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.

The food component of the poverty line resulting from the poverty analysis will be anchored to the per capita daily calorie consumption of households in the data set. Data on food consumption in the IHS was collected in two sections:

- Non-cash food consumption by 3-day recall at the time of the main household interview was registered in the K-1 section of the questionnaire.
- Cash acquisition of food was recorded in the diary of expenditures. Households were to be visited by the enumerator every three days to record cash purchases, both food and non-food, since the enumerator's last visit. If the diary could not be kept for the desired 28 days, enumerators were to ensure that the diary be kept for at least 14 days.

In order to capture the full food consumption of the household, the enumerator had to be diligent over the entire survey period. The IHS involved the use of 60 enumerators nationwide. Field supervision was unable to be as close as might have been desired. Consequently, there are several enumeration areas where the enumerator did not do a particularly good job in recording food consumption. Households from these areas should not be included in the poverty analysis, as their reported food consumption is not a true reflection of their actual consumption, both in cash and calorie terms. The selection criteria presented in this note are used to identify these households with poor food consumption information in order to exclude them from the poverty analysis.

Deriving the poverty analysis data set from the IHS c2 data set – first steps

The analysis started with the calorie consumption data for the 10,698 households of the c2 version of the cleaned IHS data set. Three selection criteria were immediately imposed:

- All household reporting no calories consumed were dropped from the data set. 1,586 households reported no calories consumed, leaving a data set of 9112 households.
- All household for which calorie computation problems were experienced in the K-1 section were dropped, thirty households in all. This resulted in a data set of 9082 households.
- An SPSS command file which identified the records of all households for which inconsistencies in expenditures and consumption or strong evidence of enumerator error was run on the data set. Five additional households were dropped as a result, leaving 9077 households.[1]

[1] This file, 'flagK1.sps' was assembled by Ellen Payongayong, a research associate from IFPRI who provided NSO with technical assistance in cleaning the data to derive the c2 version of the IHS. The file actually identifies 1,577 households as having problematic expenditure and consumption data. Most of the

It is felt that these criteria for selection are relatively uncontroversial. Consequently, the sensitivity analysis conducted to determine the final data set for the poverty analysis begins from this data set of 9077 households.

Deriving the poverty analysis data set from the IHS c2 data set – sensitivity analysis

First, a set of selection criteria which have been used on household poverty analysis data sets elsewhere was applied to the 9077 households.

- All households for which the diary of expenditure was kept for less than 15 days were dropped. 513 households met this criteria.
- All households reporting less than 500 calories per person per day or more than 5000 calories per person per day were dropped. The median per capita calorie requirement for the IHS households is between 2150 and 2210 daily. 1510 households reported less than 500 calories, whereas 790 households reported consuming more than 5000 calories per capita per day.
- All households which did not report consuming a staple food (grains or roots and tubers) were dropped. 585 households met this criteria.

As some households met more than one of these criteria, the total number of households dropped was 2763, resulting in a data set of 6314 households.

These three criteria were then subjected to a step-wise sensitivity analysis to assess whether they were reasonable.

Diary of expenditure: days kept

The 15 day cut-off was judged to be too stringent by the NSO staff involved with the IHS. The enumerators had been asked to keep the diary for a minimum of 14 days. Relaxing this criterion to 14 days from 15 increased the sample size by 232 households to 6550. Imposing the minimum of 14 days for the diary criterion on the 9077 household data set reduces it by 281 households to 8796.

An assessment was made as to whether there was scope to reduce the diary days criterion even lower. However a district assessment of the 281 households for which the diary of expenditures was kept for less than 14 days showed that 83% of such households were found in only five of the 29 IHS district strata. This was taken as clear evidence of enumerator error in those districts. 14 days was judged the appropriate cut-off.

Household per capita calorie consumption bounds

The 5000 calorie upper bound was accepted. Additional investigations will need to be made of the food baskets of the households excluded under this selection criterion to determine whether certain food-quantity to calorie conversions were in error. Bulk food purchases that were not caught in cleaning may account for the high calorie consumption. Also under-reporting of household members would lead to much higher per calorie consumption figures than are realistic.

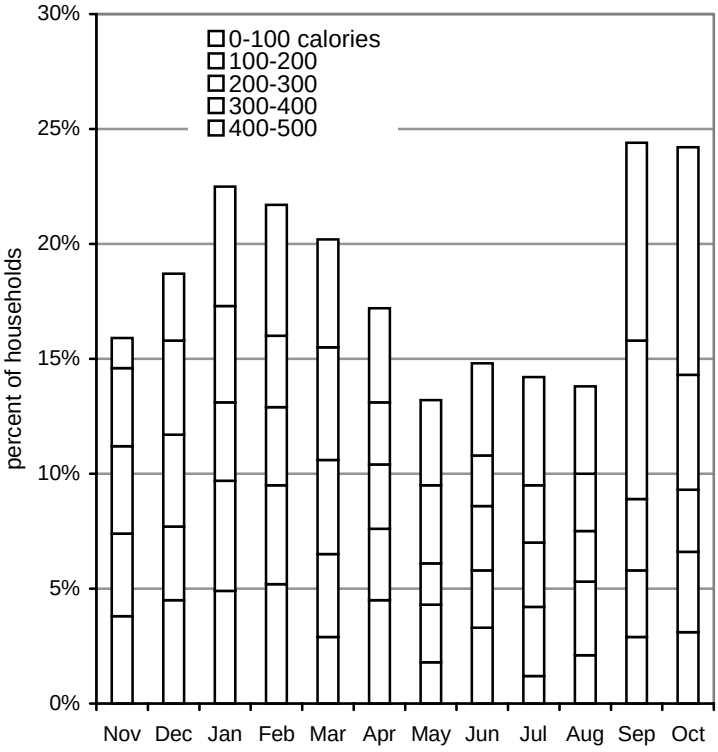
The lower limit of 500 calories was evaluated to judge whether there might be any seasonal effect which accounts for low calorie consumption by households. Specifically, do we find more households reporting low calorie consumption in the hungry months of November through February?

---

households listed in this file were already dropped here due to reporting no calorie consumption. This same file was used in the data cleaning of the K-1 non-cash food file and is the main reason that these households report no calories from K-1. In sum, this step and the first selection step are very closely correlated.

For a data set of 8054 households, the chart presents the percentage of households in a given month which report calorie consumption below 500 calories.[2] The stacked bars represent 100 calorie ranges. On first glance, there does seem to be something of a seasonal effect. The highest levels of low calorie consumption are reported in September and October as the hungry period following the IHS survey period is starting. However, the hungry season of 1997-98 can be seen as well.

Percent of households interviewed in a given month with consumption of less than 500 calories per capita per day



Yet, in examining the stacked bars, it becomes apparent that most of the seasonal component in the graph is due to those households who reportedly consume less than 200 calories per person per day. These households very likely have erroneous data, as this level of calorie consumption is an extreme starvation diet. While malnutrition is seen in Malawi, such starvation as this level of consumption would indicate is exceptional.[3]

If one sets aside households which consume less than 200 calories per person per day, in assessing this information one must still ask whether it is any more reasonable to expect households to consume 200 to 500 calories a day. The minimum recall period on food consumption in the IHS is three days. As noted above, the recommended daily calorie requirement is closer to 2200. Rates of 500 calories per person per day and under could only legitimately be recorded for households who make no purchases of food over at least a 14 day period, as these would be noted in the diary of expenditures, and who have very low levels of non-cash food consumption in the three days prior to their interview.

Is there any pattern in this graph which would indicate that a household reporting 300 or 400 per capita calorie consumption is valid data for the hungry season? Possibly such a pattern can be seen, but a decision to include these low calorie households would also lead to the inclusion of an unknown, but likely significant, number of households for which consumption data was not collected comprehensively and who, consequently, end up reporting calorie consumption at these low levels.

Examining the spatial distribution of low calorie reporting households by district, high frequencies are found in a handful of districts, indicating enumerator error would account for some of

[2] The 8054 household data set results from dropping from the 9077 data set those households which kept the diary for less than 14 days and which report consumption of more than 5000 calories per capita per day

[3] Both the 1996-97 and the 1997-98 cropping seasons in Malawi were average to relatively good, the first having somewhat excessive rains nationwide, while in 1997-98 there were localized patches of both excessive and low rainfall across the country. Food availability was not strongly constrained by rainfall factors in either year, however. Exceptionally poor harvests are <u>not</u> a factor in accounting for the pattern of this graph.

the low values. Moreover, the high levels of low calorie consumption in September and October, the final two months of the survey, could be indicative of enumerator fatigue. The enumerators may not have had sufficient enthusiasm for their job at this stage to go to the trouble to probe households on their food consumption over the past three days or to record all of the households food purchases.

The evidence does not rule out that lower bounds than 500 calories are not warranted. There may well be properly interviewed households who were existing on less than 500 calories a day at the time they were interviewed, but it is difficult to determine which households these are. That being the case, it seems best to adopt a conservative stance, assume that few households can and do survive on less than 500 calories per person per day for very long, and retain the 500 calories per capita per day level as the lower calorie consumption limit by which one selects valid households for the poverty analysis.

No staple food consumption

The third criterion used to select households for the poverty analysis was that households which did not report consuming staple foods were dropped. Staple foods include maize, wheat, rice, millet, sorghum, cassava, sweet potato, Irish potato, and products made from these items.

Starting from a first-cut valid household set of 9077 households, if one deletes households with diary reporting period of less than 14 days or per capita daily consumption of greater than 5000 or less than 500 calories, a data set of 6586 households results. Of these households, 72 report not consuming any staple foods.

The spatial distribution of these households is relatively even across districts, except for Karonga district where 6.3% of households did not report a staple. The next highest rate of households not reporting staples for a district is Kasungu with 2.1%. Cooking bananas, an item not on the list of staples nor specifically identified in the IHS, are eaten in northern Karonga district. This may constitute the staple food for these households.

The temporal distribution shows that the greatest proportions of households interviewed in a given month who did not report consuming staples are found in February and March, the tail end of the hungry season. Twice the number of households not reporting staples are found in these months than in any of the other months.

The number of households not reporting having consumed staples is small. When the poverty line is derived, given that there may be good reasons for why no staples were reported for these households, a sensitivity analysis will be run to determine whether including or excluding these 72 households in the analysis leads to any significant differences in the various poverty measures. If significant differences are seen, further investigation of the characteristics of these households will be required.

Data set for poverty analysis

If the 72 households without staples are retained, a final data set consisting of 6586 households results. The sample design of the IHS involved the use 29 districts as strata – the four urban centers of Malawi plus the 26 current administrative districts, less Balaka which was not yet established when the survey was designed. The distribution of the 6586 household sample across these 29 districts is presented in the table below.

No households remain in the data set for Ntchisi district. Moreover, relative to the population size of the district, the number of households in Nsanje, Chikwawa, Mwanza, Phalombe, Machinga,

Number of households in each IHS district in the 6586 household data set, arranged in columns by administrative region.

District	hhs	District	hhs	District	hhs
Nsanje	97	Ntcheu	147	Mzimba	185
Chikwawa	132	Dedza	310	Mzuzu City	347
Mwanza	17	Salima	192	Nkhata-Bay	122
Blantyre Rural	248	Lilongwe Rural	594	Rumphi	162
Blantyre City	414	Lilongwe City	229	Karonga	22
Zomba Rural	268	Mchinji	308	Chitipa	130
Zomba Municipality	164	Kasungu	381		
Thyolo	268	Dowa	262		
Mulanje	391	Ntchisi	nil		
Phalombe	49	Nkhotakota	185		
Machinga	194				
Mangochi	479				
Chiradzulu	325				

Number of households interviewed in each month and percent of total in the 6586 household data set.

November	685	10.4%	May	564	8.6%
December	615	9.3%	June	568	8.6%
January	519	7.9%	July	619	9.4%
February	576	8.7%	August	545	8.3%
March	560	8.5%	September	413	6.3%
April	531	8.1%	October	391	5.9%

Ntcheu, and Rumphi are exceptionally small. The expansion factors for the households in these districts – the number of households in the district which each of the valid IHS households represents – will be large. To deal with this problem in the poverty analysis, weights will be applied to households. In order to accurately undertake the poverty line analysis without allowing any small subset of households to have unwarranted influence on the results, weights will be applied. The weighting scheme used is discussed in the paper in this series on the derivation of the poverty line.

The distribution of these households by month of interview is shown in the second table. The distribution is quite even. There is a reduction in numbers in the final two months of the survey, September and October, with the highest proportion of households being interviewed in the first month of the survey, November.

To sum up, pending the analysis of the effect which the households which do not report staples will have on the poverty line and the poverty head count, it is recommended that the poverty line be derived from a final IHS household sub-sample consisting of either 6514 or 6586 households.