

Project Report on the Proposed Nutrition Surveillance System

By
Prof. J. G. Krishnayya, Medha Ashtekar

Full Report (51 PP), Rs. 250.00 (India), \$10.00 (Overseas) June 1996

Executive Summary: (from the Report Pages 4 - 7)

This study is primarily concerned with what can be done within the existing economic and social set-up to create **a model to be used at District level to forecast Nutritional stress on a village-wise basis**. It has the following general and specific objectives:-

General Objective: To develop a Nutrition Surveillance model at the village level useful for micro-planning, using simple yet sensitive indicators.

Specific Objectives 1: To study the relationship of agriculture, economic, environment and basic needs indicators to food consumption and nutrition status of the population.

2: To identify sensitive indicators associated with nutrition and health status and to develop a simple, fast and cost effective Nutrition Surveillance Model.

This simulation modelling exercise was carried out as a component of the National Institute of Nutrition project on creation of a Nutrition surveillance System which is to be deployed in all the districts and states of India. The overall project has included the collection of field data from the Villages in 2 Mandals each of 4 districts (in different agro-climatic zones of Andhra Pradesh), including obtaining a variety of economic and infrastructure and social information, as well as specific nutrition measures on the children in each Village. Additional necessary data was collected or Mandal and District level.

There may appear to be a contradiction between the massive data collection in this project and the small model that has resulted. In fact, one of the objectives of the Nutrition Surveillance Project was to identify which were the most sensitive indicators and to build a simple model using these indicators. We are happy to fulfill this objective with some help from System Dynamics technology.

Reasons for selecting System Dynamics are: System Dynamics is used to identify salient variables. The purpose of the System Dynamics Model is **understanding**. The goal of the modeling effort is to improve understanding of the relationships between feedback structure and dynamic behavior of the system.

System Dynamics focusses on the structure and behavior of the system composed of interacting feedback loops: by experimenting with the model, we identify which are the salient variables. This is the heart of the modeling exercise, as the main thrust of

Nutrition Surveillance modeling is to build the model using the minimum number of parameters.

If we try to express everything as a cause and effect relationship then Nutritional status is an effect of the socio-economic, health, cultural and location parameters. There are total n number of cause variables; some of them produce a constant type effect i.e. deterministic; while a few of them have volatile impact i.e. probabilistic, on the Nutritional status. It is important to identify deterministic and probabilistic variables separately, and treat them separately.

System Dynamics applies the notion of feedback, and allows repeated experimentation with the system, testing assumptions and altering policies. The purpose is to gain understanding, so that the problem which the model addresses is minimized and we are able to locate the salient variables. This is quite a difficult task using the usual set of differential equations; as the differential equations show the combined effect of all explanatory variables; and also the alteration of the form is not that easy.

The usual **Correlaton technique** gives the degree of relationship between various variables. This is quite helpful but it would not have allowed us to experiment with different sets of data. It does help to distinguish between static variables and dynamic variables.

Existing Data

The main emphasis of this model building exercise was to use existing data, and a minimum quantity. If accurate and proper data is supplied then a model can capture the fineness of the real world for practical purposes. Despite being handicapped by use of existing data we have formulated our model in such a way that it should be sensitive to changes in various explanatory variables, and it is expected that the Cross-sectional data across various villages from different agro-climatically socially and culturally different districts will capture all the broad features of the system. For example in our model, Income is explained in terms of Irrigated land, Unirrigated land, Rainfall, Family size, Literacy rate, so as to cover different agro-climatic and cultural regimes.

(NOTE on Data availability: Efforts were made to collect a similar data set for the "wet villages" in PUNE district, so that a comparative analysis could be performed, using the model estimated with Andhra data. To our surprise we learnt that this information (e.g. number of Pink and White card holders in each village; number of programmes operational in each village, etc.) is not available. Extension of this model would then be questionable on a national level unless the District statistical machinery is geared up to collect at least the very minimal standard set of data about each village on a quarterly or semi-annual basis. We had hoped that the National Informatics Centre's **DISNIC** database would contain the required information, but now we are doubtful.)

If our Village gets good rainfall then the Income will be good. If our Village is from a drought-prone area then reduced rainfall means the contribution from Irrigated and Unirrigated land will be lower. Alternatively, if a Village does not have a Agriculture base (which is very rare in India) then there might be a presence of some household industry, or something else that provides the livelihood to the villages, which we tried to measure as a function of literacy rate (because direct data on

Household industries is not available). If a village is socially backward then a prominent feature of the social structure of that village will be large families, which is a negative function for Income; and hence our claim is that our Model will be able to capture the agriculture and socio-economic features, which makes the Model sensitive to various levels of changes in the variables.

The Nutrition data used is Percentage of children in various grades for 3 age groups: below 1 year, 1-3 years and 3-5 years. NIN Experts say that Nutritional status of the children reflects the nutritional status of the whole society.

We have estimated equations for Irrigated land, Unirrigated land, Literacy rates, Family size, Agriculture Income even when actual values for them are available for the base year. The reason for doing so is that all these values are one-point values. For prediction purposes we need values for future years. By studying the relationships and formulating these equations we have solved this problem. The Formulation of the equations is based on the actual values of these variables.

The distinction between reality and perception of reality is vitally important at validation stage. It must be realized that not all quantities represented in the model are accessible within the system being modeled. Some information is inaccessible to the system. This brings a degree of uncertainty to the model behaviour.

The main reason for differences between the Predicted and Actual values are:

I. Any Model represents reality to the extent determined by the quality of the data on which the model is built. If the data set provided is incomplete then the model may well produce incorrect results. For example in our data set, in District 1 Mandal 2, data is not available for Literacy, Cultivators, PHC-distance, Bus-stop distance and No. of births. For District 4 Mandal 1 and Mandal 2 the data series for the variables: Births, Protected Couples (15-30), Protected Couples (1-45), Primary Schooled Females, & Secondary Schooled Females are not available.

II. For comparing predicted values and actuals we should ignore the outliers. Our model will not capture the extreme behaviour. In statistical theory the basic reason behind the existence of outliers is assumed to be errors of measurements. Take the case of Village 6 of District 2, it has a level of 97% for Normal and Grade-1 children for below 1 year, which is highly unlikely. It is most probably an error in measuring the grades by data collectors. In this way an error of measurement can bring unreliability to the model which is estimated on this data. Hence the quality of the data should be as good as possible. Of course, in the real world there are often extremely badly-situated villages, or extremely well-off villages, which will not be correctly predicted by our model either.

III. Non-availability of the data for a few particular series compelled us to use proxy variables, which in turn changes output in relative terms. For example we know from our previous studies that female literacy rate is a significant factor which affects the fertility - which in turn has a direct impact on family size. In our family size equation we could not use female literacy rate as data on female literacy is not available. Hence we ended up using a proxy variable, Total literacy rate. Total literacy rate is also an important variable but a large change in Female literacy (which would in reality be very significant at low levels) may have only a small effect on the Total literacy - and therefore on the output of our model. Actually Female literacy is a fairly critical data element in many feedback loops, as far as modeling Nutrition behaviour

is concerned. **This is probably the most important single piece of data that should be ensured to be available for model-building and similar applications in rural areas in India.**

IV. Use of generalised data like **Mandal-wise** Rainfall data is also responsible for making the model less sensitive to village-to-village differences. The availability of water in wells may well differ from village to village; this is most significant in a draught year. But this information is not available. There will always be differences in the micro-climates of neighbouring villages, depending not only on rainfall, but also on Insolation, the number of trees, and so on. All this detailed texture is lost in the model we have had to build with the data available.

Thus incomplete data, errors of measurement, use of proxy variable and geographic generalisation of variables representing important information as also the fact that villages of vastly different types are grouped together in this model are the main reasons why this model's predictions differ from the actual values. We have made an effort to improve the results by re-estimating the model with data only for the Wet Mandals. This does give better R-squared results in the equations.

However, we expect our Model to remain stable over the years as long as all externalities and the structure of the economic and social system remain the same.

The question will arise: Why not estimate the equations separately for each district? We must then have data for all the series for at least the base year 1993.

For District 1, Mandal 1, data is available for all the variables.

For District 1, Mandal 2 data is not available for Land holdings, Literate, Cultivators, PHC distance, Bus distance and No. of births.

For District 2, Mandal 1 and Mandal 2 data is available for all the series except Primary Schooled Females and Secondary Schooled Females (which can be dropped and estimation can be done); but there are only 36 data villages from both the Mandals.

Similarly for District 3, the total number of villages in Mandals 1 & 2 is just 31. If we ignore the data size and carry out the estimation, this will definitely affect the precision of the estimates. As the total number of explanatory variables are quite large this will reduce the degrees of freedom, which will reduce the precision of the estimates.

District 4: separately estimated

Using the recursive system, the model was estimated separately for District 4. District 4 has 41 data points. For District 4, Mandal 1 and Mandal 2, the following data series are, however, not available: Births, Protected Couples (15-30), Protected Couples (31-45), Primary Schooled Females, Secondary Schooled Females.

The equations for this model are presented at the end. In our opinion, however, the heterogeneity of the villages in any one district are not likely to be lower than in the data set as a whole, and we therefore prefer the alternative of using an environmental criteria alternative as below.

Wet Mandals only ..

During the 18 April 1996 meeting we learnt that the data had been collected from both Wet Mandals and Dry Mandals. After learning this, Prof. Krishnayya took the decision of estimating separate sets of equations for Wet Mandals and Dry Mandals. What we present below are Estimates for Wet Mandals only. In the Dry Mandals, we found the Data to be too widely dispersed and without any particular trends. Hence we could not estimate separate for Dry Mandals.

Recursive Estimates are therefore available separately (a) for All Villages (b) for District 4, (c) and for villages from Wet Mandals only. The model outputs are presented in the relevant chapters.

Typical Model Results (from overall Recursive Model):

The model is very sensitive to the Rainfal. For example, below we discuss the results obtained for Village 3 of District 2, Mandal 2.

For Rainfall - 842.00 mm, the Percentage of Normal and Grade1 children below 1 year is 77. Percentage of Grade2 children below 1 year is 19 and Percentage of Grade3 and Grade4 children below 1 years is 4 ... Now, if the rainfall goes down in one year by 50%, i.e. if Rainfall =421.00 mm, then Percentage of Normal and Grade1 children below 1 year becomes 60, Percentage of Grade2 children becomes 31 and Percentage of Grade3 and Grade4 children below 1 year becomes 9.

Similarly, the Percentages for age groups 1-3 years are -

	Normal and Grade1	Grade2	Grade3 & Grade4
Rainfall = 842.00	70	25	5
Rainfall = 421.00	60	32	9

and the Percentages for age group 3-5 years are -

	Normal and Grade1	Grade2	Grade3 & Grade4
Rainfall = 842.00	70	27	3
Rainfall = 421.00	59	35	6