, if more than one set of profiles - each set associated with a different dimension or risk - and associated GOM

scores are calculated for a given population, we can do regression of outcome responses on GOM scores to assess which dimensions, and which GOM-score levels, are associated with negative outcomes. These analyses are very effective way of defining target sub-populations for new disease control intervention strategies.

#### EVALUATION OF GRADE OF MEMBERSHIP (GOM)

1. In what context is GOM of use for TDR/SER projects?

GOM representations are suitable for the rapid ascertainment of high risk subpopulations that a control programme would target in attempting interventions in the presence of high mobility and lack of social cohesion.

2. What are the strengths and limitations of the method from TDR/SER's perspective?

##### Strengths

- (a) GOM is useful for characterizing very heterogeneous populations, and allows for the fact that people are not often classifiable into crisply defined and readily interpretable categories.
- (b) GOM models may be used to explore possible profile structures compatible with given data.
- (c) GOM scores can be used for further statistical analysis, e.g. regression to define target sub-populations at particular risk.
- (d) Note: Display of an individual response vector together with the individual's GOM scores and the profile description provides a very in-depth description of individual characteristics. When using GOM in clinical diagnosis this is one of its most important features.

##### Limitations

- (a) Software and manuals are not available for general dissemination.

3. What is needed to make it usable by new researchers?

- (a) Substantial software has been produced at Duke University and Yale University. However, a readily accessible "user friendly" set of programmes has yet to be developed.

4. What action should be taken by SER?

- (a) SER could sponsor a comparative analysis of data sets on one or more tropical diseases using different methods such as regression, cluster analysis, factor analysis and GOM in order to investigate the strengths and limitations of GOM in relation to other methods.
- (b) SER could prepare a collection of publishable papers that apply the technique and one or two papers that describe the theoretical basis of the model.
- (c) Test modifications in the model as it develops through appropriate field runs and comparative studies.

THE GENERALIZED AUTOMATIC TEXT ORGANIZATION AND RETRIEVAL SYSTEM (GATOR)

Description

GATOR (the Generalized Automatic Text Organization and Retrieval System) is a database system that indexes and retrieves information from machine-readable texts such as interviews and case histories.\* It can assist sociologists, psychologists, historians, clinicians, psychiatrists, curators of oral history collections, and others who have a need to organize large files of text of subsets of them.

Designed for use on a microcomputer, GATOR is flexible enough to be tailored to specific applications yet general enough to support many applications without special programming. Because it can easily be used with standard statistical packages, notably SPSS (the Statistical Package for the Social Sciences) and SAS (the Statistical Analysis System), GATOR should be especially attractive to researchers who wish to integrate both qualitative and quantitative data. Unlike other microcomputer data-base systems, GATOR does not rely on prior designation of fields. Boolean logic is used for indexing the searching text files but GATOR has the ability to index randomly organized texts making it especially powerful in retrieving and organizing information from diaries, clinical histories, transcriptions of interviews and other informal materials.

Until recently, researchers engaged in qualitative analysis have had little computer-assisted support, particularly for qualitative analysis of texts. Although text-analysis systems have existed for some time, most have not offered the flexibility needed for indexing and retrieving both written and spoken text. Most content-analysis systems function by performing automatic "lookups" in an on-line dictionary. This procedure saves time when a mass of text must be analysed, since researchers do not have to read and code the text, but many forms of text do not lend themselves to automated analysis. A procedural dictionary lookup routine in such contexts requires that specific words appear in the text. If they do not, researchers lose data and, in the process, lose considerable control over the interpretation of the text.

The structure of spoken text has particular complexities which make it difficult to index with already available software. Most computer-assisted indexing systems, including the newest ones, assume that ideas are discrete and linear, hence sequentially indexable. In spoken text, however, ideas are rarely discrete and linear. Instead, as interviewees recount a story or make a point, ideas and recollections are usually condensed and bundled. Often a block of text may contain a number of ideas that a researcher would like to index. Just as often a block of text containing a single idea may overlap other blocks of text containing other ideas. Indeed, the taut structure that one hopes to find in formal, written text rarely exists in spoken or informal text because many speakers and writers think extemporaneously, without regard to structure and polish.

---

\*Text taken from Giordano, 1988 (see references)

GATOR was first developed to analyse approximately 13,000 pages of text generated by a study of career patterns of American scientists (Zuckerman, Cole and Bruer, forthcoming). It is useful for research in which questions and lines of analysis arise that cannot be fully anticipated and that may result in modifications of methods and focus. A software system for such open-ended enquiry must be designed to allow for continuing additions and changes without requiring constant gutting and starting over. GATOR, therefore, is made up of programme modules. New programmes can be added and existing ones changed. The over-arching principle behind the development of GATOR was that the research design must influence the technology, rather than the other way around.

GATOR makes no assumptions about the context, structure, content or layout of raw text, except that keywords and their associated text are set off from the rest of the text by markers, or delimiters. Because randomly structured spoken text was the principal source of data, GATOR was designed to handle keywords and associated text of arbitrarily varying length.

Before GATOR can be used all text must be converted to machine-readable form. The most convenient and least expensive procedure for transcribing interviews (and other texts) is word processing. After each text is transcribed, it must be coded - that is, keywords must be inserted and the text associated with each keyword must be marked off. Coding can be done directly on the word processor, but it is advisable to print the transcribed text, code the printout, and then insert the key words and text markers into a copy of the interview residing on a floppy or hard disk.

GATOR builds a master file of key words and associated text based on the markers inserted into the interview text. The master file is a subset of the full record. It contains only specific keywords and their associated texts, substantially reducing the space needed for storing the data to be analyzed. GATOR needs at least two pieces of information before it can extract keywords and their associated text from the larger record file: the locations of the keywords and the locations of the text associated with the keywords.

After GATOR extracts a keyword and its associated text, it writes it, along with other information, to a master file. As interviews are processed by GATOR new keywords and text are appended to the master file. All keywords and text can be directed to one master file or to as many master files as there are overlay delimiters. For instance, all keywords and text marked off with "A" can be directed to one master file, while those marked off with a "B" can be directed to another. This may be useful when several coders are working on the same texts or interviews and each wants to create a separate master file. All other GATOR programmes used in analysis, such as the searching programmes and the report generator, refer to the master file or files, not to the interviews. The interviews can then be stored on backup disks, written to tape or deleted from disk.

For textual analysis, GATOR can be integrated with other computer programme, so long as at least one link is provided between statistical files and text files to allow for both qualitative and quantitative analysis. Cases from the statistical files that meet certain numeric criteria can be selected and their ID or cell numbers written to a disk file. GATOR can then be ordered to extract the corresponding qualitative records from its master files.

### References

5. Giordano, R., Cole, J.R., and Zuckerman, H., Text Retrieval on a microcomputer. *Perspective in Computing*, 8(1): 52-60. Software available from Ms. Debbie Gilcrest, Centre for Social Sciences, Columbia University, International Affairs Building/8th floor, 420 W 118th Street, New York 10027, USA. Telephone: (212) 854 3081.
6. Cole, J.R., and Singer, B., A theory of limited differences: explaining the productivity puzzle in science. In Zuckerman, H., Cole, J.R., Bruer, J. (eds) The Outer Circle, New York: N.W. Norton.

### Evaluation of GATOR

#### 1. In what context is GATOR useful for TDR/SER?

This programme is useful for almost all qualitative data collected for SER purposes. Compared to other methods of text retrieval it seems to retain the content better. Specific examples where it could be use are:

- (a) For studies of attitudes, perceptions and behaviour.
- (b) For health-seeking behaviour studies.
- (c) For analysis of in-depth interviews, focus groups, observational and field notes.
- (d) For studies on women and tropical diseases, almost all of which are descriptive, or have descriptive components.

#### 2. What are the strengths and limitations of the method from TDR/SER perspective?

##### Strengths

- (a) It follows the traditional steps of anthropological data analysis - going through the data and classifying or coding it.
- (b) It helps to explore issues related to a research hypothesis which may not be captured by pre-coded questions.
- (c) It forces one to know one's data, allowing the interaction of the researchers with the text to be dominant.
- (d) It is transparent in that it allows other researchers to review how particular statements were interpreted.
- (e) It can be connected to other packages like SPSS for quantifying and further statistical analysis.
- (f) It works on most word processor files.

##### Limitations

- (a) Considerable time is required to input the data into the computer.
- (b) It is labour intensive as the initial coding takes time.

- (c) If the data is poorly coded, much of the data may remain in the original file and not be analysed.
- (d) The researcher must maintain consistency in the coding categories which may become difficult with large quantities of text. This is a general limitation for analysing qualitative data.

3. What is needed to make it usable by new researchers?

It is easily available but appears to require more testing by social scientists.

4. What action should be taken by SER?

- (a) Testing the method further within social science and tropical disease settings.
- (b) Dissemination of information, possibly through demonstrations in workshops.
- (c) Encouraging use of GATOR.

HEALTH RESOURCE ALLOCATION GAME/MODEL

DESCRIPTION

The Health Resource Allocation Model (HRAM) is partially based on a card board game designed by the Royal Tropical Institute, Amsterdam. In collaboration with the London School of Hygiene and Tropical Medicine, the Swiss Tropical Institute (STI) has established a computer version that simulates the context of a real situation, Kilombero District (Tanzania).

The HRAM was designed to illustrate factors and determinants that are important in health planning. It can help:

- to explain principles of resource allocation at district level,
- to identify information needs for rational planning,
- to understand the setting up of health services for a district,
- to explore simulation models, their strengths and limitations.

The HRAM cannot consider all factors and is therefore not designed to reproduce the real world situation. The HRAM is a powerful tool for training health district managers and provides a guide in the planning of outcome-oriented operational research (mainly health systems research).

The HRAM consists of two basic sets of data: basic information about the district and variable data for the testing of health planning options. The outcome measures of the HRAM are the proportion of people that are served in the district (broken down by village) and the number of deaths the service prevented.

1. Basic Data

The basic data set contains:

- (a) Geographic (villages, roads, conditions of roads) and demographic information.

- (b) Available health facilities including their relative costs and capacity.
- (c) Morbidity and mortality patterns based on actual rates and translated into incidence rates for the health services. This set also includes activities (immunization, home visits) that are translated as incidence rates for the services. Finally, the minimum level of care for each disease/condition is specified.
- (d) Diseases and conditions including preventive activities are classified into "severity" categories. These categories reflect a composite variable consisting of the perception of disease by the population and health-seeking behaviour.

## 2. Variable data

The HRAM allows the user to set total budget for health and to allocate resources for management, drug supply, training and supervision. Once these allocations are made, the health facilities will have to be allocated in order to reach a cost-effective health care delivery system that respects equity.

## 3. Results

The HRAM subsequently assesses the health system by testing how well the system can serve the given population, including options for the referral system. The main output is the proportion of people served and the number of deaths prevented. These results are broken down by village and can be compared on the screen or on a comprehensive outprint with the overall data for the whole district. The analysis of the results provides the basis for a new health plan and permits the running of a series of sensitivity analyses for various planning options.

## 4. Future development of HRAM

The HRAM offers various possibilities for future development and the adaptation to other settings. The flexible structure of the programme allows refinement of all data files and the addition of basic or variable data files that improve the simulation capacity of the model. Emphasis will be placed on (i) extending the planning frame from one year to five years, (ii) refinement of the economic data set, (iii) the inclusion of more preventive activities in the disease/condition data set and the possibility to assess the impact of the investments in prevention, (iv) the refinement of the health-seeking behaviour module and (v) the possibility to test specific intervention strategies such as vaccination versus comprehensive mother child health services.

## 5. Information about the programme

The HRAM was written and compiled in Turbo C by STI. The present version (1.04d) still represents a developmental stage. It is accompanied by a draft manual that allows one to use the HRAM within the context of the Kilombero District, Tanzania, but does not allow the user to adapt it to his/her local setting. Interested researchers and institutions should contact STI for a copy of the model and/or proposals of collaboration in research and training:

Swiss Tropical Institute  
Socinstrasse 55a  
CH-4051 Basel/Switzerland

Tel: 061-284 8287/FAX 016-2718654/Telex 962508

Evaluation of Health Resources Allocation Game/Model (HRAM)

1. In what contexts is HRAM useful for TDR/SER projects?

- (a) HRAM can stimulate outcome-driven research in health planning and evaluation, and facilitates the construction of more sophisticated models.
- (b) The model can be adapted for tropical diseases and disease control in specific localities and situations.
- (c) It can be used to train various health cadres in health planning and to stimulate discussion among health planners concerning health planning options.

2. What are the strengths and limitations of the method from TDR/SER's perspective?

Strengths

- (a) It stimulates thinking about planning and resource allocation processes, particularly about essential data/information needs.
- (b) It provides a conceptual framework on health resource allocation.
- (c) The model can easily be extended and applied to different situations and used as a teaching and training tool.

Limitations

- (a) HRAM has a bias towards curative medicine and is thus oriented more to disease than health or disease prevention.
- (b) HRAM is restricted to the health sector for resources and does not involve other sectors such as agriculture, water resources, NGO's etc.
- (c) The model assumes overall outcome measures. For example, the number of patients treated may be an inaccurate and inappropriate indicator of the function of the health system in any situation.

3. What is needed to make it usable by new researchers?

- (a) Further development and adaptation of HRAM requires close collaboration between researchers and computer programmers.

4. What action should be taken by SER?

- (a) SER could encourage and target disease specific studies on planning disease control options in relation to resource allocation which will allow the model to move from training towards planning for policy orientation.
- (b) As HRAM provides a conceptual framework it can be used by SER to guide operational and health systems research, particularly at the planning stage of a project.
- (c) To test modifications in the model as it develops through appropriate field runs and comparative studies.

LIST OF PARTICIPANTS

1. Aguilar, Marcelo, Instituto Juan Cesar Garcia, Quito, Ecuador
2. Anyangwe, Stella, Public Health Unit, CUSS, Yaoundé, Cameroon
3. Barry, Michele, Yale University School of Medicine, New Haven, CT, USA
4. Berggren, Warren, Save the Children, Westport, CT, USA
5. Berkman, Lisa, Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT, USA
6. Blumenthal, Ursula, Department of Epidemiology and Population Studies, London School of Hygiene and Tropical Medicine, London, UK
7. Bonilla, Elssy, Department of Economics, Andes University, Bogota, Columbia
8. Briceno-Leon, Roberto, Laboratorio de Investigaciones Sociales, Universidad Central de Venezuela, Caracas, Venezuela
9. Chuit, Roberto, Servicio Nacional de Chagas, Cordoba, Argentina
10. Khan, M.E., The Population Council, Baroda, India
11. Kleinman, Arthur, Department of Anthropology, Harvard University, Cambridge, MA, USA
12. Knapp, Michael, East Haven, CT, USA
13. Makuch, Robert, Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT, USA
14. Manderson, Lenore, Tropical Health Education Programme, University of Queensland Medical School, Brisbane, Queensland, Australia
15. Pelto, Pertti, J., Department of Community Medicine, UConn Health Center, Farmington, CT, USA
16. Pino, Marco, Instituto Juan Cesar Garcia, Quito, Ecuador
17. Rao, Shobha, Department of Biometry and Nutrition, Maharashtra Association for the Cultivation of Science, Pune, India
18. Sawyer, Diana, CEDEPLAR/UFGM, Belo Horizonte, Brazil
19. Singer, Burton H., Chairman and Associate Dean for Public Health, Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT, USA
20. Singhanetra-Renard, Anchalee, Department of Geography, Chiang Mai University, Chiang Mai, Thailand
21. Tanner, Marcel, Swiss Tropical Institute, Basel, Switzerland
22. Vlassoff, Carol, World Health Organization, Geneva, Switzerland
23. Waxman, Matthew, New Haven, CT, USA

24. Weiss, Mitchell, Department of Social Medicine, Harvard Medical School, Boston, MA, USA
25. Wijeyaratne, Pandu, Health Sciences Division, International Development Research Centre, Ottawa, Canada
26. Wilson, Mark, Department of Epidemiology and Public Health, New Haven, CT, USA
27. Zimicki, Susan, Centre for International Health & Development Communication, Philadelphia, PA, USA

\*\*\*