Looking within the household: Impacts for individual outcomes using aggregate household data*

Vincenzo Di Maro†

Preliminary and Incomplete

May 2007

Abstract

While most development policies target individuals, the estimation of the program impact on individual outcomes might be *a priori* impossible due to lack of individual level data on key variables, such as food consumption. This paper deals with an approach to estimate the program impact on individual impacts when only aggregate household data is available. A methodology for inferring individual outcomes from household level data is described and used to estimate the impact of an intervention (PROGRESA) on individual caloric intake. Our main results show that the program is having a stronger impact at younger ages (both for males and females) and for females up to age 30 (mothers). These findings are remarkably in line with the program's design. Another issue explored here is whether there is asymmetric information within the household as regards food consumption. Preliminary evidence show that women might not have a complete information on food intakes of adult male members within the household.

Keywords: Program Impact, Nutrient Intake, PROGRESA, Semiparametric methods

*I thank Orazio Attanasio and Nithin Umapathi for useful suggestions and discussions. Comments from seminar participants at UCL, London and University Parthenope, Naples have been very helpful.

†University College London

1 Introduction

Most development objectives focus on the well-being of individuals. The welfare of an individual is largely based on the set of economic and social interactions in which he/she is involved. These interactions can affect, and be affected by, the creation, existence, and dissolution of institutions within which the individual is situated, of which family and households are the most important ones. Within family and households we commonly refer to the processes of allocation of resources among individuals and the outcomes of those processes as *intrahousehold resource allocation*.

Development policies, while commonly targeting individuals, might be missing two important points as far as the individual dimension is concerned. At the design stage policies do not always acknowledge the intrahousehold resource allocation. One of the main results of the vast literature[1] on resource allocation models predicts that neglecting patterns of intrahousehold inequalities can lead to policy failure[2].

Another problem arises at the interventions' evaluation stage: even if the program is targeting individuals, it might be impossible to estimate the impact of the program on individual outcomes simply because of lack of data at the individual level, with this being especially true for nutritional outcomes. Obviously, the choice of collecting household level data (and not individual data) is commonly dictated by reasons that are not directly related to the program's evaluation process: household budget surveys are less expensive and require less time for the interviews; individual surveys might be too intrusive, for example as regards eating habits; information collected in household budget surveys might be more comparable between countries.

This paper focuses on this second issue since it shows how it is possible to estimate the program impact on individual outcomes when only aggre-

[1] See the review in Haddad, Hoddinott and Alderman, 1997

[2] Examples are in Haddad and Kanbur (1992), Pelletier, Msukwa and Ramakrishan (1991), Senauer and Garcia (1992), Beaton and Ghassemi (1982), Kennedy and Alderman (1987), Apps and Savage (1989) and Cox and Jimenez (1992)

gate household data is available. In particular, a methodology for inferring individual outcomes from aggregate household data, introduced by Chesher (1997), is used to estimate the impact of PROGRESA[3] on individual caloric intake. Key features of the PROGRESA's design and evaluation sample make it very suitable for the purposes of this paper: first, while all the members of the household can potentially benefit from the program's benefits, PROGRESA has a particular focus on improving educational, health and nutritional status of specific members within the household: children and pregnant and lactating mothers. In addition to this and as regards nutrition outcomes, consumptions of food are only observed at the household level in the evaluation sample. Another key aspect refers to the assignment of the program that is completely exogenous being based on randomization between treatment and control localities.

The positive and significant impact of PROGRESA on average household caloric availability is an established result in the literature: Hoddinott and Skoufias (2004) find that by November 1999, beneficiary households in treatment localities obtained 6.4% more calories than did comparable households in control localities. Here we try to shed some light on how this increased caloric availability is shared within the household. Our results show that the program is having a stronger impact on caloric availability for younger ages (both male and females) and for females up to age 30 (mothers). This result is remarkably in line with the very program design, which focuses particularly on welfare of children and their mothers.

One of the findings motivated the analysis of another issue explored here: we find that while the estimated calorie-age profile for females displays reasonable values in terms of per capita daily calorie, the profile for the sample of males shows unreasonably low values (especially for adults; see results in section 4). We argue that this can be the result of asymmetric information within the family regarding food consumptions. In particular, when the respondent to the survey questions is female (in our sample 85% of respon-

[3]PROGRESA, now known as *Oportunidades*, is an intervention targeting poor households in rural Mexico

dents) she might not hold a complete information of intakes of adult male members of the household and therefore she might be understating their food consumption. Some preliminary evidence confirms this pattern.

The rest of the paper is organized as follows. Section 2 describes PROGRESA's program design and evaluation sample. The methodology for inferring individual outcomes when only household level data is available is described in section 3. Section 4 present and discusses our results. In section 5 we present some evidence of the possible presence of asymmetric information regarding food consumptions. Finally, section 6 draws some concluding remarks.

2 Progresa (now *Oportunidades*): program and evaluation sample

The Mexican governments started PROGRESA in 1997 with the aim of improving educational, health and nutritional status of poor households in rural areas. The program provides cash transfers linked to children's enrollment and regular school attendance and to regular visits to health centers.
As regards specifically the program's nutritional component[4] special attention is given to the prevention of malnutrition in infants and small children, which is a crucial determinant of their future development. The program provides food supplements to pregnant and lactating women and to children between two and five years of age if any signs of malnutrition are detected.[5] In addition beneficiaries are required to attend nutrition and health lectures (known as *platicas*[6]), which are mainly directed to mothers.

Our estimates will be exploiting one key features of PROGRESA: randomization between treatment and control group of the localities in the evaluation

[4]For more details on the other program's components, see Skoufias(2005)

[5]Also children living in non-PROGRESA households will receive the supplement if malnutrition is detected

[6]Up to 25 themes are discussed, including nutrition, hygiene, infectious diseases, immunization, family planning, and detection and prevention of chronic disease

sample. This sample consist of 24077 households from 506 localities divided in 320 treatment and 186 control localities and in poor household (52% eligible) and not poor (42% not eligible). The eligibility status was decided in 1997 before the start of the program and it is based on household socio-economic characteristics. Only eligible households in treatment localities receive the program. As regards randomization, Berhman and Todd (1999) examines the characteristics of treatment and control groups in terms of age, education, access to health care and income and conclude that the groups are very similar and do not indicate any systematic differences at locality level.

Households in the sample have been interviewed every six months between March 1998 and November 2000 and then again in 2003; here we use the wave November 1999, that is 2 years after the program's inception. See table 1 for some sample's descriptive statistics and figure 1 for the age-sex structure of our sample.

Table 1: Evaluation sample statistics

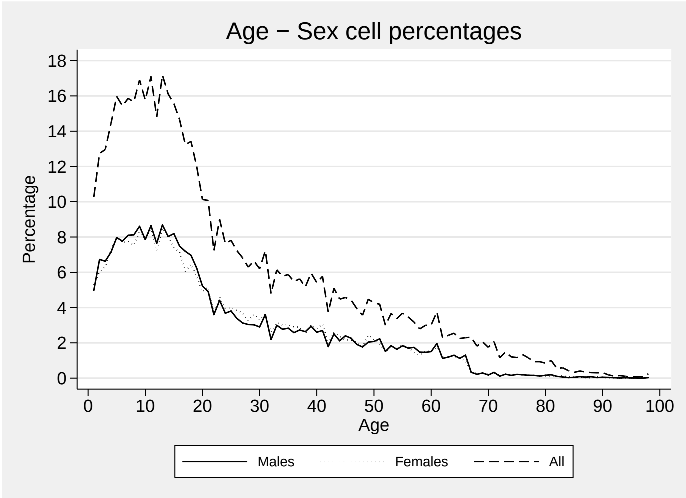
	Treatment		**Control**	
	Poor	**Not poor**	**Poor**	**Not poor**
Number of households	7837	7019	4682	4539
Number of localities	320		186	
Average size of locality (households)	46.42		49.57	

3 Inferring individual data from household-level data

The method developed here draws on Chesher (1997), which stresses that this is the only method available for estimating individual nutrient consumption when only household totals are available.

First, we model the household food acquisition process: the average rate of nutrient consumption by individual p conditional on household composi-

Figure 1: Age-Sex structure



tion and other characteristics can be expressed as:

$$E(C_p|x,z) = f(x_p,z) \tag{1}$$

where x_p denotes the characteristics of person p and z denotes household characteristics, such as region of residence and income.

Household nutrient consumption C is then the sum of individual consumption, as in 2

$$E(C|x,z) = \sum_{p=1}^{P} f(x_p,z) \tag{2}$$

When only data for aggregate household nutrient consumption we do not observe the individual consumptions $f(x_p,z)$, However, at the household level the following moment conditions holds

$$E\left\{\left[C - \sum_{p=1}^{P} f(x_p,z)\right] g(x,z)|x,z)\right\} = 0 \tag{3}$$

where $g(x,z)$ is an arbitrary function.

Condition (3) allows us to estimate the individual consumption function. The first step in the estimation is to make a simplifying assumption about the function. In particular, we assume that $f(x_p, z)$ is a multiplicatively separable function of individual and household characteristics, or $f(x_p, z) = f(x_p)u(z)$, and we allow allow for different functions for male and females:

$$f(x_p) = s_p f_M(age_p) + (1 - s_p) f_F(age_p)$$

where $s_p = 1$ if person p is male and $s_p = 0$ otherwise and $f_M(.)$ and $f_m(.)$ are age-intake functions for respectively males and females.

The objective of estimating $f(x_p)$ can be pursued through a non linear approach. A first convenient simplification is to approximate $f(x_p)$ by a step function with points of increase at integer years of age. This task can be performed defining the following vector of binary indicators for each gender

$$w_p = (w_{p,0}, ..., w_{p,97})$$

where $w_{p,a} = \mathbf{1}_{[a \leq a_p < a+1]}$ and the binary indicators are defined for $a = 1, ..., 97$ since ages recorded in our sample span this range.

Accordingly, the calorie consumption-age relationship can be approximated by the discrete form $f(a_p) = w_p^{'} \beta^s$ where $\beta^s = (\beta_1^s, ..., \beta_{97}^s)$ is a vector of age- and sex-specific average intakes. The model for the average household calorie consumption can then be written as follows

$$E(C|x, z) = \left[\beta_0 + \sum_{p=1}^{P} \{s_p w_p^{'} \beta^M + (1 - s_p) w_p^{'} \beta^F\}\right] exp(z^{'}\gamma) \tag{4}$$

or, in more compact form

$$E(C|x, z) = (\beta_0 + n_M^{'} \beta^M + n_F^{'} \beta^F) exp(z^{'}\gamma) \tag{5}$$

where n_s is a vector containing counts of household members of sex s at each integer year of age.

Finally, a consistent estimator of β and γ can be obtained by using a non-linear least squares estimator, as defined in (6) where we use household income as household characteristic z_h

$$(6) \qquad \arg\min_{\gamma,\beta^M.\beta^F}\left[\sum_{h=1}^{H}\{C_h-(\beta_0+n_M^{'}\beta^M+n_F^{'}\beta^F)exp(z_h^{'}\gamma)\}\right]$$

where h identifies households $h = 1, ..., H$.

Consistent estimators of γ,β^M and β^F can be obtained by standard non-linear least squares methods. One problem with this approach is that it is likely that the estimated calorie-age profile will exhibit too much variability (or more variation than the one implied by the underlying nutrient-age relationship) with this happening because of the estimator above fitting a curve only taking into account goodness-of-fit. Since we are interested also in the smoothness of the estimated relationship a roughness penalty approach is employed.[7]

As regards the validation of this procedure, Naska, Vasdekis and Trichopoulou (2001) have compared age-gender specific food availability based on data collected at the household level with individual nutrition surveys for the same population finding that the individualization procedure seems to work quite well in practice.[8]

4 Results

The estimation approach described in 3 allows us to estimate the individual consumptions $f(x_p, z)$. In particular, we use age and sex as individual characteristics x_p and household income as household characteristic z, therefore

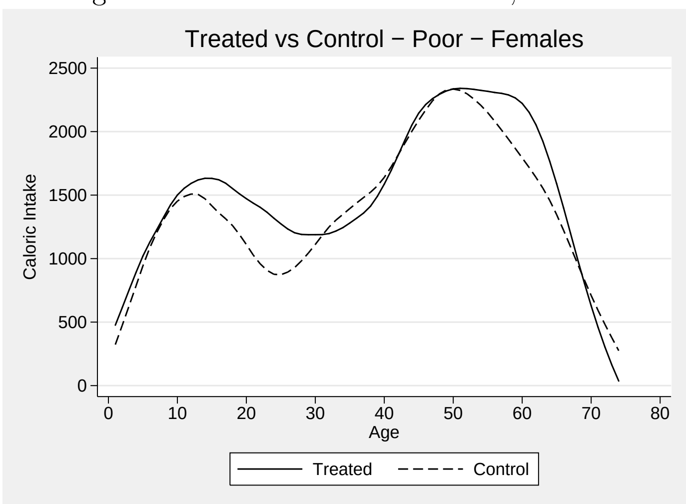
[7] This approach tries to compromise between the two, often conflicting, aims in curve fitting: goodness-of-fit and smoothness. For more details see Green and Silverman (1994) on the approach in general and Chesher (1997) for implementation issues.

[8] The study uses household budget surveys and individual nutrition surveys of four European countries: Belgium, Greece, Norway and UK

the output of our estimation is a calorie-age profile for females and males, respectively.
This is repeated both for the treated (households eligible for the program, that is poor, and living in PROGRESA communities) and control group (poor and not in PROGRESA communities). The comparison of the treatment and control groups' profiles will deliver an estimate of the PROGRESA impact on individual calorie consumption.

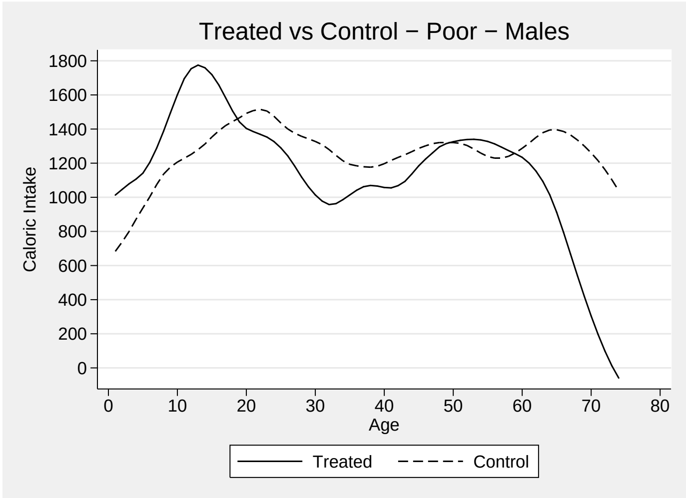
In figure 2 and 3 we report the estimated relationship between per capita daily caloric consumption and age both for the treated and control group respectively for females and males. One main pattern that emerges is that for younger ages the estimated calorie intake for treated group is higher than control group's. Another interesting finding is that while the estimated values for caloric intake for females (see figure 1) are not far from meaningful values (for example, daily recommended intake for girls age 0-5 is around 890 kcal and 2000 kcal for adults), the values for males (see figure 2) are not in line with benchmark intakes (recommended intake for a male age 15-19 is 3000 kcal). This issue is explored further below.

Figure 2: Estimated Caloric Intake, Females



The program impact (in our case treatment effect on the treated, TT) is

Figure 3: Estimated Caloric Intake, Males



the difference at each age point between the estimated caloric consumption of treatment and control group. We compute this difference both for females (see figure 4) and males (see figure 5) together with 2 standard error pointwise confidence bands.

Our main findings are that program seems to have a positive and sizeable impact only at younger ages, both for males and females. Particularly for females positive impact lasts till age 30. These results are remarkably in line with the very program design: PROGRESA wants to have an impact on the nutritional status of poor families, particularly of children and their mothers.

As a robustness check we also re-estimate the calorie-age profile not allowing any longer for a different function between males and females; the estimated calorie intake profile is in figure 5 and the impact in figure 6. These pooled results are consistent with findings above: impact is positive and significant only for younger ages.

5 Asymmetric information on food consumption?

One issue left unaddressed above is the fact that for the male calorie-age estimated profile we find values that are not close to reasonable ones (for ex-

Figure 4: Impact, Females

Figure 5: Impact, Males

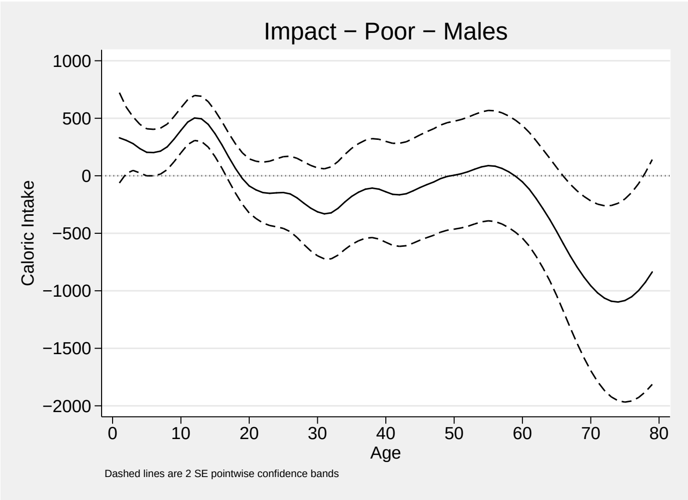


Figure 6: Estimated Caloric Intake, Pooled

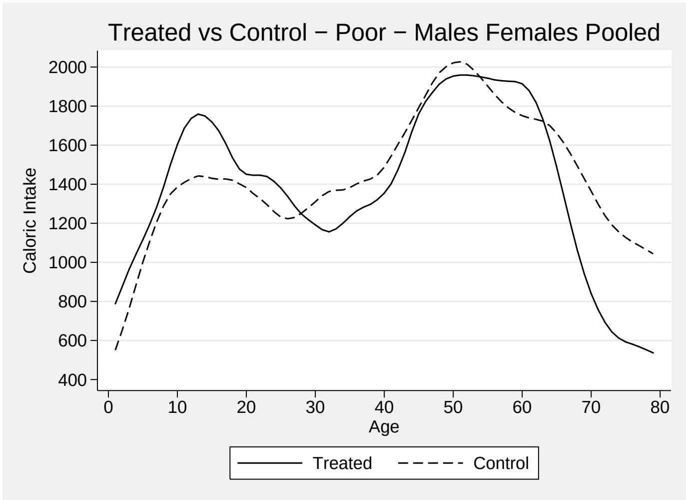
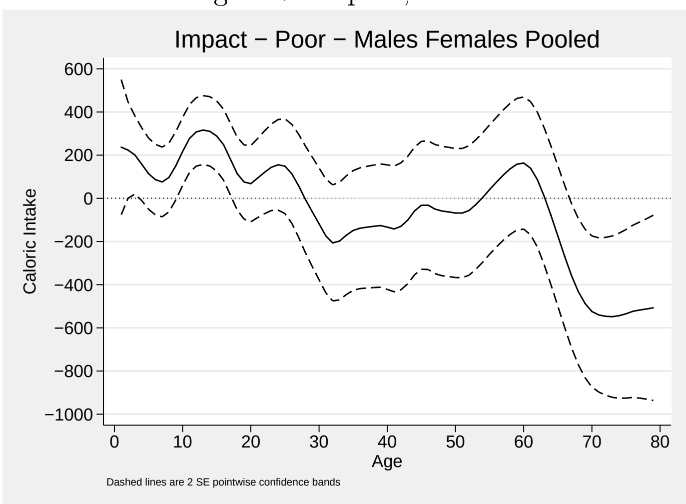


Figure 7: Impact, Pooled



ample to recommended daily intakes). One possibility is that some type of measurement issue is biasing the estimates for our sample of males. Due to randomization we can safely assume that this possible bias is affecting in the same way treated and control group, with this meaning that the estimates of program impact are still unbiased. However, it is interesting to explore further the issue.

One possible explanation of this observed pattern is that there is asymmetric information regarding food consumptions; respondent (to food consumption questions) might not have a good information of household activities made by specific age-sex groups within the family. Our sample is very "asymmetric" in terms of who responds since respondent is a woman in 85% of questionnaires. Hence, it might be that respondent women do not hold a complete information on intakes of adult male members within the household and therefore they are understating their food consumption. A previous study (Boozer and Goldstein, 2003) explores a similar issue with data from Ghana where husbands and wives were interviewed separately and each respondent was asked to report its own expenditure, the expenditure of their spouse (cross-reporting), and the expenditure of any other person in the household that was used for household consumption. A major finding is that some components of consumption are "private" in nature, and thus essentially unobserved in the cross reports.

We try to test whether this measurement issue is present in our sample with a simple (and preliminary) strategy: for each sub-sample (females and males) we repeat the estimation allowing for a further disaggregation: "only female respondents" and "only male respondents". In figure 8 and 9 is then possible to assess whether and how the estimated profile changes according to the sex of the respondent. One interesting finding is that when respondent is male the shape for adults is different both for the female and male sample. In particular, the estimated caloric intake for the sample of males is substantially higher(somewhat closer to reasonable values) for adults when the respondent is male (see figure 9). Another finding is that the shape of the profile do not seem to change with the sex of the respondent for younger ages (below age 20) and for elderly people (above age 60).

Figure 8: Calorie-Age profile, 3 Samples, Females

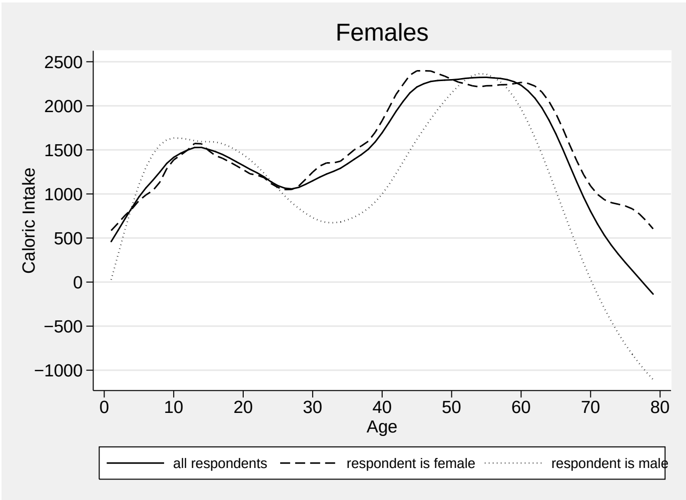
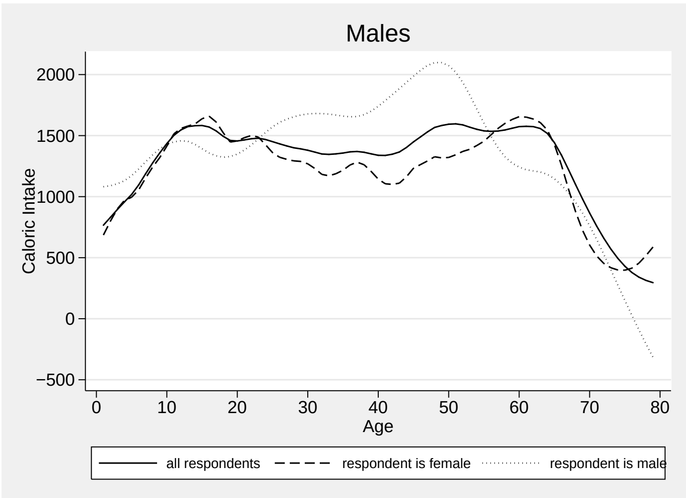


Figure 9: Calorie-Age profile, 3 Samples, Males



In conclusion, we find some evidence of under-reporting of food intake (caloric intake) of other-sex adult members in the household. In particular, since most of the respondents in our sample are females, women seem to have distorted information on food intake of male adults, with this explaining the unreasonable low values we find for the estimated calorie-age profile for males. We are aware that this explanation is tentative and preliminary (many selection issues need to be addressed), however it seems to make a case for further research on this issue.

6 Conclusion

This paper studies the impact of interventions on specific members within the household. While most programs target individuals, estimating the impact at the individual level might be *a priori* impossible due to lack of individual level data on key variables, such as food consumption. We show here how it is possible to estimate the program impact on individual outcomes when only aggregate household data is available. In particular, we employ a methodology for inferring individual outcomes from household level data in order to estimate the impact of PROGRESA on individual caloric intakes.

Our main results show that the program is having a stronger impact at younger ages (both for males and females) and for females up to age 30 (mothers). PROGRESA, by its very design, tries to improve the educational, health and nutritional status of poor rural household with a primary focus on children and pregnant and lactating women, therefore our findings suggest that the program is remarkably achieving its targets.

Another issue explored here is the possible presence of asymmetric information within the household as regards food consumption. In particular, women (85% of respondents in our sample are females) might not have a complete information on food intakes of adult male members within the household. We find some (preliminary and tentative) evidence that women respondents understate adult males' food consumption.

References

Apps P. and E. Savage, 1989, Labour supply, welfare rankings, and the measurement of inequality, *Journal of Public Economics* 39(3):335-364

Beaton G. and H. Ghassemi, 1982, Supplementary feeding programs for young children in developing countries, *American Journal of Clinical Nutrition* 34(4)(supplement):836-916

Berhman, J.R. and P.E. Todd, 1999, Randomness in the Experimental Samples of PROGRESA, International Food Policy Research Institute, Washington, D.C.

Boozer, M. and M. Goldstein, 2003, Poverty Measurement and Dynamics, Mimeo Yale University

Chesher, A., 1997, Diet Revealed? Semiparametric Estimation of Nutrient Intake-Age Relationships, *Journal of the Royal Statistical Society* Series A (Statistics in Society), Vol. 160. No. 3, 389-428

Cox D. and E. Jimenez, 1992, Social security and private transfers in developing countries: The case of Peru, *World Bank Economic Review* 6(1):155-171

FAO - LATINFOODS, 2002, Tabla de Composición de Alimentos de América Latina, http://www.rlc.fao.org/bases/alimento/busca.asp

Green, P. J. and B. W. Silverman, 1994, *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman and Hall, London

Haddad L., Hoddinott J. and A. Alderman, 1997, *Intrahousehold Resource Allocation in Developing Countries. Models, Methods, and Policy*, IFPRI, John Hopkins University Press, Baltimore and London

Haddad L., and R. Kanbur, 1992, Intrahousehold Inequality and the Theory of Targeting, *European Economic Review* 36 (2-3, April):372-78

Kennedy E. and H. Alderman, 1987, Comparative analyses of nutritional effectiveness of food subsidies and other food-related interventions, IFPRI Research Institute and WHO-UNICEF, Washington D.C

Pelletier D., Msukwa L. and U. Ramakrishan, 1991, Nutrition in Project Planning: Intrahousehold Risks and Determinants, *Food Policy* 16(2): 127-139

Senauer B. and M. Garcia, 1992, An Intrahousehold Analysis of a Philippine Food Subsidy Program, Paper presented at the IFPRI/World Bank Conference on Intrahousehold Resource Allocation, 12-14 Feb, Washington D.C. Mimeo

Vasdekis, V.G.S., Stylianou, S. and A. Naska, 2001, Estimation of age- and gender-specific food availability from household budget survey data, *Public Health Nutrition* 4(5?), 1149-1151