The methodology for the poverty analysis of the Malawi Integrated Household Survey, 1997-98, *with provisional poverty lines*

This is the fourth in a series of documents describing the poverty analysis of the Integrated Household Survey (IHS), 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.

In this document the derivation of a poverty line from the IHS data will be described. The method described here builds on the output of what was described in the three previous papers in this series: i) the calculation of a welfare indicator for every household based on per capita daily household consumption and expenditure; ii) the determination of the per capita recommended daily calorie requirements for each household; and iii) the selection of the households from the IHS which are judged to have reliable expenditure and consumption data.

The poverty line is that level of welfare above which one can consider that the basic needs of a household or individual are adequately met. The poverty line will be established with reference to the welfare indicator calculated for each of the IHS households. Comparing the welfare indicator for a household to the poverty line, if the welfare indicator of a household is greater than the poverty line, that household is not considered to be poor.

In observing households around us, all of us carry in our minds some sort of idea of a poverty line which enables us to describe some households as poor, while others as not poor. Such poverty lines are defined in a very subjective way. What is sought in this analysis, however, is to develop an objective poverty line which can be replicated to generate comparable poverty measures for Malawi either through time to determine trends in poverty or across space to compare the general welfare level of Malawians to that of populations in other countries. In brief, the following steps will be taken to derive the poverty line:

- The objective core of this poverty line will be the per capita recommended daily calorie requirement for the households in the IHS data set used here. These requirements have been established by nutrition researchers.
- This recommended calorie requirement will be used to establish the food component of the poverty line by determining what it costs for a poorer household in Malawi to acquire sufficient calories to meet the poverty line. The cost for each calorie will be determined by calculating the value of each calorie consumed by these poorer households.
- More than simply food is needed to meet the basic needs of a household. There is a non-food component to the poverty line as well. Unfortunately, no independent objective criteria exists by which one can establish what should make up the non-food component of the poverty line. The method adopted here will be to examine the non-food consumption of households whose <u>total</u> consumption and expenditure is in the neighborhood of the <u>food</u> component of the poverty line. Since these households are sacrificing nutritionally necessary food consumption to consume these non-food items, the items can be considered basic necessities for household welfare. The value of these items will make up the non-food component of the poverty line.
- Summing the food and non-food components results in the poverty line. The poverty status of

each household now can be assessed by comparing the level of its welfare indicator to the poverty line.

Step by step description of the methodology to derive the poverty line

First, two key analytical points: At several steps in the poverty analysis, the mean or the median value of a variable is to be used in the computations. In the example presented in the body of this paper, the median will be used. The mean, however, will be presented alongside the median in tables. For the reasoning behind this choice of which measure of central tendency to use, see the relevant section of the appendix to this paper. In the sensitivity analyses which conclude this paper, the results of running this same analysis using means rather than medians are presented.

Secondly, the sample households are weighted based on the population of the district stratum of which they are a part. However, in seven of the 29 districts with low numbers of sample households, for purposes of weighting these districts are merged with adjoining districts. The weighting scheme is described fully in the appendix. In the sensitivity analysis section at the end of this paper, the effects on the results of the poverty analysis from using the simple weighting scheme are also shown.

Valid households

The analysis being described was done using the SPSS statistical software program. The first step is to build a household level working file by assembling the following information for each of the households which were judged to have data of sufficient quality for the analysis:

- per capita daily calorie consumption as calculated from the food consumption and expenditure information in the IHS.
- welfare indicator: daily per capita total expenditure and consumption.
- per capita food consumption and expenditure.
- per capita non-food consumption and expenditure.
- number of persons in the household

These data are produced by the process described in the paper in this series on calculating a household welfare indicator.

Example

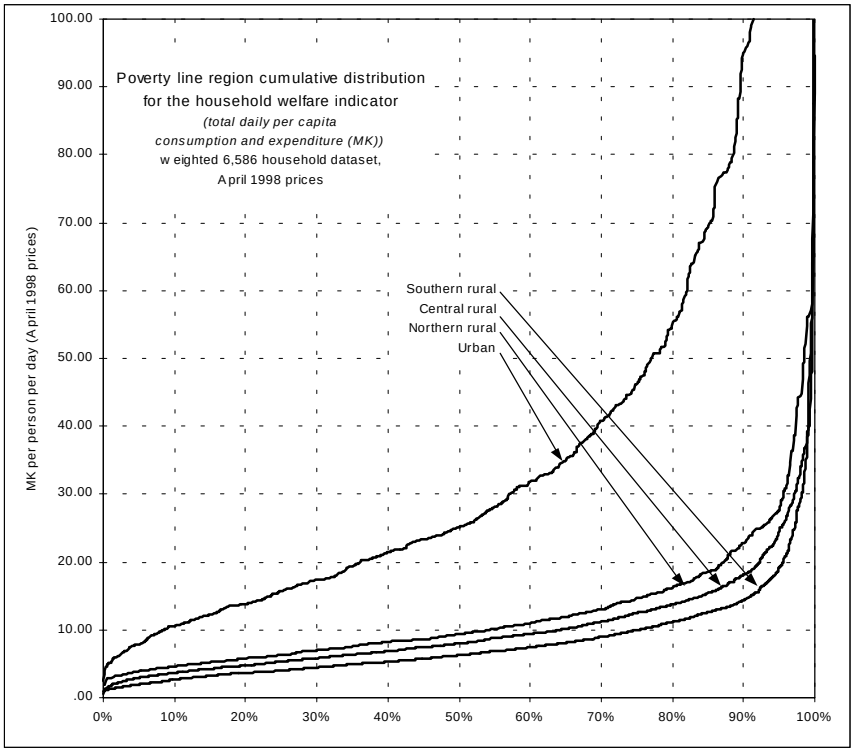
In selecting the valid households for the poverty analysis, as described in the paper on that topic, two possible samples for analysis were to be investigated, one of 6586 households in which 72 households were retained who reported having consumed no staples, and one of 6514 in which these households were dropped. Here the larger data set will be used.

The spatial composition of this data set by poverty line regions is as follows:

Region	households		individuals	
Southern rural	2468	37.5%	10,597	36.6%
Central rural	2379	36.1%	10,979	37.9%
Northern rural	810	12.3%	3,368	11.6%
Urban	929	14.1%	4,002	13.8%
TOTAL	6586	100.0%	28,946	100.0%

The weighted cumulative distribution of the welfare indicator in each poverty line region is

shown in the graph below.



In the examples accompanying each of the sections is sketched out the poverty analysis of this data set. The results of the same poverty analysis on the smaller 6514 household data set is presented for comparison later in a sensitivity analysis.

Per capita daily calorie requirements

Into the working file is merged information for each of the households on their per capita recommended daily requirements (RDR) for calories. The RDR for a household is determined by its demographic characteristics and by the activity level of the adults in the household. The paper in this series on calculating the per capita calorie RDR presents two possibilities: one in which all adults are assigned a moderate activity level, a second in which rural adults are assigned an RDR which is based on working at a high activity level for 3 months of the year, corresponding to the early to mid-cropping season, and at a moderate activity level for the other 9 months. The appropriate RDR as chosen by the analyst should be merged into the working file.

Example

In the example presented in the body of the text, the RDR calculation is used in which all adults are assigned the moderate activity level calorie requirement. The mean and median per capita calorie

Per capita daily calorie RDR for total 6586 hh sample

Region	Median daily per capita RDR	Mean daily per capita RDR
Southern rural	2174	2192
Central rural	2140	2175
Northern rural	2171	2212
Urban	2250	2288
National	**2167**	**2198**

N.B.: weighted by hh size and for sample design

RDR for the data set is presented in the table below. In the appendix, the results of a poverty analysis using the alternate RDR are presented.

Calculating the food poverty line

In order to derive the food poverty line – the food consumption and expenditure portion of the poverty line – one needs to determine the value of each calorie which poorer households in the population consume. Poorer households are chosen for this determination on the assumption that these people will be acquiring their calories as cheaply as possible, given local taste preferences. Richer households, in contrast, usually spend more for their calories. For example, while poorer households will eat maize flour and cassava, richer households might substitute more expensive wheat flour and Irish potatoes.

The cost per calorie for the poorer households is then multiplied by the recommended daily calorie requirement for those same households to come up with the food poverty line. The reason the RDR for the poorer households is used rather than that of the entire sample is so that the food poverty line reflects the prevailing demographic conditions of the poorer households and their consequent calorie needs.

'Poorer' households – identifying them in the data set

The poorer households can be identified in many ways. The poorer households in the Malawi poverty analysis will be defined as those whose reported calorie consumption is less than their recommended daily calorie requirement.

An alternative approach is to define poorer households as those whose welfare indicator – total daily per capita consumption and expenditure – is less than the median welfare indicator for the entire data set. However, this definition was shown to be unsatisfactory in undertaking a poverty analysis across several poverty line regions, as is being done in Malawi. If there are different price structures in the regions, using the welfare indicator criteria will result in a bias against selecting the poorer households in the highest cost regions.

Example

The table below shows the number and percentage of all households in each region of the 6586 dataset who are included in the two definitions of poorer households noted above. There is a clear anti-urban bias when the welfare indicator is used as the defining variable. This concept can be visualized above in the cumulative distribution graphs of the welfare indicator in each region.

In the sensitivity analyses, the results are presented of the same poverty analysis run using the median welfare indicator to define the poorer households.

Cost per calorie and recommended daily calorie requirement for poorer households

Regional composition of the poorer households under two definitions

Region	'poorer' defined using calorie consumption		'poorer' defined using welfare indicator		*total households*
Southern rural	1669	67.6%	1469	59.5%	*2468*
Central rural	1478	62.1%	1063	44.7%	*2379*
Northern rural	514	63.5%	225	27.8%	*810*
Urban	695	74.8%	55	5.9%	*929*
TOTAL	4356	66.1%	2812	42.7%	*6586*

N.B.: unweighted values

Food poverty line computation values

	Median values			Mean values	
Region	**Cost per 1000 calories (MK)**	**Per capita calorie RDR** *(poorer HHs)*	**Food poverty line (MK/person/day**	**Cost per 1000 calories (MK)**	**Per capita calorie RDR** *(poorer HHs)*
Southern rural	3.01	2167	6.53	3.56	2186
Central rural	3.62	2140	7.76	4.21	2175
Northern rural	4.08	2179	8.90	4.47	2218
Urban	7.53	2250	16.95	8.47	2288

In order to derive the cost per calorie, the poorer households in each region are selected. The reported calorie consumption for these households is divided by the food consumption and expenditure component of the welfare indicator to come up with a cost per calorie for the household. The weighted median cost per calorie for each region is then used in the computations.

The per capita recommended daily requirement for calories to be used to establish the food poverty line in each region is simply the weighted median per capita RDR for the poorer households in that region.

Calculation of the food poverty line

The food poverty line for all households in a region is the product of the price per calorie and the recommended per capita daily calorie requirement for poorer households in a region.

Example

For the 6586 household data set, the table above presents the results from these calculations. Note that the calorie cost is in terms of 1000 calories. The food poverty line is only presented for the median value, as this is the food line used in this example.

Non-food component of the poverty line

The food component of the poverty line has an objective basis in being linked to the recommended daily calorie requirements of individuals in the households. The non-food component is more fuzzy, as it is difficult to quantify what exactly are the minimum non-food requirements of an individual.

For the non-food component of the poverty line, here we use the value of the non-food consumption of households whose total consumption and expenditure – the household welfare indicator – is in the neighborhood of the food poverty line. This is done on the assumption that the non-food consumption of these households reflects the minimum amount necessary. These households have chosen to consume non-food goods rather than food when they are objectively in need of additional food consumption. This choice indicates the importance of these non-food items to the welfare of these households.

The neighborhood is defined as households whose welfare indicator is within 20% of the food poverty line. The weighting scheme gives greater weight to the non-food consumption of households whose welfare indicator is closer to the food poverty line. For example, households whose welfare indicator is 18 to 20 percent greater or less than the food poverty line are given a weight of one, whereas households whose welfare indicator is within 2 percent of the food poverty line are given a weight of ten for the computation.

An alternative method to define the non-food component of the poverty line is to consider those

Non-food poverty lines (MK/person/day)

	Median values		Mean values	
Region	**Based on welfare indicator**	**Based on food consumption & expenditure**	**Based on welfare indicator**	**Based on food consumption & expenditure**
Southern rural	1.23	1.51	1.48	2.58
Central rural	1.51	1.97	1.78	3.49
Northern rural	2.26	2.60	2.50	5.03
Urban	8.43	18.31	8.35	38.81

households whose food consumption and expenditure is in the neighborhood of the food poverty line. This will lead to a higher poverty line, but makes more intuitive sense. The principal drawback to this method is that in countries, such as Malawi, where poverty is prevalent, the higher total poverty line which results may generate poverty head counts so high that they are of no value for policy making purposes.

Both methods could be used to set two poverty lines. Both poverty lines would have the same food component. However, the 'normal' poverty line would have a non-food component based on the non-food consumption of households whose food consumption is close to the food poverty line. The 'ultra-poor' poverty line would have a non-food component based on the non-food consumption of households whose total consumption and expenditure is close to the food poverty line.

Example

The table above presents the non-food component of the poverty line using the two methods sketched above. In the sensitivity analysis section are presented the results of an analysis using food consumption and expenditure to identify households for purposes of defining the non-food component of the poverty line.

The poverty line and poverty head counts

The poverty line is simply the sum of the food and non-food components of the line.

Once the poverty line is established, households in each region are categorized as poor and non-poor depending on whether their total consumption and expenditure, their welfare indicator, is below or above the poverty line. The poverty head count can then be computed, indicating the proportion of individuals or of households below the poverty line.

Example

The poverty lines and individual and household poverty headcounts in the table below result from the example presented here. Recall that the following analytical decisions were made:

Poverty lines and poverty head counts

Region	**Poverty line** (MK/person/day)	**Individual poverty headcount**	**Malawi's poor in region (individual)**	**Household poverty headcount**	**Malawi's poor households in region**
Southern rural	7.76	62.2%	43.4%	56.7%	45.0%
Central rural	9.27	58.8%	37.7%	53.7%	36.4%
Northern rural	11.16	60.6%	10.2%	50.2%	9.7%
Urban	25.38	50.8%	8.7%	44.3%	8.9%
MALAWI	-	59.6%	-	53.6%	-

- Sample of 6586 IHS households,
- Median values of variable distributions used throughout,
- Recommended Daily Calorie Requirements based on moderate activity level for all adults,
- 'Poorer' households defined as those whose reported calorie consumption is less than their calorie requirements,
- Non-food component of the poverty line determined by considering the non-food consumption and expenditure of households whose total consumption and expenditure is close to the food poverty line, and
- Analytical weights for survey design based on merging 14 of the 29 district strata into 7 paired districts (see appendix).

Sensitivity analysis

Analytical decisions had to be made in the poverty analysis. These have been noted in the text above. This section presents what the results of the poverty analysis would be if the same analysis would have been run using the alternative choice for each decision.

The following table presents the analytical choices made. The second column presents that which was replicated in the example shown in the text. The third column gives the alternative method.

Analytical choice	Example	Alternative
Number of households in the data set	6586 households	6514 households
Central tendency measure used in the analysis	Median	Mean
Recommended Daily Calorie Requirement calculation	Moderate activity level for all adults	Rural adults have 1/4 high activity & 3/4 moderate activity levels to account for cropping season labour demands
Definition of 'poorer' households	Per capita calorie consumption of household less than its RDR	Welfare indicator for household below national median
Non-food component calculated for which households	Welfare indicator near food poverty line	Food expenditures near food poverty line
Weighting of sample households to make representative of the population	Weights based on lumped district population	Weights based on simple district population

Assessment of alternative methodologies

The results of the application of the alternative methodologies are shown in the table below. In the interest of simplicity and space, no analyses to examine interactions of these alternative methods was done. Only the single choice of interest is changed from the basic analysis presented above.

6614 households – There is very little difference in the results of the poverty analysis using 6614 households or using 6586 households. The 72 households included in the 6586 household data set are poorer than the general population, given that the poverty head count goes down by a fraction in all regions when the 72 are dropped from the analysis. That said, there seems no compelling reason to drop these 72 households from the analysis.

Mean values used in the calculations – The national poverty head count rises by 10.2 percent if means rather than medians are used in the poverty line calculations. This is to be expected given the nature of the consumption and expenditure data. Such data tends to be skewed, with the bulk of observations being low levels of consumption and expenditure, with outlying observations at much

Results of the sensitivity analyses of the poverty analysis

Analytical methodology scenario	Region	Poverty head count (% of individuals)	Poverty line (MK/person/ day)	Food poverty line (MK/person/ day)	Non-food poverty line (MK/person/ day)
Example above	Southern rural	62.2	7.76	6.53	1.23
	Central rural	58.8	9.27	7.76	1.51
	Northern rural	60.6	11.16	8.90	2.26
	Urban	50.8	25.38	16.95	8.43
	MALAWI	**59.6**	-	-	-
6514 households in sample	Southern rural	62.1	7.74	6.51	1.23
	Central rural	58.6	9.25	7.75	1.50
	Northern rural	60.1	11.03	8.88	2.15
	Urban	50.6	25.36	16.93	8.43
	MALAWI	**59.4**	-	-	-
Mean values used in calculations	Southern rural	73.3	9.60	7.79	1.82
	Central rural	70.0	11.21	9.15	2.06
	Northern rural	67.9	12.56	9.92	2.64
	Urban	56.9	29.47	19.37	10.09
	MALAWI	**69.8**	-	-	-
Seasonally adjusted RDR	Southern rural	63.5	7.89	6.62	1.27
	Central rural	60.8	9.50	7.93	1.57
	Northern rural	60.9	11.26	8.96	2.30
	Urban	50.8	25.38	16.95	8.43
	MALAWI	**60.9**	-	-	-
'Poorer' households as those	Southern rural	47.8	6.05	5.11	0.94
below median welfare Indicator	Central rural	40.2	6.87	5.74	1.13
	Northern rural	43.0	8.51	6.89	1.62
	Urban	12.2	11.34	7.86	3.48
	MALAWI	**40.8**	-	-	-
Non-food calculated using hhs	Southern rural	64.3	8.04	6.53	1.51
with food expenditures at food	Central rural	62.7	9.73	7.76	1.97
poverty line	Northern rural	62.4	11.49	8.90	2.60
	Urban	65.2	35.26	16.95	18.31
	MALAWI	**63.6**			
Weights on simple district pop.	Southern rural	65.1	8.10	6.77	1.33
	Central rural	59.3	9.23	7.73	1.50
	Northern rural	60.0	11.22	8.92	2.30
	Urban	50.8	25.38	16.95	8.43
	MALAWI	**60.9**	-	-	-

higher levels. These outliers raise the means relative to the medians.

Using the means in the analysis would translate into a higher per capita RDR and higher cost per calorie for the poorer households. The resultant higher food line would lead to the analysis to compute the non-food component of the poverty line being run on somewhat richer households who have higher non-food consumption. Overall the poverty line would go up.

The exceptionally large difference between the median and the mean head counts bears closer attention. The focus on the poorer households to derive the food poverty line should have removed much of the skewness from the distribution. Why this did not should be examined further.

Seasonally adjusted rural adult RDR – Differences should only be found in the rural regions. The largest shift is seen in the Central region, where the head count increased by 2.0 percent.

However, the overall change, although upwards as expected, is not very great. Either RDR could be used in the final analysis. For cross-country comparisons, however, it may be better to use the moderate activity RDR, as this straight-forward RDR is more likely to permit international comparisons to easily be made.

'Poorer' households defined using median welfare indicator – This change in the analysis leads to the largest shift in the poverty headcount: nationally it moves down by 18.8 percent. Earlier it was highlighted that this method has an anti-urban poor bias. This is clearly seen, with the urban poverty head count dropping by 38.6 percent and the urban poverty line dropping 55 percent relative to the base poverty analysis. The shifts in the rural regions, while not as large, are still substantial.

As noted earlier, this method of defining the poor is confounded by the fact that there are real differences in the cost of living between the four poverty line regions. Defining the poorer households on nutritional grounds is to be preferred.

Calculating the non-food component using households whose food consumption and expenditure is close to the food component of the poverty line – The national poverty head count goes up by 4.0 percent using this method. However, the boost in the head count is primarily found in the urban centers. The poverty head count for the urban region rises by 14.4 percent, to the level of the rural regions. Similarly, while the non-food component of the poverty line in the rural areas increases by less than MK0.50, in the urban centers it increases by almost MK10.00.

Further discussion and investigation is needed to judge the merits of using this method to determine the non-food component of the poverty line. As noted, this method makes intuitive sense. However, it is unclear why the largest impact of this method is seen in the urban region.

Using simple district populations to calculate analytical weights – The overall head count rises by 1.3 percent when simple district weights are used, although the poverty head count in the Northern rural region dropped slightly. The analysis used for the example was based on analytical weights which paired seven districts with low numbers of sample households with adjoining districts with a larger number of sample households.

Although this is discussed in more detail in the appendix to this paper, the impetus for using weights based on paired districts stems from the very large expansion factors for the sample households in the districts with low numbers of sample households. This means that these few households would have an inordinately large effect on any weighted statistics used in the analysis. Any households in these districts with values that are outliers would skew the regional median or means for those values. The districts were joined together in order to lower these expansion factors.

The results show that there are outlier values among the sample households from districts with low numbers. Only the rural regions are affected, as it is only here where districts were joined. It is only in the Southern rural region that one sees any noteworthy change. Here the poverty line increases by MK 0.34 (4.4 percent rise), while the head count rises by 2.9 percent. Given these shifts, it seems prudent to continue to use the weights that are based on joined districts.

In sum, the analysis used for the example in the text seems to be the proper one to use. There is scope for closer examination of the underlying data to understand why using means rather than medians leads to such a difference in the poverty head count.

Likewise, the non-food component derivation methods should be looked at more closely. Using households whose food consumption and expenditure is at the poverty line to calculate the non-food component leads to a much higher non-food component to the poverty line in the urban region primarily. If a satisfactory explanation for this effect is found, two poverty lines could be established:

a ‘normal’ poverty line using this method, and an ‘ultra’ poverty line using households whose total consumption and expenditure is at the food poverty line to calculate the non-food component, as was done in the example.

The other alternative methods considered either lead to no significant differences in the results or are clearly flawed.

Consequently, provisionally we should accept the poverty head count of the analysis conducted in the example with the 6,586 household data set – a national poverty head count of 59.6 percent.

Appendix : Methodological considerations in deriving the poverty line

Two important considerations underlying the poverty analysis are presented in this appendix: whether to use median or mean values in the computations and weighting data from the sample households in the analysis so that they will be representative of the population of the district in which they are found.

Using median or mean values in poverty line analysis computations

In several computations in the poverty analysis, one makes use of a measure of central tendency for a variable – the per capita calorie requirement, the price per calorie for the poorer households, and the non-food component of the poverty line. The two choices for this measure are the mean – the average value – or the median –the middle value of a ranked distribution of a variable.

In a skewed distribution the outlying values of a variable will have a greater weight in the computation of the mean than will values closer to the center of the distribution. If the distribution is strongly skewed, the resultant mean will be less representative than we might desire of the cases at the non-skewed end of the distribution where the majority of cases lie.

The median, in contrast, is not affected by the outliers of a skewed distribution. Consequently, when analyzing a skewed distribution the median is the preferred measure to use. Household consumption and expenditure distributions are typically skewed, where one finds a few households reporting very high levels. The median is used in the example presented in this paper. However, where median values are presented for use in the analysis, the corresponding means are also shown.

It should be noted, however, that there may be methodological reasons for preferring to use means to medians in an analysis. The principal consideration in this regard is in determining the confidence intervals for the poverty estimates which one generates. Computationally, this is more easily done using means than using medians.

Weights to correct for under or over-representation of sample households relative to the population as a whole

Survey sample design

The sample design for the IHS was based on 29 strata – the four major urban centers of Malawi, plus the 26 administrative districts, less Balaka (which was not yet established at the time of the survey design and was still part of Machinga). The data generated by the IHS was to be representative down to the level of these 29 districts.

In each of the 25 generally rural administrative districts, one or more Traditional Authorities (TA) were selected randomly, with the probability of a particular TA being selected being proportional to its population relative to the population of the district as a whole. Roughly for every 50,000 to 60,000 households in a district, one TA was selected.

Within the selected TAs in the 25 districts, 12 Enumeration Areas (EA) were selected on the basis of probability of selection proportional to the population of an EA. The 12 EAs were selected in each selected TA for the 12 months of the survey. Twenty households were then randomly selected in each EA from a full listing of all households in the EA. These twenty households would all be interviewed in a single month of the survey year. The enumerator would go to another EA in that TA in the following month until all 12 EAs had been enumerated over the survey year.

In the four urban centers a two-stage sample selection was done, rather than the three stage

sample selection procedure used in the rural districts. EAs in the city were selected on the basis of probability of selection proportional to the population of the EA. Ten households were then randomly selected in each selected urban EA from complete lists of households in the EA. The number of EAs chosen in a city was roughly proportional to the population of the city: Blantyre – 60 EAs, Zomba – 24, Lilongwe – 36, Mzuzu – 24.

Expansion factors

The households selected in each district stratum are representative of households in that district. As the stages of the sample selection process were done on the basis of probability of selection proportional to the size of the population, one can simply divide the number of households in a district by the total number of households selected in that district for the IHS to derive the *expansion factor* for each sample household in the district. The expansion factor tells you how many households in the general population each sampled household represents. All survey households within a district have the same expansion factor.

If all households in the IHS sample had the same expansion factor, the sample would be self-weighting and no weights would have to be assigned. As described above, the original sample design would have led to a roughly self-weighted sample, where each IHS household would represent approximately 175 households in the general population.[1]

The planned total number of households to be interviewed was 12,960. At the end of the survey year, just under this number of questionnaires were submitted for data entry. However, upon cleaning, it became clear that about 15 percent of these questionnaires were of insufficient quality for analysis. The cleaned data set released in early May 2000 has 10,698 households. Additional cleaning was then undertaken on this set of households to determine whether all had suitable expenditure and consumption data for poverty analysis. About 3000 households were judged to be questionable in this regard, so are not being used for the poverty analysis.

The cleaning of the IHS data revealed certain districts where disproportionate numbers of questionnaire were judged to be of poor quality. Consequently, with the dropping of large numbers of survey households from some districts and relatively few from others, any possible self-weighting to the sample design breaks down. IHS households in different districts will have different expansion factors – that is, sample households in different districts will represent significantly different numbers of households. This fact needs to be taken into account in the poverty analysis in order for the results to be representative at spatial scales broader than the district level.

However, several districts are problematic for having very low numbers of survey households relative to the size of their population. Ntchisi district is the extreme case, with no survey households having reliable consumption and expenditure data. The other problematic districts are Chikwawa, Mwanza, Phalombe, Machinga, Ntcheu, and Rumphi. Given the large expansion factors which the households from these districts will have, any households from these districts with extreme values for variables which are used in the poverty analysis will significantly affect the results of the analysis.

In order to minimize the potential outlier effect from these districts, in the analysis these seven districts are merged with adjoining districts with comparable agro-ecological conditions and dominant economic activities which are in the same poverty line region. For example Nsanje and Chikwawa are 'lumped' together. The expansion factor for the joined district is: (total hhs in Nsanje + total hhs

[1] However, as can be seen in the table below, expansion factors likely would have had to have been applied for the poverty analysis, even if all 12,960 households were properly surveyed. There is considerable variation in expansion factors even with the designed number of households, e.g. compare Salima with Mzuzu City.

in Chikwawa) / (total survey hhs in Nsanje + total survey hhs in Chikwawa). By doing this, the 'good' district of the two will moderate the high expansion factor of survey households in the problematic district. The partner districts for the other problematic districts are Blantyre Rural for Mwanza, Mulanje for Phalombe, Mangochi for Machinga, Dedza for Ntcheu, Dowa for Ntchisi, and Mzimba for Rumphi.

Example

For the 6586 household data set, the expansion weights for each district as originally designed are shown in the table below in the fourth column. In column seven the expansion weights which would be used if survey households in each district were treated separately are presented. Note that for Dowa district, the population is the sum of the population of Dowa and that of Ntchisi district. The expansion weights above 500 are underlined. The final column present the expansion weights when 14 of the 29 districts are joined together to derive a set of expansion factors with fewer extreme values. Chikwawa and Nsanje paired retain an unfortunately high expansion factor. However, there are no alternative pairings for these two Lower Shire Valley districts.

Expansion factors for IHS poverty analysis

District	Planned number of survey hhs	1997 household population	Planned expansion factor	Number of survey hhs	1997 household population (Ntchisi with Dowa)	Simple district expansion factor	Number of survey hhs – joined districts	Number of hhs in joined district (14 districts paired)	Joined district expansion factor
Nsanje	240	44,746	186	97	44,746	461	229	119,446	522
Chikwawa	480	74,700	156	132	74,700	566	-	w/Nsanje	-
Mwanza	240	31,542	131	17	31,542	1855	265	102,404	386
Blantyre Rural	480	70,862	148	248	70,862	286	-	w/Mwanza	-
Blantyre City	600	116,045	193	414	116,045	280	414	116,045	280
Zomba Rural	720	117,911	164	268	117,911	440	268	117,911	440
Zomba Municipal.	240	14,043	59	164	14,043	86	164	14,043	86
Thyolo	720	107,389	149	268	107,389	401	268	107,389	401
Mulanje	720	102,425	142	391	102,425	262	440	158,410	360
Phalombe	240	55,985	233	49	55,985	1143	-	w/Mulanje	-
Machinga	720	148,057	206	194	148,057	763	673	294,044	437
Mangochi	720	145,987	203	479	145,987	305	-	w/Machinga	-
Chiradzulu	480	55,160	115	325	55,160	170	325	55,160	170
Ntcheu	480	83,511	174	147	83,511	568	457	193,832	424
Dedza	480	110,321	230	310	110,321	356	-	w/Ntcheu	-
Salima	240	60,006	250	192	60,006	313	192	60,006	313
Lilongwe Rural	1200	207,598	173	594	207,598	349	594	207,598	349
Lilongwe City	360	93,199	259	229	93,199	407	229	93,199	407
Mchinji	480	70,874	148	308	70,874	230	308	70,874	230
Kasungu	480	102,819	214	381	102,819	270	381	102,819	270
Dowa	480	88,963	185	262	124,405	475	262	124,405	475
Ntchisi	240	35,442	148	-	w/Dowa	-	-	w/Dowa	-
Nkhotakota	240	55,189	230	185	55,189	298	185	55,189	298
Mzimba	480	109,641	228	347	109,641	316	369	135,799	368
Rumphi	240	26,158	109	22	26,158	1189	-	w/Mzimba	-
Mzuzu City	240	17,745	74	122	17,745	145	122	17,745	145
Nkhata-Bay	240	35,581	148	162	35,581	220	162	35,581	220
Karonga	240	35,616	148	130	35,616	274	130	35,616	274
Chitipa	240	25,090	105	149	25,090	168	149	25,090	168
TOTAL	**12960**	**2,242,605**	**173**	**6586**	**2,242,605**	**341**	**6586**	**2,242,605**	**341**