Inferring morbidity and mortality rules for household level simulation models in rural Zimbabwe

J A Wright, Institute of Ecology and Resource Management, University of Edinburgh, Agriculture Building, West Mains Road, Edinburgh EH9 3JG, Scotland, UK.
S W Gundry, Institute of Ecology and Resource Management, University of Edinburgh, Agriculture Building, West Mains Road, Edinburgh EH9 3JG, Scotland, UK.
J Worrall, Institute of Ecology and Resource Management, University of Edinburgh, Agriculture Building, West Mains Road, Edinburgh EH9 3JG, Scotland UK.
A Kelly, Institute of Ecology and Resource Management, University of Edinburgh, Agriculture Building, West Mains Road, Edinburgh EH9 3JG, Scotland UK.

Abstract. This paper reports the development of a set of rules that determine the incidence of illnesses and accidents, with outcomes, for subsistence farming households in the Buhera District of Zimbabwe. This rule-set is part of a comprehensive database of rules within a model of the food and nutrition system for the study district. The model simulates the behavioural responses in 40,000 households to spatial and temporal change in environmental, economic, and social conditions. The rules are inferred from an analysis of a longitudinal survey of sample households in the district and will also be inferred from analyses of secondary data, reported research findings from similar locations and from encoding of local 'expert' opinions. As many of the variables concerned are categorical, decision rules are identified from field survey data using non-parametric statistics. Initially, respondents' diseases are grouped into seven categories and the probabilities of affliction for each category, for each age cohort and for each survey round are estimated together with the mean affliction period and outcome probabilities. A set of probabilistic and deterministic rules is then formulated and for successive time-steps, these are applied to all households and the health status of each is updated. The work illustrates how a set of simple rules can be formulated from diverse sources and applied to a large data set as part of a simulation model. Rule formulation is a means of formal decomposition of complex multivariate relationships, as found in the food and nutrition system.

1. INTRODUCTION

Field surveys for food security employ a variety of quantitative and qualitative techniques to understand local socio-economies. Initially, studies used questionnaire-based household surveys but analysis of these proved laborious and often yielded misleading results [Chambers, 1981]. As a supplement to the questionnaire survey, more qualitative techniques were deployed. Interviews with 'key informants' – experts with detailed knowledge of a particular area – were increasingly used [Seaman et al, 1993] and Rapid Rural Appraisal surveys became common. With quantitative studies and expert opinion now both seen as important, food system models in developing countries therefore need to be able to incorporate these disparate types of information. This paper examines a method of incorporating both quantitative results from a field survey and expert opinion within a rule-base that is used to simulate patterns of under-nutrition in a district of Zimbabwe. The model consists of a set of modules that predict the nutritional status of the district's population and one particular module – that concerning infection and recovery from illness and injury – is considered here.

The infection and recovery module has been developed from the results of a field survey in the Buhera District of Zimbabwe. Buhera forms part of Zimbabwe's Communal Lands, lower potential agricultural land occupied by smallholder subsistence farmers. Few households produce enough grain to feed themselves throughout the year, and most rely on income from off-farm employment to purchase grain through the market [Corbett, 1994]. Migration out of the district to urban areas, commercial farms, and mines in search of work is commonplace. Although living conditions have improved since independence, in 1992 73% of households still had inadequate sanitation and 59% had to travel more than 500 metres to their water supply [Government of Zimbabwe, 1993]. Health problems related to this environment are exacerbated by the spread of the HIV/AIDS virus, with prevalence in the 20-30 year age cohort approaching 25% in some areas [Gregson et al, 1996]. The poor health status of many individuals has consequences for their nutritional status, as the 'malnutrition-infection complex' results in weight loss. Tagwireyi and Greiner [1993], for example, have highlighted episodes of diarrhoea resulting from poor weaning practices as being one of the major causes of under-nutrition in children under 5 years.

2. INTERFACING OF RULE-BASE AND RELATIONAL DATABASE

As noted by Gundry et al [1997], the simulation strategy adopted here makes use of a modular approach based on existing software packages. Two existing applications are employed: a relational database is used to hold details of households and individuals within the district, and a proprietary expert systems package is used to implement the rules.

2.1 Rule-base

The use of an existing package to develop the rules for the simulation has several key advantages. Firstly, sets of rules can be developed independently of the simulation package described here using stand-alone software and a well-tested interface. Secondly, the use of proprietary software solves several major design problems. The proprietary rule-base software used is capable of indexing sets of rules to facilitate rapid processing of large data sets. Once the rules have been developed, this means that they can be applied to a large number of households rapidly. The rule-base software also checks newly created rules for logical consistency, to ensure that any new input of opinion is valid. Finally, the rule-base is also capable of explaining the reasoning behind a particular output, making the model's operation more transparent to the user.

The final model will consist of a set of inter-related rules, organised into modules. Separate modules are planned for on-farm production, food access through the market and aid system, livestock and nutritional status.

2.2. Population database

The population database at present consists of a sample of 354 households from within the district (approximately 1% of the total population in 1992), though eventually processing of attributes for the entire district is envisaged. The database consists of five principal sets of tables: two tables describe the characteristics of households and two more tables the characteristics of individuals within these households. Dynamic characteristics of households and individuals that change during the course of a model run are stored separately from static characteristics. Dynamic characteristics are referenced both by a unique identifier representing the household or individual and by the time-step in which they were recorded. Initially, the population database contains dynamic characteristics for the first time step only, and then changes for each time step are appended to the dynamic characteristic tables during each simulation run. An additional table of health centres within the district will be used to store the number of sick, healthy, and malnourished individuals attending at each time step. This information will then be compared to actual data on changes in prevalence of under-nutrition and certain illnesses collected at these health centres [Wright et al, 1997b].

2.3. Interaction of rule-base and population database

The large number of records in each simulation run means that speed of processing is an important consideration. Consequently, several different methods for coupling the population database to the rule-base were investigated. These methods are illustrated in Figure 1. A prototype system, based on each of these methods of coupling the two packages has been developed to identify the most effective interfacing mechanism. In method 1(a), the relational database package acts as the client, and interacts with the rulebase software using Object Linking and Embedding (OLE) automation. With this method, a text file of household and individuals' characteristics is created from within the database software. An application program written within the relational database then instructs the rulebase package to import this text file and process it, according to a given set of rules. Results are again written to a text file, and this is imported back into the database and appended to the table containing dynamic characteristics of individuals. To date, this has proved the fastest method of interfacing the two packages.

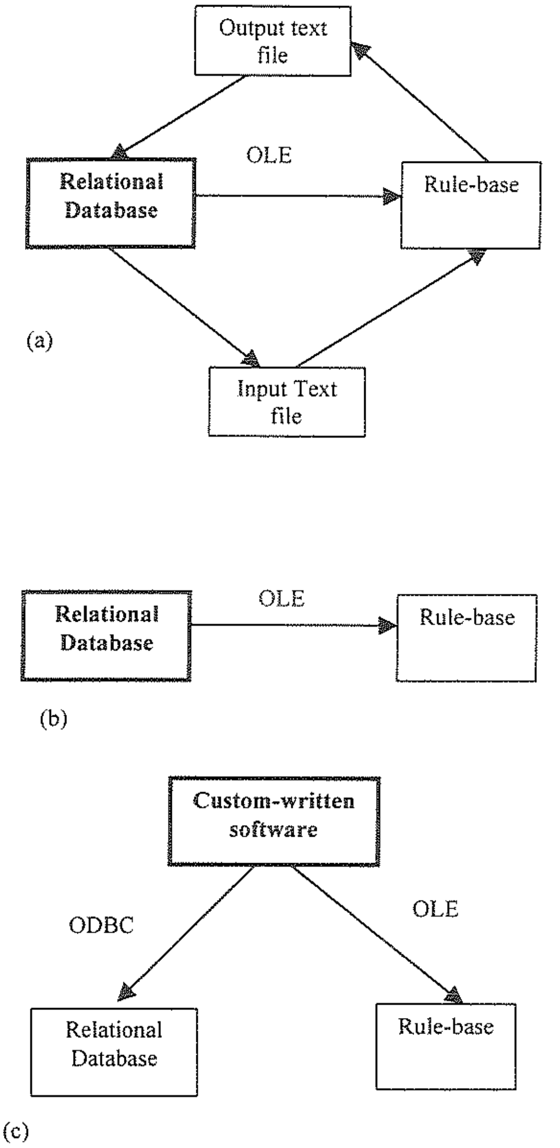


Figure 1: Methods of interfacing the relational database and rulebase: (a) from within the relational database using text files; (b) from within the relational database using OLE2; (c) from a separate custom-written program. The client program is shown in bold in each case.

In method 1(b), the relational database again controls the rulebase, but this time, however, details of households and individuals are passed to the rulebase through OLE rather than via a text file. In practice, a prototype model version implemented using this method proved slow, since there was a large processing overhead associated with communicating between the two packages. Consequently, this method of interfacing database with rules has been rejected.

In method 1(c), a program written separately from both packages in Borland Delphi was used as the client application. Information was read from the relational database using the Open Database Connectivity (ODBC) standard and then passed household by household to the rule-base via OLE. The same mechanism was used to extract results from the rule-base and append the results to the table of dynamic individual characteristics in the relational database. As with method (b), in a prototype model based on this approach, a heavy processing overhead was encountered in communicating between the two packages so this coupling method was also rejected.

A fourth possibility – using the rule-base itself as the controlling client application – is also currently under investigation. However, the more limited functionality of the rule-base programming environment relative to that in the relational database makes this a less attractive option.

The next section illustrates how these two packages can be combined using the example of rules governing health status.

3. ESTIMATION OF CHANGES IN HEALTH STATUS

3.1 Background

Underlying causes of nutritional status can be divided into three main categories showing links between inadequate basic health services, insufficient household food security and inadequate child care and feeding. These factors produce two immediate causes affecting nutritional status namely dietary intake and health and disease. Disease contributes to growth failure in several ways, since a sick individual may require more food because of fever yet intake is often reduced due to lack of appetite and an inability to digest food. This inability to digest food because of illness can affect growth in children, leading to stunting or weight-for-age malnutrition. Malnutrition, in turn, puts the child at even greater risk of contracting new infections thus creating a vicious and compounding cycle [Tagwireyi and Greiner ,1993]. The purpose of this module is to establish a value for health status, which in adults affects the income-generating capacity of households due to time off work, as well as influencing nutritional status directly.

The field survey that forms the basis for the infection and recovery module took place between 1994 and 1995. 354 households participated in the survey, representing approximately 1% of the district's population. Households were chosen on the basis of a stratified random sample that took into account agro-ecological zone, access to healthcare, and access to water [Wright et al, 1996]. Once every four months, an adult carer in each household was asked to list by individual any health problems that had been experienced over the preceding two weeks. The adult carer described the duration of any symptoms experienced and any time taken off work or school. The interview was repeated on three occasions, enabling seasonal variation in the prevalence of health problems to be identified.

The raw data recorded illness in the form of 95 codes representing illnesses or symptoms, many of which had similar causes and effects. This broad classification was simplified into seven main groups of similar illnesses:

1. General aches, pains and ear, nose & throat (ENT) problems
2. Diarrhoea, sickness, or stomach problems
3. Respiratory disorders
4. Injuries
5. Non-infectious internal problems
6. Other infectious diseases
7. Skin or eye problems

Some of the complaints, notably respiratory infections, diarrhoea/stomach complaints, and injuries, can be directly compared to health problems identified at health centres and recorded in the National Health Information System (NHIS). With the inclusion of suitable simulation of patterns of health care attendance, output for these complaints from this model can be validated against that recorded in the NHIS at health centres [Wright et al, 1997b].

Despite the high prevalence of AIDS/HIV, this complaint was never reported in the questionnaire because of the sensitivity of the issue and problems of diagnosis. Although HIV/AIDS sufferers are not therefore explicitly represented within the data, the impact of HIV/AIDS on health status is implicitly captured through secondary infections in the seven categories above.

In many instances, these complaints affect different age cohorts to different extents, and the prevalence of certain of these complaints varies depending on the time of year. Wright et al [1997a] have shown that the prevalence of diarrhoea reported by the National Health Information System in Zimbabwe peaks during the rainy season between November and February, whilst peak malaria prevalence occurs just after this period. Similarly, peak respiratory infection occurs over the cooler winter months and this situation is reflected within the questionnaire survey here (see graph C of Appendix A). Similarly, the Government of Zimbabwe [1988] found a particularly high prevalence of diarrhoea amongst children under 5 years: in a two-week survey period, 20% of under 5s had suffered from bouts of diarrhoea.

3.2. Generation of probabilities

The probability of any member of a household, of any age, in a particular round, suffering from any of the seven illness categories was calculated based on the following general equation:

$$P(\text{Ill}_{ria}) = \frac{\text{Number of individuals}_{ria}}{\text{Total number of individuals present in}_{ra}}$$

where r = round
i = illness
a = age cohort.

For simplicity, cases where individuals were suffering from more than one of the seven complaints were ignored. Thus, only the main illness or symptom listed in the questionnaire was considered in the above equation. The relationship between age, season, and health status can been seen in appendix A.

The total population within the sample (both healthy and sick) was also found to vary between seasons. This results from the traditional pattern of migration from Communal Areas such as Buhera to cities and commercial farms in search of work [Corbett, 1994]. Many such migrants return to their families during peak agricultural labour times, such as planting in October and November. This means that the age-sex structure of the district's population varies by season, together with health status. This migratory behaviour is being modelled in a separate module that handles grain trading and food access through markets.

3.3. Duration and economic impact of illness

Data of number of days ill for each individual for each illness of each age bracket was collected and the mean and standard deviation calculated. This created a set figure for the number of days ill for an individual suffering from a particular illness/symptom. The same analysis was carried out for days off work and/or household chores with the exception of the under 5's. In order to place these data into the model, which is dekad based, the results were scaled down proportionally to a ten-day period. These data will be used to simulate the effect of illness on income-generating and agricultural labour capacity within the household within the grain trading module described earlier.

4. IMPLEMENTATION OF HEALTH MODULE WITHIN THE RULE BASE

These probabilities were encoded within the rule-base package described earlier, as shown in Figure 2. Inference within the rule-base begins with the text file created from within the relational database being read into the package ('read_from_file'). The number of individuals is read in, and information about each individual is processed. The age category is calculated from the individual's age ('age_category'), and then the illness category is calculated from the probabilities described above ('ill_category'). Once the illness category has been calculated, then all variables are reset for the next individual in the household ('clear_loopvars'). Program execution continues until all households within the file have been processed.

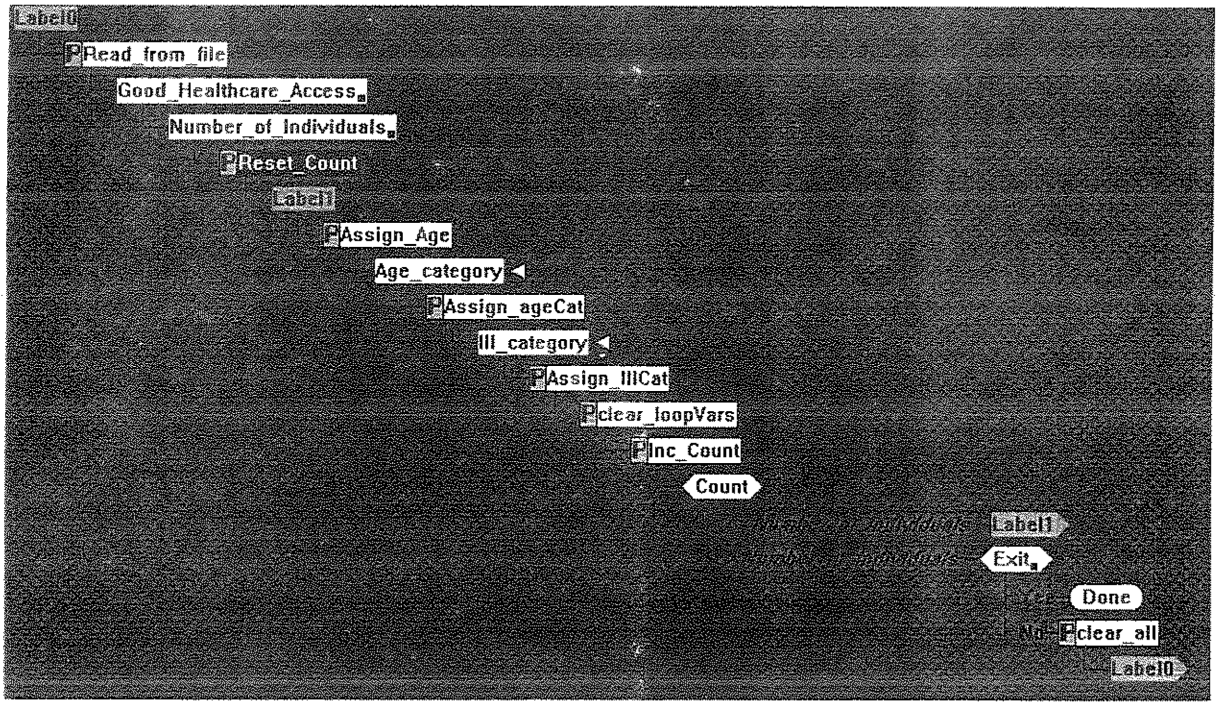


Figure 2: Diagram of implementation of health status module within rule-base (P represents small sections of programming code, hexagons represent conditional statements, whilst boxes marked with a triangle represent inference from rules)

5. CONCLUSION

The simulation approach described here couples together existing software packages, rather than developing new software where this is unnecessary. The two packages used here are a proprietary relational database and rule-base software. Both these tools can be used to edit and view elements of the model, as well as accessing these through the model itself.

Although rule-base and expert systems technology has been available since the early 1980s, the use of these techniques in simulation remains relatively rare. Rule-bases constructed for large-scale simulation need to be capable of rapid processing, and suitable indexing of different decision paths is therefore required. In addition, when a simulation system is created from existing proprietary software packages, an efficient mechanism for coupling these together is required. Initial testing of three different configurations here suggested that data transfer via simple text files represented an effective way of passing information between relational database and rule-base.

Having developed this simple health status module, future work on this simulation model aims to build similar modules to investigate patterns of grain trading, arable production, livestock management and nutritional status [Gundry et al, 1997]. It is envisaged that the construction of these modules may involve the use of more qualitative data sources, such as the opinions of 'key informants'.

Acknowledgements

This research was supported by the European Union, under the Science and Technology for Developing Countries Programme (STD 3). The authors would like to thank Mrs. J. Nyatsanza, G. Mudimu, and G. Makombe of the University of Zimbabwe for their invaluable assistance in Zimbabwe.

References

Chambers, R., *Rural development: putting the last first.* Longman, 180 pp., London.

Corbett, J., Livelihoods, food security, and nutrition in a drought-prone part of Zimbabwe. ESCOR report no. R4685. University of Oxford, 129 pp, Oxford, 1994.

Government of Zimbabwe, *Primary health care / maternal and child health / expanded programme on immunisation surveys.* Ministry of Health, Harare, 1988.

Government of Zimbabwe, *Census Provincial Profile: Manicaland.* Central Statistical Office, 147 pp, Harare, 1993.

Gundry, S.W., Wright, J.A., and Ferro-Luzzi, A., Simulating the food and nutrition system in rural Zimbabwe to support targeting of emergency food aid. MODSIM 97 International Congress on Modelling and Simulation, Hobart, Australia, 8-11th December, 1997

Gregson, S., Zhuwau, T., Anderson, R., and Chandiwana, S., *The early socio-demographic impact of the HIV-1 epidemic in rural Zimbabwe: summary report of findings from the Manicaland study of HIV-1 and fertility.* University of Oxford, 37 pp, Oxford, 1996

Gundry, S.W., Wright, J.A., and Ferro-Luzzi, A., Simulating the food and nutrition system in rural Zimbabwe to support targeting of emergency aid. MODSIM 97 International Congress on Modelling and Simulation, Hobart, Australia, 8-11th December, 1997.

Seaman, J., Holt, J. and Allen, P., *A New Approach to Vulnerability Mapping for Areas at Risk of Food Crisis.* Save the Children Fund, 11 pp, London, 1993.

Tagwireyi, J. and Greiner, T., *Nutrition in Zimbabwe: an Update.* The World Bank, 127 pp, Washington D.C., 1994.

Wright, J., Hoyles, P.J., Makombe, G., Gundry, S.W., Mudimu, G. and Nyatsanza, J., The use of Geographical Information Systems for household survey sample stratification: an application from Zimbabwe in nutrition and food systems research. Development Studies Association Conference, University of Reading, 18-20th September, 1996.

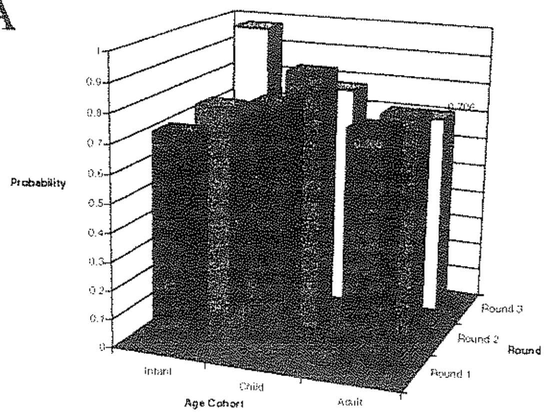
Wright, J, Gundry, S.W., Hoyles, P.J., Nyatsanza, J. and Vaze, P., Seasonal aspects of under-nutrition in young children in Zimbabwe. Development Studies Association Conference, University of East Anglia, 11-13th September, 1997.

Wright, J., Gundry, S.W., Ferro-Luzzi, A., and Hoyles, P., Validation of a complex spatial model of the food and nutrition system in Zimbabwe. MODSIM 97 International Congress on Modelling and Simulation, Hobart, Australia, 8-11th December, 1997.

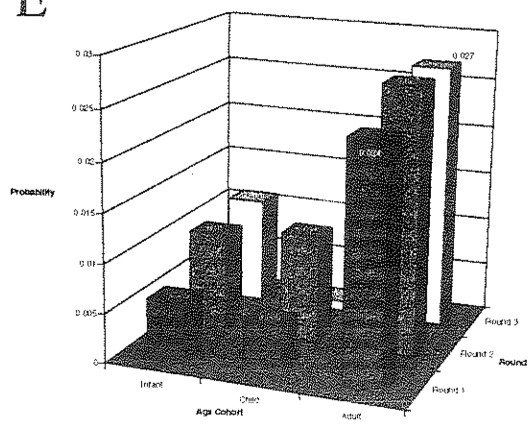
Appendix A: Relationship between illness, age, and season for:

A. Those not suffering any illness;
B. General illnesses or ENT problems;
C. Diarrhoea and stomach problems;
D. Respiratory complaints;
E. Injuries;
F. Non-infectious internal medical problems;
G. Infectious diseases;
H. Skin and eye complaints.

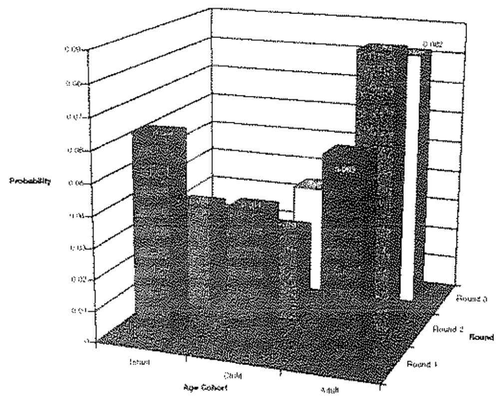
A



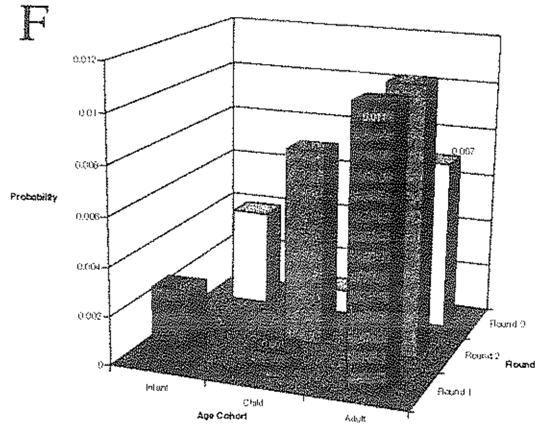
E



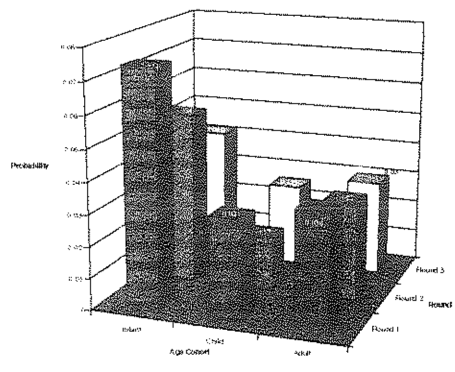
B



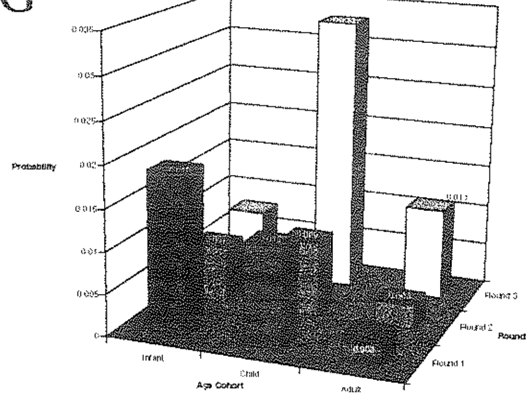
F



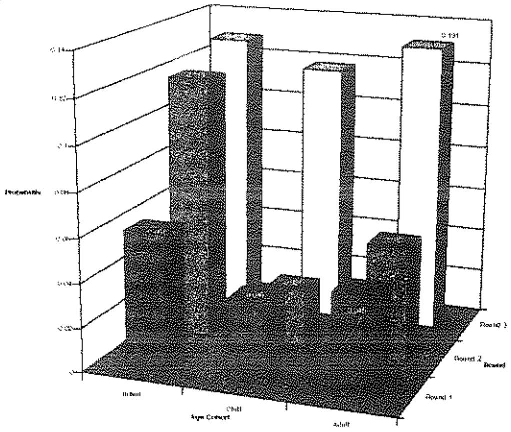
C



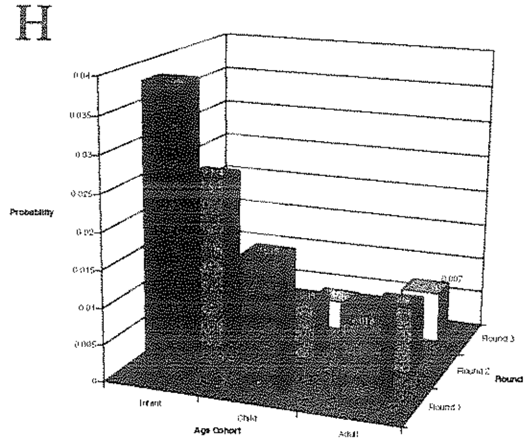
G



D



H



APPENDIX A