Assessing poor or non-poor bias in the criteria used for selecting sample households for the poverty analysis of the Malawi Integrated Household Survey, 1997-98

The Integrated Household Survey (IHS) was a comprehensive socio-economic survey of the living standards of households in all districts of Malawi. The National Statistical Office administered the IHS questionnaire to about 12,900 households over a 12 month period, November 1997 to October 1998. The data was cleaned between May 1999 to April 2000. 10,698 households remained in the data set when the 'c2' version of the data was released in early May 2000. However, comprehensive and reliable information on consumption and expenditures is not available for all of these households. Only IHS households for which we have reliable information will be used in the poverty analysis.

An earlier report entitled *Criteria used for selecting sample households for the poverty analysis of the Malawi Integrated Household Survey, 1997-98* presented the selection criteria used to arrive at a sub-set of 6586 IHS households which could confidently be used in the analysis. A common comment made by reviewers of the earlier document was that some indication should be provided as to whether the households being dropped were significantly different in their poverty status from those which were retained for the analysis.

Household characteristics

Expenditure and consumption data cannot be used to assess a poor or non-poor bias in the households being dropped: They are being dropped because the information we have on them in this regard is judged to be of poor quality. Consequently we must use other household characteristics. A set of twenty-one household characteristics which have been shown in the past to have a strong correlation with the poverty status of Malawian households were extracted from the IHS data set for use in this assessment. These are described in Table 1.

In the same table an *a priori* judgment is made as to whether a poor or a non-poor bias is indicated if the mean value for the characteristic for one sub-set of households is significantly higher than the mean for the other. For example, it is expected that poorer households are disproportionately female-headed. Thus, if the proportion (the mean of the dummy variable) of female headed households in the set of household dropped using a particular selection criterion is significantly higher than the proportion of female households in the retained households, the dropped households are judged likely to be disproportionately poor. Consequently, a non-poor bias would result in the data set retained for analysis.

These judgments as to the bias which can be inferred from the differences in the means are clearly open for discussion and debate as they represent an informed but nonetheless subjective assessment.

Selection criteria

Households were dropped based on four separate criteria. These four criteria were applied in a step-wise manner to the initial 10,698 households making up the c2 data set of the IHS:

1. IHS households which were judged by the team responsible for cleaning the IHS to have obviously flawed expenditure or consumption information, but otherwise good information.

Table 1: Household characteristic variables used in bias assessment of selection criteria

Variable	Variable type	Description	Implication of higher mean
FEMHHH	Dummy	Female-headed	Poorer
HHH25_60	Dummy	Head aged 25 to 60	Less poor
MARRIED	Dummy	Married head	Less poor
DPNDRAT	Continuous	Dependency ratio	Poorer
HHHED	Rank	Highest education level for head	Less poor
HHHLIT	Dummy	Literate head	Less poor
MOTHERED	Rank	Educational level of mother	Less poor
BIRTHRTE	Continuous	Years between births for mothers	Less poor
PROPDEAD	Continuous	Proportion of children born in HH who are now dead	Poorer
HAZLT2	Dummy	Stunted child in HH	Poorer
PCLAND	Continuous	Per capita acres cultivated	Less poor
HYBMAIZE	Dummy	Grows hybrid maize	Less poor
TOBACCO	Dummy	Grows tobacco	Less poor
FERTUSE	Dummy	Uses fertilizer	Less poor
PCLVSVAL	Continuous	Value of livestock owned (MK)	Less poor
PCINC	Continuous	Monthly per capita cash income (MK)	Less poor
GOVTEMP	Dummy	HH member employed by government	Less poor
OWNTAP	Dummy	Water source is own tap	Less poor
BICYCLE	Dummy	Bicycle owned by HH member	Less poor
RADIO	Dummy	Radio owned by HH member	Less poor
PCCLOTH	Continuous	Per capita value of clothing purchased past 3 months (MK)	Less poor

1621 households were selected to be dropped, with 9077 retained.

2. Households for which the diary of expenditures was not kept for the required 14 days. This indicated enumerator error in data collection. Applied to the 9077 households remaining from the first selection step, 281 households were dropped, with 8796 retained.

1. Households which reported per capita daily calorie consumption of greater than 5000 calories were dropped. Recommended adult daily calorie requirements in Malawi are between 2000 and 3000. Consumption by household members of more than 5000 calories per day over several days is clearly unreasonable. Applying this selection criterion to the 8796 households resulted in 742 dropped households, with 8054 retained.

2. Households which reported daily per capita calorie consumption of less than 500 were dropped. While it is possible for a small number of IHS households to have correctly reported such low values – subsistence producers with no food stocks to consume and who made few food purchases over the at least 14 days in which they were surveyed – in most cases such low reported calorie consumption is due to error. When applied to the 8054 households remaining, 1468 households were dropped, while 6586 remained.

A comparison of the characteristics of the final 6586 to all households dropped from the original 10,698 household data set was also conducted.

Methodology

At each stage of the selection process a simple means comparison (t-test) statistical procedure was carried out on the retained and the dropped sets of households on each of the variables. A 0.05

probability level was used to judge whether the difference of the means was significant. Following the step-wise application of the selection criteria, a comparison was made of the characteristics of all of the households dropped to those of the households retained for the poverty analysis.

Results

The results are shown in Table 2 below. For the first three selection criteria applied, while there are significant differences between the means for more than half of the variables, there is no clear trend. The bias which one would infer is not consistent across variables: As many variables indicate that the dropped households are disproportionately non-poor as indicate that they are poor.

In the case where households were dropped if they reported calorie consumption of less than 500 per capita per day, however, the evidence is quite strong that the households dropped under this criterion are uncharacteristically poor. For the 16 variables for which there is a significant difference in the means between the dropped and retained households, 15 of the 16 indicate that the dropped households are poorer.

The results of the comparison of all households dropped to those retained also shows that the dropped households are likely to be poorer than the households retained for the analysis. While for a slight majority of the variables there was no significant difference in the means, there are significant differences between the dropped and retained households, particularly in the sex of the household head and in the education of the household head or his wife (where the head is male). Only two variables indicated that the dropped households might be wealthier than those retained.

Implications

The households whose consumption and expenditure is used to compute the poverty line are somewhat wealthier than the population as a whole. How much wealthier they might be is unclear. If we had adequate consumption and expenditure data on all of the households, given that for many of the variables the difference in means was not significant, one would expect that the overall difference in the welfare indicator means for the two groups would not be very large.

In any case, the effect of the non-poor bias in the analytical data set on the derivation of the poverty line should be of little consequence. The poverty line is derived using a basic needs approach anchored to the recommended daily calorie requirement of individuals in the sample households. Households are judged poor if they are not meeting their recommended daily calorie requirements, plus an allowance for non-food consumption. The resultant poverty line should be consistent whether 30 percent, 50 percent, or 80 percent of the households in the analytical data set have consumption and expenditure levels which would place them below the poverty line derived through the analysis.

Although the poverty line will be robust regardless of the proportion of the households analyzed which are poor, the resultant poverty head count which one derives for the nation from this non-poor biased data set will be erroneous: A lower poverty head count than is likely the case will result.

To determine an accurate poverty head count, a proxy welfare indicator – per capita daily consumption and expenditure in Malawi kwacha – will be assigned to each of the 4112 dropped households. This will be done by undertaking a regression analysis on the characteristics of the 6586 retained households, using their actual welfare indicator as the dependent variable. The resultant model will be applied to the dropped households using the same characteristics to derive a proxy welfare indicator for these households. Making use of their proxy welfare indicators with the poverty

line derived from the analysis of the 6586 households, the poverty status of these 4112 households will be determined. A poverty head count for the nation as a whole then will be derived from the complete IHS data set of 10,698 households.

The import of the analysis presented in this paper is that the poverty head count for Malawi based on the 10,698 household IHS sample should be slightly higher than that which was derived from the analysis of the 6586 households – 59.6 percent. As the dropped households are somewhat poorer than the 6586 households, their inclusion should increase the poverty head count somewhat.

Table 2: Results of the means comparisons on selected poverty related household characteristics between dropped and retained households

Household characteristic	HHs dropped stepwise - due to clearly unreliable consumption and expenditure information			HHs dropped stepwise - diary of expenditures kept for less than the required 14 days			HHs dropped stepwise - unreasonably high per capita daily calorie consumption (> 5000)			HHs dropped stepwise - unreasonably low per capita daily calorie consumption (< 500)			Comparison of HHs remaining (6,586 HHs) to all those dropped		
	Mean - Retained HHs	Mean - Dropped HHs	Bias in Dropped HHs	Mean - Retained HHs	Mean - Dropped HHs	Bias in Dropped HHs	Mean - Retained HHs	Mean - Dropped HHs	Bias in Dropped HHs	Mean - Retained HHs	Mean - Dropped HHs	Bias in Dropped HHs	Mean - Retained HHs	Mean - Dropped HHs	Bias in Dropped HHs
Female-headed	0.25	0.25	None	0.25	0.27	None	0.25	0.24	None	0.24	0.29	Poor	0.24	0.26	Poor
Head aged 25 to 60	0.76	0.77	None	0.76	0.81	Non-Poor	0.77	0.72	Poor	0.77	0.77	None	0.77	0.76	None
Married head	0.74	0.73	None	0.74	0.68	Poor	0.75	0.64	Poor	0.76	0.72	Poor	0.76	0.71	Poor
Dependency ratio	1.05	1.15	Poor	1.05	0.99	None	1.10	0.55	Non-Poor	1.07	1.20	Poor	1.07	1.05	None
Highest education level for head	1.50	1.32	Poor	1.49	1.87	Non-Poor	1.47	1.65	Non-Poor	1.52	1.27	Poor	1.52	1.40	Poor
Literate head	0.73	0.69	Poor	0.73	0.78	None	0.73	0.72	None	0.75	0.65	Poor	0.75	0.69	Poor
Educational level of mother	0.94	0.91	None	0.93	1.22	Non-Poor	0.94	0.84	Poor	0.96	0.82	Poor	0.96	0.89	Poor
Years between births for mothers	5.07	5.09	None	5.06	5.64	Non-Poor	4.99	6.19	Non-Poor	5.03	4.81	Poor	5.03	5.15	None
Proportion of children born in HH who are now dead	0.17	0.18	None	0.17	0.16	None	0.17	0.22	Poor	0.17	0.15	Non-Poor	0.17	0.17	None
Stunted child in HH	0.20	0.23	Poor	0.20	0.17	None	0.22	0.09	Non-Poor	0.21	0.22	None	0.21	0.20	Non-Poor
Per capita acres cultivated	0.52	0.59	Non-Poor	0.53	0.44	None	0.50	0.83	Non-Poor	0.52	0.39	Poor	0.52	0.55	None
Grows hybrid maize	0.29	0.01	Poor	0.29	0.15	Poor	0.29	0.29	None	0.32	0.20	Poor	0.32	0.14	Poor
Grows tobacco	0.14	0.24	Non-Poor	0.15	0.00	Poor	0.15	0.13	None	0.16	0.09	Poor	0.16	0.15	None
Uses fertilizer	0.29	0.47	Non-Poor	0.29	0.12	Poor	0.29	0.28	None	0.32	0.16	Poor	0.32	0.30	Poor
Value of livestock owned (MK)	345.05	369.26	None	336.52	612.00	Non-Poor	304.98	678.85	Non-Poor	323.39	222.38	Poor	323.39	389.28	None
Monthly per capita cash income (MK)	181.66	308.41	Non-Poor	176.10	355.79	Non-Poor	149.54	464.36	Non-Poor	149.60	149.28	None	149.60	282.98	Non-Poor
HH member employed by government	0.07	0.04	Poor	0.07	0.12	Non-Poor	0.06	0.08	None	0.07	0.06	None	0.07	0.06	None
Water source is own tap	0.06	0.03	Poor	0.06	0.12	Non-Poor	0.06	0.08	None	0.06	0.05	None	0.06	0.05	Poor
Bicycle owned by HH member	0.31	0.38	Non-Poor	0.31	0.23	Poor	0.31	0.32	None	0.32	0.26	Poor	0.32	0.31	None
Radio owned by HH member	0.07	0.07	None	0.07	0.14	Non-Poor	0.07	0.07	None	0.07	0.06	Poor	0.07	0.07	None
Per capita value of clothing purchased past 3 months	94.68	1.67	Poor	92.84	152.31	Non-Poor	83.27	196.79	Non-Poor	89.88	53.57	Poor	89.88	65.70	Poor
Number of households	*9,077*	*1,621*		*8,796*	*281*		*8,054*	*742*		*6,586*	*1,468*		*6,586*	*4,112*	
Bias summary:		**None**	**8**		**None**	**6**		**None**	**9**		**None**	**5**		**None**	**10**
		Poor	**8**		**Poor**	**5**		**Poor**	**4**		**Poor**	**15**		**Poor**	**9**
		Non-poor	**5**		**Non-poor**	**10**		**Non-poor**	**8**		**Non-poor**	**1**		**Non-poor**	**2**
			21			**21**			**21**			**21**			**21**