Global Bureau, Office of Agriculture and Food Security

POLICY SYNTHESIS

for Cooperating USAID Offices and Country Missions

(http://www.aec.msu.edu/agecon/fs2/psynindx.htm)

Number 54 August 2000

Office of Sustainable Development

USAID/Mozambique

DEVELOPING COST EFFECTIVE METHODS FOR ESTIMATING HOUSEHOLD INCOME AND NUTRIENT INTAKE ADEQUACY

By

David Tschirley and Donald Rose

Food Security II Cooperative Agreement between U.S. Agency for International Development, Global Bureau, Economic Growth and Agricultural Development Center, Office of Agriculture and Food Security and Department of Agricultural Economics, Michigan State University

BACKGROUND: Governments, donors, and NGOs in developing countries spend billions of dollars every year on efforts to improve the well-being of rural households. Most of these interventions have the ultimate goal of reducing poverty, and many include specific objectives of improving household income, dietary intake, or nutritional status. Since an accurate assessment of these outcomes is costly and time-consuming, much research has attempted to identify simple indicators which are correlated with the variables of interest. For example, measures of household assets have been used as indicators of income levels (Little 1997; see Swindale and Ohri-Vachaspati 1999 on food consumption indicators). Evaluations of the interventions can then rely in part on changes in these indicators to indicate whether progress is being made against objectives. Unfortunately, these indicators are often not well-focused on project activities, and may not correlate well with actual objectives. Changes in household assets do not capture year-to-year variations in household incomes, and the purpose of the intervention may have nothing to do with increasing the assets in question.

This paper reports on work in Mozambique to develop an approach which more accurately assesses project outcomes. Specifically, the methods reported here use easy-to-collect information in combination with statistical methods to develop models that predict actual household income and nutrient intake adequacy (prediction models). If these methods are successful and transferable to other locales, they would allow evaluators to report directly on the outcomes of interest, rather than the indicators, at a small fraction of the cost of collecting and processing full income or food consumption surveys.[1]

OBJECTIVES AND METHODS:

Objectives: The objectives of this research were to:

1. Develop general approaches to using easy-to-collect information to generate quantitative estimates of household income and nutrient intake adequacy;
2. Apply these general approaches in Mozambique;
3. Evaluate how well the approaches perform across geographic zones; and
4. Draw lessons for how to improve and adapt the approaches to other settings.

Methods: Prediction models as developed in this research are one part of a package of procedures that NGOs, donors, governments, or research institutions can use to monitor rural household income and income

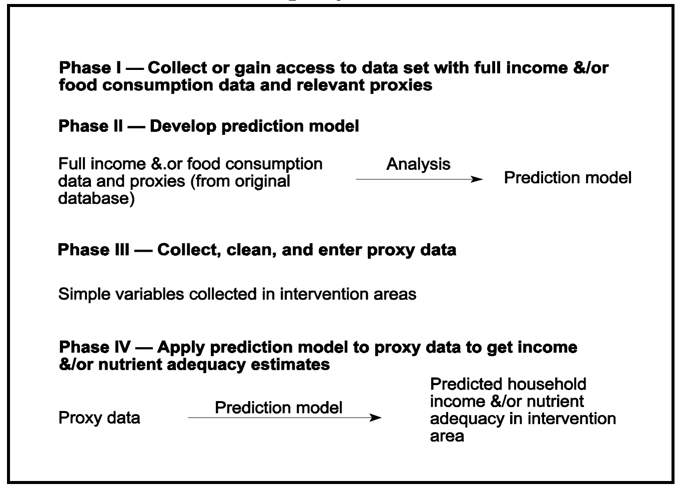
[1] For other multivariate approaches to developing poverty indicators, see Glewwe 1990; Glewwe and Kanaan 1989; Grosh and Baker 1995; Wodon 1997; Hentschel et al. 1998; and Minot 2000.

components, or household nutrient intake adequacy, using easy-to-collect variables. The income models and nutrient adequacy models require separate data sets and analytical procedures, but the general approach for each is very similar. Each model is a set of algebraic equations relating easy-to-collect proxy variables to the outcome variables of interest (income components or nutrient intake adequacy).[2] These algebraic relationships are developed using standard econometric techniques applied to household data sets which contain both types of data. In the income models, the components sum to total income.

Figure 1 provides an overview of the process for developing and using the models. An income prediction model would require a full income survey including the likely proxies for the income components, while the nutrient adequacy prediction model would require a food consumption survey and the likely proxies. These Phase I surveys, and the intensive data analysis of Phase II, need to be done once at the beginning of the project cycle. Once the detailed data sets are collected and the models are estimated (Phases I and II), one needs only to conduct the proxy survey (Phase III), then apply the model to the proxy data (Phase IV) to obtain estimates of the variables of interest.

The simple proxy surveys will take one-quarter or less time than the full income or consumption surveys, and need be conducted only once a year, or however often the institution wishes to track household income or nutrient adequacy. Data cleaning takes only a very small fraction of the time, due to the simple nature of the variables. Generating the estimates in Phase IV is nearly instantaneous, since the models developed in Phase II are incorporated into a computer program which conducts all needed manipulations on the proxy data and generates the estimates. For validation purposes,

Figure 1. Overview of Process to Develop and Apply Income or Nutrient Adequacy Prediction Models



the Phase I survey should be conducted again at a later time and, if needed, the prediction models should be recalibrated.

The complete package which defines the methodology includes: 1) sampling guidelines for the periodic proxy surveys; 2) model questionnaires for these surveys (one for income, another for nutrient intake adequacy); 3) the set of econometric models relating the proxy variables to the variables of interest; 4) computer programs based on these models that use the proxy data to generate the quantitative estimates of the variables of interest; and 5) a manual for operating the package.

We applied these general procedures in Mozambique using two data sets. The nutrient intake adequacy prediction models are based on data collected during 1994-96 in two provinces of northern Mozambique. The Nampula/Cabo Delgado (NCD) study was originally designed to identify the impacts of smallholder cotton schemes on household incomes and food security.[3] Using repeated visits on close to 400 households in 16 villages, the study collected

[2] We use "proxy variables" to mean easy-to-collect variables that form the basis of a prediction model. They are proxies for the more difficult-to-collect variables that would be needed to measure our outcomes of interest directly.

[3] See MAP/MSU 1996; Strasberg 1997 for methodological details on the original study. See Rose et al. 1999 for a detailed exploration of household food and nutrient consumption behavior based on this data set.

information on demographic characteristics, agricultural production and sales, expenditures on food and other necessities, and daily food consumption at three different periods during the year — May ("harvest"), September ("post-harvest"), and January ("hungry season"). Household food consumption was measured using a 24-hour recall technique administered by a trained interviewer to the person in charge of food preparation. These interviews were made on two separate visits during each period and included volumetric measurement of foods consumed.

To develop a prediction model, we used linear regression techniques in which the household intake of a nutrient was the dependent variable and the independent variables were the frequencies of consumption of foods from 11 different food groups, and household size.[4] The result of this technique is a simple algebraic relationship that relates each set of easy-to-collect proxy variables to each of the outcome variables of interest - nutrient adequacy at the household level for energy, protein, vitamin A, and iron. We also developed a Mozambican Diet Quality Index (MDQI). Which combines information about the intake of these four nutrients, as well as seven others, to provide an overall description of diet quality.

The income prediction models are based on data collected jointly by Michigan State University and USAID-funded NGOs operating in Mozambique. A random sample was drawn of 490 households stratified into seven geographic zones. Each household was interviewed twice during 1998. These interviews generated detailed information on household demographics, agricultural production and marketing (including livestock), off-farm income earning activities, and household assets.[5]

Two models were developed. INCPROX estimates 10 components of household income and, by summing these components, derives an estimate of total income. The components defined for Mozambique were the value in US$ of production of dry staple food crops, "other" crops (principally "cash" crops), all staple food crops in a fresh state, vegetables, fruit, cashew, and livestock, plus the value of any fisheries catch, earnings from wage labor (cash and in-kind), and earnings from microenterprises. There is no single "right" definition of components; those chosen should be relevant for understanding the rural household economy, and should be predictable with some accuracy using simple proxy variables. INCPROX Lite is a simpler alternative which requires fewer proxy variables and returns an estimate of total income with no breakdown by component.

In estimating each of the 10 income components for INCPROX, we chose proxy variables of three general types: (1) measures of the intensity of the household's involvement in each economic activity; (2) measures of the resources that the household could bring to bear on this activity (production function variables); and (3) zone variables which allowed the relationship between the proxy variables and component income to vary across geographic zones. Examples of measures of intensity (for the food crop production component) are the number of food crops the household cultivated, and the number of food crops that it sold. Production function variables were the same across all agricultural components: land (proxied by the number of fields cultivated); labor (the number of non-elderly adults resident in the household); and capital (defined as the number of types of farm implements that the household owned).

[4] The 11 food groups are grains, tubers, beans, nuts and seeds, animal products, vitamin A-rich fruits and vegetables, vitamin C-rich fruits and vegetables, other fruits and vegetables, sugars, oils, and other foods. See Rose and Tschirley 2000 for more detail.

[5] See Tschirley, Rose, and Marrule 2000 for detail on sample and overall study design.

Table 1. Measured Frequency of Low Intakes Compared with Predicted Frequency from the Prediction Model [1]

Nutrient	Post-Harvest Season		Hungry Season	
	Measured (% low)	Predicted (% low)	Measured (% low)	Predicted (% low)
Energy	25.1	25.9	58.4	61.6
Protein	7.7	11.9	55.2	62.4
Vit. A	97.7	99.2	81.6	88.0
Iron	20.2	17.9	53.6	53.1
Overall diet [2]	47.4	47.9	78.7	80.0

[1] A low intake refers to intakes less than 75 % of the recommendation.

[2] The MDQI was used to summarize information from the four nutrients and seven others in this table. Low quality diets are those with scores less than 7.5 on this 10-point index.

Quantitative production variables for maize grain and seed cotton were included in the analysis. These variables are more complex to collect and process than typical proxy variables, but are needed because production levels can fluctuate substantially from year-to-year based on rainfall and other factors. Including these quantities helps capture variation in yield levels of other crops within their category, and greatly improves the performance of the models.

KEY FINDINGS:

Nutrient Adequacy: Table 1 shows measured vs predicted frequency of low intakes for the four nutrients and for overall diet quality during the post-harvest and hungry seasons. For each nutrient, the predictions are better for the post-harvest season than for the hungry season, but regardless of season, predicted prevalences are quite close to actual.

Income: For each geographic zone, Table 2 presents means of calculated household income and predicted income from INCPROX and INCPROX Lite. It also shows the ranking of those means across the seven zones. In general the two approaches do quite well distinguishing income levels by zone. Specifically, INCPROX Lite results in the same income ranking as calculated income (though specific estimates differ), while INCPROX switches zones 3 and 5 but otherwise ranks all zones correctly. The model performs best in those zones whose income is closest to the mean across all zones.

We examined how well INCPROX ranked zones within income components. For example, which zones earn the most and least from non-staple crops? The models correctly ranked the zone in 51 of 70 possibilities (10 components across 7 zones), and 68 of the 70 rankings were correct within one place. In other words, ranking errors were nearly always the switching of adjacent zones. We next examined how well INCPROX ranked income components within zones. In this case, 42 of the 70 components were correctly ranked, and 61 of these rankings were correct within one place. Again, ranking errors were nearly always the switching of adjacent components.

CONCLUSIONS AND RECOMMENDATIONS:

Nutrient Adequacy Models: Results from these models are encouraging and suggest that the collection of very simple food consumption data may allow effective monitoring of nutrient intake adequacy once such models are developed. The geographically

Table 2. Zone-by-zone Comparison of INCPROX and INCPROX Lite in Level and Ranking of Predicted Income

Geographic Zone		Calculated Income		INCPROX Estimate			INCPROX Lite Estimate		
Name	Number	Income (US$/hh)	Rank	Income (US$/hh)	Rank	% Error[2]	Income (US$/hh)	Rank	% Error[3]
Manica	7	536.35	1	483.03	1	-9.9%	509.98	1	-4.9%
Central Sofala/Manica	6	482.92	2	464.09	2	-3.9%	425.79	2	-11.8%
Zambezi Valley	1	419.33	3	390.11	3	-7.0%	379.47	3	-9.5%
Cotton Belt	4	309.61	4	316.16	4	2.1%	306.50	4	-1.0%
Central Zambêzia	2	281.93	5	282.37	5	0.2%	289.88	5	2.8%
N. Zambêzia-S. Nampula	3	218.42	6	227.68	7	4.2%	239.20	6	9.5%
Coastal Nampula	5	200.66	7	233.36	6	16.3%	214.00	7	6.6%
All Zones[1]		299.18		299.18			299.18		

[1] Mean is weighted by zone level sample weights
[2] Mean absolute error = 6.23%
[3] Mean absolute error = 6.59%

restricted sample used to develop these models did not allow testing their performance over space, so this should be a key area for further research, in addition to their performance over time.

Income Models: Results for these models also suggest that INCPROX could be an effective tool to monitor household incomes and income components. However, the performance of the INCPROX components over geographic zones highlights the need for greater flexibility in the models over space, to ensure that the relative rankings of economic activities are more consistent. Other key areas for further research and improvement include:

- Testing how well the models perform over time by collecting a comparable data set 4-5 years after the original and evaluating model performance; and
- Incorporating as proxy variables households' subjective assessment of the relative importance of different economic activities; this work is currently being done in Kenya.

Adapting the Approaches to Other Contexts: Developing useful income prediction models in other countries will require good knowledge of the rural economy of the new locale, especially the extent to which cropping patterns and other income earning activities vary geographically. This knowledge will drive the definition of a relevant set of income components and the proxy variables needed to predict them. New nutrient adequacy models will require similar knowledge of how consumption patterns vary over space.

The new models will perform better if there is a close link between the design of the questionnaire for the original data set and the anticipated design of the proxy questionnaire: how a question is asked affects the answer received, so as much as possible, analysts should anticipate the proxy variables and obtain them in the same way in the original questionnaire as they anticipate doing in the proxy questionnaire.

Monitoring or Impact Evaluation?: These models are designed to capture the association between the outcome variables of interest and the proxy variables, and to predict as accurately as possible. As such, they can be used directly to monitor these outcomes. The models are not designed to allow conclusions regarding cause and effect; to use these models for impact evaluation (for example, to measure the impact of an NGO or government's agricultural production and marketing assistance on agricultural and overall household income, and on nutrient intake), they need to be integrated into an approach which includes an appropriate evaluation design.[6] Within such an integrated approach, use of income or nutrient adequacy prediction models can allow more frequent monitoring (because it will be less costly and time consuming), provide a richer set of monitoring results, and reduce the cost of the impact evaluation.

References:

Glewwe, P. 1990. *Efficient Allocation of Transfers to the Poor: The Problem of Unobserved Household Income.* Working Paper No. 70, Living Standards Measurement Study. Washington, D.C.: World Bank.

Glewwe, P., and O. Kanaan. 1989. *Targeting Assistance to the Poor: A Multivariate Approach Using Household Survey Data*. Policy, Planning and Research Working Paper 225. Washington, D.C.: World Bank.

Grosh, M., and J. Baker. 1995. *Proxy Means Tests for Targeting Social Programs: Simulations and Speculation*. Working Paper No. 118, Living Standards Measurement Study. Washington, D.C.: World Bank.

Hentschel, J., J. Lanjouw, P. Lanjouw, and J. Poggi. 1998. *Combining Census and Survey Data to Study Spatial Dimensions of Poverty.* Policy Research Working Paper No. 1928. Washington, D.C.: World Bank.

Little, Peter. 1997. Income and Assets as Impact Indicators. Assessing the Impact of Microenterprise Services (AIMS). Washington, D.C.: Management Systems International.

[6] For a good introduction to this topic, see Ravallion 1999. Download from www.worldbank.org/research/, choosing "poverty," then searching for "Ravallion" under "Policy Research Working Papers." See also Riely et al. 1999.

MAP/MSU. 1996. *Smallholder Cash Cropping, Food Cropping, and Food Security in Northern Mozambique: Research Methods*. Working Paper No. 22. Maputo: Ministry of Agriculture and Fisheries.

Minot, N. 2000. Generating Disaggregated Poverty Maps: An Application to Vietnam. *World Development* 28.2: 319-331.

Ravallion, Martin. 1999. *The Mystery of the Vanishing Benefits: Ms. Speedy Analyst's Introduction to Evaluation*. Policy Research Working Paper 2153. Washington, D.C.: World Bank.

Riely, Frank, N. Mock, B. Cogill, L. Bailey, and E. Kenefick. 1999. *Food Security Indicators and Framework for Use in the Monitoring and Evaluation of Food Aid Programs*. Washington, D.C.: Food and Nutrition Technical Assistance Project. www.fantaproject.org

Rose, D., P. Strasberg, J.J. Jeje, and D. Tschirley. 1999. *Household Food Consumption in Northern Mozambique: A Case Study in Three Northern Provinces*. MAP/MSU Research Report No. 33. Maputo: Directorate of Economics, Ministry of Agriculture and Fisheries.

Rose, D., and David Tschirley. 2000. *A Simplified Method for Assessing Dietary Adequacy in Mozambique*. MAP/MSU Research Report No. 36. Maputo: Directorate of Economics, Ministry of Agriculture and Fisheries.

Strasberg, P. 1997. Smallholder Cash Cropping, Food Cropping, and Food Security in Northern Mozambique. Ph.D. dissertation, Michigan State University.

Swindale, Anne, and Punam Ohri-Vachaspati. 1999. *Measuring Household Food Consumption: A Technical Guide*. Washington, D.C.: Food and Nutrition Technical Assistance Project. www.fantaproject.org

Tschirley, David, Donald Rose, and Higino Marrule. 2000. *Methodology for Estimating Household Income in Rural Mozambique Using Easy-to-Collect Proxy Variables*. MAP/MSU Research Report No. 38. Maputo: Directorate of Economics, Ministry of Agriculture and Fisheries.

*Work for this summary was conducted under the Food Security II Cooperative Agreement (PCE-A-00-97-00044-00) between Michigan State University and the United States Agency for International Development, through the Office of Agriculture and Food Security in the Economic Growth Center of the Global Bureau (G/EG/AFS). Supplemental funding for this research was also provided to the FS II Cooperative Agreement by the Africa Bureau, through the Food Security and Productivity Unit of the Sustainable Development Division, Productive Sector, Growth and Environment (AFR/SD/PSGE/FSP).

Tschirley is Visiting Associate Professor, Michigan State University; and Rose is Visiting Associate Professor, University of Cape Town. The views expressed in this document are exclusively those of the authors.