Small-area Estimation of Poverty and Malnutrition in Cambodia

CAMBODIA
DECEMBER 2012

The National Institute of Statistics, Cambodia
in collaboration with
The United Nations World Food Programme

Prof Stephen Haslett, A/Prof Geoffrey Jones and Alison Sefton
Institute of Fundamental Sciences - Statistics
Massey University
New Zealand

SUMMARY

Small-area estimates (SAE) of poverty and malnutrition in Cambodia are produced at commune level by combining survey data with auxiliary data derived from the 2008 Cambodia Census of Population and Housing (Census2008). A model for predicting log average per capita household expenditure is estimated from the 2009 Cambodia Socio-economic Survey (CSES2009) based on the Cambodia National Institute of Statistics calculation of expenditure in each of the households sampled in CSES. The model is applied to household-level census data to estimate poverty incidence, gap and severity. FAO has used CSES2009 to derive estimates of caloric intake in the form of kilocalories consumed per capita for each sampled household; and a survey based model for kilocalorie consumption is also applied to household-level census data to investigate the feasibility of predicting kilocalorie consumption per household; when compared with a kilocalorie cut-off norm this could potentially be used to estimate undernourishment at commune level. We find however that there remains considerable unmodellable uncertainty in the kilocalorie data, so that, since the small-area estimates of undernourishment at commune level are not sufficiently reliable, they have not been included in the report. Models for predicting standardized height-for age and weight-for-age are estimated from both the 2008 Cambodia Anthropometrics Survey (CAS2008) and the 2010 Cambodia Demographic Health Survey (CDHS2010), each being applied to child-level census data to estimate incidence of stunting and underweight; the separate estimates from each source are combined using inverse-variance weighting to produce a single set of estimates for each of stunting and underweight. Estimates of wasting, though desirable, are not produced here because of the inadequacy of predictive models for height-for-weight from both CAS2008 and CDHS2010. The small-area estimation procedure used in this study does not produce *measures* of poverty, caloric intake or child malnutrition at the local level. Rather the procedure applied here is able to *estimate* welfare outcomes – based on a statistical model estimated in the relevant household survey. These estimates of wellbeing are measured with error, and the degree of imprecision will vary as a function of a wide variety of factors, most notably the degree of disaggregation at which indicators of wellbeing are being estimated. In this study it was found that estimates at the level of a commune– which comprises on average around 1700 households – are generally reasonably precise. Estimates at village or enumeration area level are far less precise. The precision of estimates varies with the specific indicator of wellbeing, and precision is generally better with consumption poverty estimates than with estimates of caloric intake and child malnutrition, because there are fewer survey

variables that can be matched with the census in the latter models. Comparisons are made with the poverty estimates derived from the 2009 Commune Database (CDB2009), and with the earlier small-area estimates of poverty and malnutrition detailed in Fujji (2003).

ACKNOWLEDGMENTS

The statistical analysis for and writing of this report has been undertaken by Professor Stephen Haslett, A/Prof Geoffrey Jones and Alison Sefton of the Institute of Fundamental Sciences - Statistics, Massey University, New Zealand. The authors have benefited greatly from collaborative work with staff at the United Nations World Food Programme (WFP), in particular the Head of the Vulnerability Analysis and Mapping (VAM) Unit Mr Kurt Burja, together with Mr Yav Long and Mr Chanvibol Choeur. We gratefully acknowledge the considerable assistance provided at all stages by the NIS officers, particularly Saint Lundy.

The completion of both this and the earlier 2008 and 2010 feasibility reports follows extensive consultation with the National Institute of Statistics (NIS), other staff from the Ministry of Planning (MoP), International Fund for Agricultural Development (IFAD), Statistics Sweden, UNICEF, World Bank, and the World Food Programme (WFP) - which commissioned this research.

National Institute of Statistics

H. E. San Sy Than, Director General, NIS, Ministry of Planning
Saint Lundy, Deputy Director, General Statistics, NIS, Ministry of Planning
Ing Sokun, Officer, General Statistics, NIS, Ministry of Planning
Khin Song, Deputy Director, General Statistics, NIS, Ministry of Planning
Nguon Sovann, Vice Chief, Bureau of Survey Planning, NIS, Ministry of Planning
Sin Serey Vuth, Chief, Bureau of Statistical Information, NIS, Ministry of Planning

National Strategic Development Plan (NSDP) Secretariat, Ministry of Planning

H. E. Tuon Thavrak, Director General, General Directorate of Planning, Ministry of Planning
Hildegard Lingnau, Senior Advisor to Ministry of Planning, Centrum für Internationale Migration und Entwicklung (CIM)
Ramanathan Natarajan, International Consultant

Statistics Sweden – International Consulting Office

Sten Backlund, Chief Advisor, International Capacity Building Project at NIS
Agneta Sandqvist, Advisor Household Surveys, International Capacity Building Project at NIS
Lars Soderberg, Advisor ICT, International Capacity Building Project at NIS
Tiina Orusild, Senior Advisor Statistical Methodology, SCB/National Institute of Statistics
Hans Pettersson, Senior Statistician at Statistics Sweden, SCB/National Institute of Statistics
Birgitta Mannfelt, Statistics Sweden, Chief Advisor at NIS, Cambodia..
Anders Sundstrom, Statistics Sweden, Advisor at NIS, Cambodia..

United Nations Development Programme / International Fund for Agricultural Development

Ung Dara Rat Moni, IFAD/UNDP Advisor

United Nations World Food Programme

Jean-Pierre de Margerie, Country Director, Cambodia
Thomas Keusters, Country Director, Cambodia
Bradley Busetto, Officer-in-Charge, Cambodia
Coco Ushiyama, Deputy Country Director, Cambodia
Kurt Burja, Programme Officer, Vulnerability Analysis and Mapping (VAM), Cambodia
Yav Long, National Program Officer, VAM, Cambodia
John Jeong, Programme Officer, VAM, Cambodia
Kim Ratha, Senior Programme Assistant, VAM, Cambodia
Michael Sheinkman, Senior Regional Programme Advisor, Vulnerability Analysis and Mapping (VAM), WFP Bureau for Asia, Thailand

World Bank

James Knowles, International Consultant.
Tim Conway, Senior Poverty Specialist
Neak Samsen, Poverty Specialist
Carlos Sobrado Senior Poverty Economist, PREM/ East Asia and the Pacific, Cambodia

UNICEF

Joel Conkle, Nutrition Specialist

Cambodia National Committee for Sub-National Democratic Development (NCDD)

Ny Borath

Contents

1. Introduction

1.1 Background

Cambodia borders Thailand to the north and west, Laos to the northeast, and Vietnam to the east and southeast, with a 443-kilometre coastline along the Gulf of Thailand. It has an area of 181,035 square kilometres, including 4,520 square kilometres of lakes and inland waterways. The principal inland water bodies are the Mekong River, the Tonle Sap (Great Lake) and the Tonle-Bassac River which together form a network of river channels, levees and river basins stretching across the entire lowlands.

The population in 2008 was 13.4 million. Cambodia is one of the least developed countries, ranking 139 out of 187 countries on the Human Development Index (UNDP, 2007). Maps of population density, ecological zones, and the administrative units including commune boundaries are given in Appendix D.0.

Economic poverty is widespread. The lack of secure land tenure, remoteness from markets and services, lack of productive assets, low levels of education, and high dependency ratios are all factors contributing to the poor living conditions of the rural population. Analysis from the 2009 Cambodian Socio-economic Survey, using the National Institute of Statistics calculation of expenditure per household for the CSES sample, estimated poverty incidence at 22.9 percent on the basis of an expenditure-based daily poverty line of 6347 Riel in Phnom Penh, 4352 Riel in other urban areas and 3503 Riel in rural areas. Poverty is particularly concentrated in the rural areas (24.6 percent).

Small-area estimation is a mathematical and statistical method that models data collected from one or more data sources, to produce estimates, for example of poverty, that are more accurate at small area level than using only data collected from each small area. The additional accuracy is achieved in many such models by “borrowing strength” for the estimate for a particular small area by using information from areas to which it is similar. Some small-area estimation techniques combine data from different sources. For example, census and new survey information may be combined to update estimates from the original census. Alternatively, and this is more usually the case for malnutrition estimates, a statistical model is fitted to survey data collected around the same time as the census, and this model is used to predict a variable not collected in the census, based on variables that are collected in both survey and census.

The first study involving small-area estimation of malnutrition estimates in Cambodia is the *Micro-level Estimation of the Prevalence of Stunting and Underweight Among Children in Cambodia* from the Ministry of Health, Cambodia / World Food Programme / Measure DHS+ - ORC Macro (2003). This study uses the World Bank method for small-area estimation to provide preliminary small-area estimates for stunting and underweight in children. The statistical models used are not given, and the detailed methodology is not discussed, but maps are provided at commune level and averages of estimated accuracy of the small-area estimates (as measured by their estimated standard errors given the fitted regression model is correct) are provided with discussion.

The World Bank method, popularly known as the Elbers Lanjouw and Lanjouw (ELL) method, has been commonly used in small-area estimation of poverty measures. In poverty studies, the most usual variable predicted is expenditure (or its logarithm) based on a model which includes education, age of household members, number of people in the household and type of house construction, among other variables. Poverty incidence, gap and severity are derived from the household level predictions of per capita expenditure. The poverty estimates are often mapped in detail, which is why this technique is sometimes given the generic title, “poverty mapping”. The maps can make interpretation simpler, but the central point is not the maps *per se*, but that poverty and relative poverty can be assessed at a much finer level at a much lower cost than by increasing the sample size sufficiently or rerunning the census. The statistical modelling has a cost, of course, but this is much lower than for a survey that is sufficiently large that it can produce estimates at this fine level. The cost of small-area estimation can be saved many times over by having better information at a finer level and maps for use in aid allocation.

The initial, national, small-area estimation of poverty in Cambodia was undertaken by Fujii (2002) for the World Food Programme, with support from the World Bank, using the 1998 population census and the 1997 Socio-economic Survey (CSES). By fitting a set of separate statistical models for expenditure on the logarithmic scale to sample information within strata for the CSES, applying these multiple models to the census data to predict expenditure at household level for all households, and summing transformations of the predictions, small-area estimates of poverty incidence, gap and severity were derived, and mapped at commune level. The methodology used was a standard application of the World Bank method (Elbers, Lanjouw and Lanjouw, 2001,

2003), which is now available as free software (PovMap – Zhao, 2006) from the World Bank website. Variations of the Elbers, Lanjouw and Lanjouw (ELL) method have been implemented for the World Bank in a number of other countries including Thailand (Healy, 2003), South Africa (Alderman et al., 2002), Brazil (Elbers et al. 2001), the Philippines (Haslett and Jones, 2005), and for the World Food Programme in Bangladesh (Jones and Haslett, 2003) and Nepal (Jones and Haslett, 2006).

More recently, Pinney (2007) has undertaken a small-area estimation exercise in Cambodia to update Fujii's estimates. Pinney has used the 2003/4 CSES and (rather than the population census, which as is common internationally is only conducted every ten years) has also used the commune database, also known as the Seila database or Seila commune database, or the National Committee for Decentralisation and Deconcentration (NCDD) database. The NCDD database is an annual census of villages and provides household information on a limited number of variables, which restricts the strength or predictive capacity (as measured by the percentage of variance that can be explained, usually denoted R^2) for statistical modelling, or predictions based on it. Pinney fits a multiple regression to the CSES data based on the variables also in the NCDD database, but without including the random effects (which would allow estimates of standard error via modelling of an additional commune or village level random component, fitted for example using the bootstrap as in ELL). The methods used by Pinney are potentially useful for providing an update to the 1997/8 estimates of Fujii, but the limited number of variables available for modelling may limit utility. The lack of information about standard errors is also a restriction, because poverty estimates are consequently of uncertain accuracy, so that it must remain unknown whether the method can provide sound poverty estimates at commune or district level.

The April 2007 World Food Programme report, *Integrated Food Security and Humanitarian Phase Classification (IPC) Pilot in Cambodia*, provides the most complete currently available comprehensive food security and vulnerability analysis. It has a direct focus on food, reflecting WFP's mandate. It contains a series of useful maps in appendices, including expenditure poverty (from CSES 2003/4) and underweight, stunting, and wasting in children. See also map on p44 – "Integrated Food security and Humanitarian Phase Classification (valid until 31.08.07) in Cambodia (as of 26.02.07)". None of these maps is however at commune level, so the need for small-area estimates of poverty remains. It has a useful reference list but no statistics, or relevant methodological details or content, although see Section 1.2 Methodology, which outlines a "meta analysis

approach".

This report and the *Micro-level Estimation of the Prevalence of Stunting and Underweight among Children in Cambodia* mentioned above warrant general comment about the relationship between small-area estimation and mapping. Small-area estimation of poverty, especially if extended from poverty incidence gap and severity, plus kilocalories, to stunting, underweight and wasting in children (as in Jones and Haslett, 2006), provides a detailed perspective on the spatial distribution of poverty. Other variables are also important however (e.g. health information, rainfall, and other Geographical Information System (GIS) data), even if these cannot be produced at such a fine level. For most users of this information, an atlas of maps is of much more general use than a detailed technical report on small-area estimation methodology, even if the technical report also contains finer level tabulated detail. The detailed small area report is however essential, as it provides a clear indication of the methodological foundation for small area maps (often called poverty maps) that are included in the atlas. Without sound use of small area methodology, and publication of the technical report that outlines that methodology, the accuracy and utility of a more generally-used atlas must remain in doubt.

In September 2007, the *Statistical Master Plan for Cambodia* was published by the National Institute of Statistics, Ministry of Planning. This document outlines the development of statistical functionality at NIS. Page 20, as part of section 6.3 "Censuses and surveys", contains detail on CSES as point 95, and Demographic and Health Surveys (DHS) as point 94. On page 21, there is Table 2, "Indicative Timetable for censuses and household surveys 2006-2015". Small areas, but not small-area estimation, are mentioned in item 89, p19.

1.2 Geographic and administrative units

For administrative purposes, Cambodia is divided into a total of 24 *provinces*, which are sub-divided into 193 *districts*. Within each district there are a number of *communes*, each comprising several *villages*: the smallest administrative unit. For some purposes, such as census enumeration and sampling frames for surveys, the larger villages are split into *enumeration areas*, but these are not in general well-defined administrative boundaries.

Table 1.1 shows the total number of each of these units in Cambodia, and their approximate sizes in terms of average number of households.

Table 1.1 Approximate number of administrative units at different levels.

	province	district	commune	village	ea
Number	24	193	1621	14073	28455
Mean no. households	117000	15000	1700	200	100

Key: ea= census enumeration area

The communes in Cambodia are commonly classified as belonging to one of three *regions*: the capital Phnom Penh, Other Urban and Rural. The districts are for some purposes grouped into five *Ecological Zones*: the capital Phnom Penh, the Coastal, Plains, Plateau/Mountain, and Tonle Sap zones. These ecological zones are characterized by differing economic conditions and activities.

Some knowledge exists on the general spatial pattern of poverty and malnutrition in Cambodia. Recent surveys (see Section 3) give estimates of economic and nutritional status for the whole country and for each province or group of provinces. However the accuracy of such estimates at a particular level depends crucially on the effective sample size at that level, so that at the district level and below the standard errors of survey-based estimates become too large to be useful because each is based on a small number of observations.

Effective targeting of development assistance, as advocated by the PRSP, requires a nation-wide overview of poverty and nutrition status at sub-district level. Estimates need to be precise, i.e. with small standard errors, so that the areas with the greatest need are identified correctly. Our analysis includes an investigation using small-area estimation methods of how finely the estimates of poverty and malnutrition indicators may be disaggregated while still maintaining a reasonable level of precision.

1.3 Poverty maps

The statistical technique of small-area estimation (Ghosh and Rao, 1994, Rao, 1999; Rao, 2003) provides a way of improving survey estimates at small levels of aggregation, by combining the survey data with information derived from other sources, typically a population census. A variant of this methodology has been developed by a research team at the World Bank specifically for the small-area estimation of poverty measures (Elbers,

Lanjouw and Lanjouw, 2001, 2003). The ELL method has been implemented in a number of countries including Thailand (Healy, 2003), Cambodia (Fujii, 2004), South Africa (Alderman et al., 2002) and Brazil (Elbers et al. 2001), Bangladesh (Jones and Haslett, 2003) and the Philippines (Haslett and Jones, 2005a). The methodology is described in detail in the next section. Some additional general methodological issues are covered in Haslett and Jones (2005b; 2010) and Haslett, Isidro and Jones (2010). Outputs, in the form of estimates at local level together with their standard errors, can be combined with GIS location data to produce a "poverty map" for the whole country, giving a graphical summary of which areas are suffering relatively high deprivation.

Our main purpose in producing such maps is to aid the planning of development assistance programmes. They could in addition prove useful as a research tool, for example by overlaying geographic, social or economic indicators.

1.4 Measures of poverty, under-nourishment and malnutrition

Poverty can be defined in a number of ways. The most common is the cost-of-basic-needs (CBN) approach, in which poverty lines are calculated to represent the level of per capita expenditure required to meet the basic needs of the members of a household, including an allowance for non-food consumption. First a food poverty line is established, being the amount necessary to meet basic food requirements. Then a non-food allowance is added, an amount equal to the typical non-food expenditure of households whose food expenditure is equal to the food poverty line. Because prices vary among geographical areas, poverty lines can be calculated separately for different regions for which price information is available. In Cambodia, these regions are Phnom Penh, Other Urban, and Rural. An important assumption in poverty mapping is that the prices faced by households are fairly homogenous within each region.

Thus in the CBN approach poverty measures are functions of household per capita expenditure. *Poverty incidence* for a given area is defined as the proportion of individuals living in that area who are in households with an average per capita expenditure below the poverty line. *Poverty gap* is the average distance below the poverty line, being zero for those individuals above the line. It thus represents the resources needed to bring all poor individuals up to a basic level. *Poverty severity* measures the average squared distance below the line, thereby giving more weight to the very poor. These three

measures can be placed in a common mathematical framework, the so-called FGT measures (Foster, Greer and Thorbeck, 1984):

$$P_\alpha = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{z - E_i}{z}\right)^\alpha \cdot I(E_i < z) \qquad (1.1)$$

where N is the population size of the area, E_i is the expenditure of the ith individual, z is the poverty line and $I(E_i < z)$ is an indicator function (equal to 1 when expenditure is below the poverty line, and 0 otherwise). Poverty incidence, gap and severity correspond to α = 0, 1 and 2 respectively. In our analysis we have produced estimates of all three measures down to commune level, using both the total poverty line and the food poverty line.

Three measures of malnutrition are considered, based on measurements of a child's height, weight and age. Stunting or low height-for-age is defined as having a height at least two standard deviations below the median height for a reference population. Underweight or low weight-for-age is similarly defined. Wasting is based on standardized weight-for-height, and low values can be a measure of acute malnutrition in some situations. The data used as a reference standard in these definitions was established in 1975 by the National Center for Health Statistics/Centers for Disease Control in the USA (Hamill, Dridz, Johnson, Reed et al., 1979). Implicit in the use of a single international reference standard is the assumption that variations in height and weight for children below five years are caused largely by environmental rather than genetic factors, although even without this assumption it can provide a fixed reference point in international comparisons.

In this report we consider the nutrition status of children below the age of 60 months (i.e. five years). Within a particular region stunting is defined as the proportion of such children with a standardized height-for-age (HAZ) value below –2: children below –3 are considered "severely stunted". Similarly underweight is the proportion with a standardized weight-for-age (WAZ) value below –2, and severe underweight below –3. Stunting can be regarded as evidence of chronic malnutrition. Underweight reflects both chronic malnutrition and acute malnutrition. It is a current condition resulting from inadequate food intake, past episodes of under-nutrition or poor health conditions. Wasting is the proportion with a standardized weight-for-height (WHZ) value under -2, and severe wasting below -3. Wasting can be an indicator of acute malnutrition. Our original aim in this report was to construct commune-level maps for these three measures.

However we were unable to find good predictive models for weight-for-height, so small area-estimates of wasting have not been produced.

Caloric intake is measured on the basis of kilocalorie consumption. The SAE methodology was also applied to predict kilocalorie consumption per person in the population census. This was then compared to the minimum dietary energy requirement (MDER) set for Cambodia which is 1770 kilocalories per person. All members of a household are considered under-nourished if predicted kilocalorie intake per person for that household is below this norm. In principle, in the same way that poverty can be summarized on the basis of the poverty *incidence*, *gap* and *severity*, caloric intake can be summarized on the basis of these three measures.

2. Methodology

We present in this section a brief overview of small-area estimation and the ELL method. Details of the implementation in Cambodia are given in Section 4.

2.1 Small-area estimation

Small-area estimation refers to a collection of statistical techniques designed for improving sample survey estimates through the use of auxiliary information (Ghosh and Rao, 1994; Rao, 1999; Rao, 2003). We begin with a target variable, denoted Y, for which we require estimates over a range of small subpopulations, usually corresponding to small geographical areas. (In this report Y is log-transformed per capita expenditure for poverty measures, and standardized height-for-age or weight-for-age for the malnutrition indicators, stunting and underweight.) Direct estimates of Y for each subpopulation are available from sample survey data, in which Y is measured directly on the sampled units (households or eligible children). Because the sample sizes within the subpopulations typically will be very small, these direct estimates will have large standard errors and hence not be reliable. Indeed, some subpopulations may not be sampled at all in the survey. Auxiliary information, denoted X, can be used under some circumstances to improve the estimates, giving lower standard errors.

In the situations examined in this report, X represents additional variables that have been measured for the whole population, either by a census or via a GIS database. A relationship between Y and X of the form

$$Y = X\beta + u$$

can be estimated using the survey data, for which both the target variable and the auxiliary variables are available. Here β represents the estimated regression coefficients giving the effect of the X variables on Y, and u is a random error term representing that part of Y that cannot be explained using the auxiliary information. If we assume that this relationship holds in the population as a whole, we can use it to predict Y for those units for which we have measured X but not Y. Small-area estimates based on these predicted Y values will often have smaller standard errors than the direct estimates, even allowing for the uncertainty in the predicted values, because they are based on much larger samples.

Thus the idea is to "borrow strength" from the much more detailed coverage of the census data to supplement the direct measurements of the survey.

2.2 Clustering

The units on which measurements have been made are often not independent, but are grouped naturally into clusters of similar units. Households tend to cluster together into villages or other small geographic or administrative units, which are themselves relatively homogenous. Put simply, households that are close together tend to be more similar than households far apart. When such structure exists in the population, the regression model above can be more explicitly written as

$$Y_{ij} = X_{ij}\beta + c_i + e_{ij} \tag{2.1}$$

where Y_{ij} represents the measurement on the *j*th unit in the *i*th cluster, c_i the error term held in common by the *i*th cluster, and e_{ij} the household-level error within the cluster. The relative importance of the two sources of error can be measured by their respective variances σ_c^2 and σ_e^2. Ghosh and Rao (1994) give an overview of how to obtain small-area estimates, together with standard errors, for this model. Where individual level data is available, as it is for stunting, underweight and wasting in children under five, an additional error term at child level within household is added. In the general explanation given below we focus on equation (2.1) in order to establish general principles useful for distinguishing the characteristics of variation at 'higher' and 'lower' levels. When there are three error terms rather than two, the three form a sequence in which the cluster remains the highest level of aggregation, household takes an intermediate status, and individual level variation is at the finest level. There is also the possibility of including a small area level error term at the greatest level of aggregation. Doing so does not affect the small-area estimates themselves, but does have the potential to increase standard error estimates, perhaps markedly. The small area models of Rao (2003) contain such an error term, but those of Elbers, Lanjouw and Lanjouw (2003) do not. Checking for the size of the small area-level error variance is strongly recommended, because if it is sufficiently large its omission leads to small-area estimates with understated standard errors and hence overstated accuracy. The issue is addressed for small-area estimation in Nepal in Jones and Haslett (2006), where the effect of the small area variance on the standard error estimates was found to be negligible. For Cambodia, see Section 5 below. Theoretical aspects of this question are discussed in detail in Haslett and Jones (2010).

We note that the auxiliary variables X_{ij} may be useful primarily in explaining the cluster-level variation, or the household-level variation. The more variation that is explained at a particular level, the smaller the respective error variance, σ_c^2 or σ_e^2. The estimate for a particular small area will typically be the average of the predicted *Y*s in that area. Because the standard error of a mean gets smaller as the sample size gets bigger, the contribution to the overall standard error of the variation at each level, household and cluster, depends on the sample size at that level. The number of households in a small area will typically be much larger than the number of clusters, so to get small standard errors it is of particular importance that, at the higher level, the unexplained cluster-level variance σ_c^2 should be small. Two important diagnostics of the model-fitting stage, in which the relationship between *Y* and *X* is estimated for the survey data, are the R^2 measuring how much of the variability in *Y* is explained by *X*, and the ratio $\sigma_c^2 \ / \ (\sigma_c^2 + \sigma_e^2)$ measuring how much of the unexplained variation is at the cluster level. Note that although σ_c^2 and σ_e^2 are parameters they are different for different models with different regressors. GIS data and cluster-level means can be particularly useful in lowering this ratio. Some care is required when using R^2as a diagnostic however, because it very much depends on the level of aggregation, and the level of aggregation in the fitted model is very much less than that of the small-area estimates. So, while high R^2 values .are good, they are not essential, provided the variances at the finest level are sufficiently larger than those at more aggregated levels. This diminution of R^2 is especially apparent where person level data is being used (as for stunting, underweight and wasting), rather than household level data (as for kilocalories and expenditure modelling, where the variation within household, which may be large, is effectively omitted from the estimation of R^2 from the model due to data aggregation to household level).

Another important aspect of clustering is its effect on the estimation of the model. The survey data used for this estimation cannot be regarded as a simple random sample, because they have been obtained from a complex survey design which although it is random, nevertheless involves stratification and cluster sampling. To account properly for the complexity of the survey design requires the use of specialised statistical routines (Skinner et al., 1989; Chambers and Skinner, 2003, Lehtonen and Pakhinen, 2004, Longford, 2005) in order to get consistent estimates for the regression coefficient vector β and its variance V_β.

2.3 The ELL method

The ELL methodology was designed specifically for the small-area estimation of poverty measures based on per capita household expenditure. Here the target variable Y is log-transformed expenditure, the logarithm being used to make more symmetrical the highly right-skewed distribution of untransformed expenditure. It is assumed that measurements on Y are available from a survey.

The first step is to identify a set of auxiliary variables X that are in the survey and are also available for the whole population. It is important that these should be defined and measured in a consistent way in both data sources. The model (2.1) is then estimated for the survey data, by incorporating aspects of the survey design for example through use of the "expansion factors" or inverse sampling probabilities. The residuals $\hat{u}_{ij}$ from this analysis are used to define cluster-level residuals $\hat{c}_i = \hat{u}_{i.}$, the dot denoting averaging over j, and household-level residuals $\hat{e}_{ij} = \hat{c}_i - \hat{u}_{ij}$.

It is usually assumed that the cluster-level effects c_i all come from the same distribution, but that the household-level effects e_{ij} may be heteroscedastic. This can be modelled by allowing the variance σ_e^2 to depend on a subset Z of the auxiliary variables:

$$g(\sigma_e^2) = Z\alpha + r$$

where $g(.)$ is an appropriately chosen link function, α represents the effect of Z on the variance and r is a random error term. Fujii (2004) uses a version of the more general model of ELL involving a logistic-type link function, fitted using the squared household-level residuals. Fujii's model is:

$$\ln\left(\frac{\hat{e}_{ij}^2}{A - \hat{e}_{ij}^2}\right) = Z_{ij}\alpha + r_{ij} \qquad (2.2)$$

From this model the fitted variances $\hat{\sigma}_{e,ij}^2$ can be calculated and used to produce standardized household-level residuals $\hat{e}_{ij}^* = \hat{e}_{ij} / \hat{\sigma}_{e,ij}$. These can then be mean-corrected or mean-centred to sum to zero, either across the whole survey data set or separately within each cluster.

In standard applications of small-area estimation, the estimated model (2.1) is applied to the known X values in the population to produce predicted Y values, which are then averaged over each small area to produce a point estimate, the standard error of which is

inferred from appropriate asymptotic theory. In the case of poverty mapping, our interest is not always directly in Y but in several non-linear functions of Y (see Section 1.4). The ELL method obtains unbiased estimates and standard errors for these by using a bootstrap procedure as described below.

2.4 Bootstrapping

Bootstrapping is the name given to a set of statistical procedures that use computer-generated random numbers to simulate the distribution of an estimator (Efron and Tibshirani, 1993). In the case of poverty mapping, we construct not just one predicted value

$$\hat{Y}_{ij} = X_{ij}\hat{\beta}$$

(where $\hat{\beta}$ represents the estimated coefficients from fitting the model) but a large number of alternative predicted values

$$Y_{ij}^{b} = X_{ij}\beta^{b} + c_i^{b} + e_{ij}^{b} \ , \quad b = 1, \ldots B$$

in such a way as to take account of their variability. The statistical analysis of the chosen model for Y yields information on how to appropriately insert variability into the calculation of the predicted values. We know for example that $\hat{\beta}$ is an unbiased estimator of β with variance V_β, so we draw each β^b independently from a multivariate normal distribution with mean $\hat{\beta}$ and variance matrix V_β. The cluster-level effects h_i^b are taken from the empirical distribution of h_i, i.e. drawn randomly with replacement from the set of cluster-level residuals $\hat{h}_i$, since the appropriate cluster level residual is known only for the clusters in the sample not all the clusters in the census. To take account of unequal variances (heteroscedasticity) in the household-level residuals, we first draw α^b from a multivariate normal distribution with mean $\hat{\alpha}$ and variance matrix V_α, combine it with Z_{ij} to give a predicted variance and use this to adjust the household-level effect

$$e_{ij}^{b} = e_{ij}^{*b} \times \sigma_{e,ij}^{b}$$

where e_{ij}^{*b} represents a random draw from the empirical distribution of e_{ij}^{*}, either for the whole data set or just within the cluster chosen for h_i (consistently with the mean-centring of Section 2.3).

Each complete set of bootstrap values Y_{ij}^{b}, for a fixed value of *b*, will yield a set of small-area estimates. In the case of poverty estimates we exponentiate each *Y* to give predicted expenditure $E_{ij} = \exp(Y_{ij})$, then apply equation (1.1). This is not equivalent to totalling the Y_{ij} in each small area and exponentiating, which is one reason that fitting the model at household (or individual level in the case of a three level model) is the better alternative. The mean and standard deviation of a particular small-area estimate, across all *b* values, then yields a point estimate and its standard error for that area.

2.5 Interpretation of standard errors

The standard error of a particular small-area estimate is intended to reflect the uncertainty in that estimate. A rough rule of thumb is to take two standard errors on each side of the point estimate as representing the range of values within which we expect the true value to lie. When two or more small-area estimates are being compared, for example when deciding on priority areas for receiving development assistance, the standard errors provide a guide for how accurate each individual estimate is and whether the observed differences in the estimates are indicative of real differences between the areas. They serve as a reminder to users of poverty maps that the information in them represents estimates, which may not always be very precise. A particular way of incorporating the standard errors into a poverty map is suggested in Section 4.

The size of the standard error depends on a number of factors. The poorer the fit of the model (2.1), in terms of small R^2, large σ_c^2 or σ_e^2, or a large $\sigma_c^2/(\sigma_c^2+\sigma_e^2)$ ratio, the more variation in the target variable will be unexplained and the greater will be the standard errors of the small-area estimates. The population size, in terms of both the number of households and the number of clusters in the area, is also an important factor. Generally speaking, standard errors decrease proportionally as the square root of the population size. Standard errors will be acceptably small at higher geographic levels but not at lower levels. If we decide to create a poverty map at a level for which the standard errors are generally acceptable, there will be some, smaller, areas for which the standard errors are larger than we would like.

The sample size used in fitting the model is also important. The bootstrapping methodology incorporates the variability in the estimated regression coefficients $\hat{\alpha}$, $\hat{\beta}$. If the sample size is small these estimates will be very uncertain and the standard errors of the small-area estimates will be large. This problem is also affected by the number of

explanatory variables included in the auxiliary information, X and Z. A large number of explanatory variables relative to the sample size increases the uncertainty in the regression coefficients. We can always increase the apparent explanatory power of the model (i.e. increase the R^2 from the survey data) by increasing the number of X variables, or by dividing the population into distinct subpopulations and fitting separate models in each, but the increased uncertainty in the estimated coefficients may result in an overall loss of precision when the model is used to predict values for the census data. We must take care not to "over-fit" the model.

There will be some small uncertainty in the estimates, and indeed the standard errors, due to the bootstrapping methodology, which uses a finite sample of bootstrap estimates to approximate the distribution of the estimator. This could be decreased, at the expense of computing time, by increasing the number of bootstrap simulations B.

Finally, the integrity of the estimates and standard errors depends on the fitted model being correct, in that it applies to the census population in the same way that it applied to the sample. This relies on good matching of survey and census to provide valid auxiliary information. We must also take care to avoid, as much as possible, spurious relationships or artefacts which appear, statistically, to be true in the sample but do not hold in the population. This can be caused by fitting too many variables, but also by choosing variables indiscriminately from a very large set of possibilities. Such a situation could lead to estimates with apparently small, but spurious, standard errors. For this reason the final step in poverty mapping, field verification, is extremely important.

The requirement for variables to match in this way between survey and census is one reason that special care must be taken if survey and census are not from the same period. The changes between periods can be structural changes, i.e. the interpretation of a particular variables has changed, or simply a change in level. Both types of change have the potential to add to standard errors of estimates, and in some cases to produce bias.

3. Data Sources

3.1 Cambodia Socio-economic Survey, 2009 (CSES2009)

The Cambodia Socio-economic Survey was carried out for the first time in 1993/94 by the National Institute of Statistics (NIS), and repeated in 1996, 1997, 1999, 2004, 2007 and 2009. The 2007 survey was a smaller interim survey using a sub-sample of primary sampling units from the 2004 survey, designed to produce statistics only for the whole country and the Phnom Penh, Other Urban and Rural regions; the 2009 survey, like 2004, was designed to give reliable estimates down to province level.

The CSES broadly follows the methodology of the World Bank's Living Standard Measurement Survey. It contains an integrated household questionnaire designed to collect data at both household and individual level on socio-demographic characteristics in addition to detailed information about expenditure and food consumption patterns. Consumption is recorded using both recall and diary methods.

The sample design for CSES2009 used a stratified cluster sampling technique. The strata were the urban and rural parts of each province, giving a total of 48 strata. The Primary Sampling Units (PSUs) were villages, although the larger villages were subdivided based on enumeration areas from the 2008 census. In the first stage a total of 720 PSUs (240 urban and 480 rural) were chosen by stratified random sampling, using Probability Proportional to Size (PPS) sampling with the number of households as a measure of size. Then a systematic sample of either 10 (urban) or 20 (rural) households was taken within each sampled PSU. This gave a total sample size of 12000 households. Ultimately a total of 11971 households were enumerated.

In our analysis, we identified a few strata with only one PSU. We merged these with geographically adjacent strata in order to be able to calculate standard errors for estimated model parameters using standard survey regression techniques.

Because the sample size at a particular level has an important bearing on the precision of estimates at that level, we present in Table 3.1 a summary of the coverage of CSES2009 at various levels and the mean and minimum number of households and PSUs at each level. The number of provinces, districts and communes in CSES2009 can be compared with the numbers in Cambodia as a whole via Table 1.1. The coverage is adequate at

provincial level, except for a few of the smaller provinces where the sample sizes are small. Twenty-two of the 193 districts are not sampled, and at least one of the others has only one PSU. Thus we cannot expect to get precise estimates directly from CSES2009 at district or sub-district levels.

Table 3.1 Structure of CSES2009 at various levels.

	province	district	commune	village
Contains	24	171	621	715
Mean households	499	70	19.3	16.7
Min households	39	19	9	8
Mean PSUs	30	4.2	1.2	1.01
Min PSUs	3	1	1	1

Key: PSU=primary sampling unit

The target variable available in CSES2009 and used in this study is monthly per capita consumption expenditure, averaged at the household level. Calculation of total household-level consumption expenditure and the regional poverty lines was conducted by the NIS. Table 3.2 below gives the official poverty lines.

Table 3.2 Poverty lines (Riel per person per day)

	Phnom Penh	Other Urban	Rural
Total poverty line	6347	4352	3503
Food poverty line	3121	2607	2300

3.2 Cambodia Anthropometrics Survey, 2008 (CAS2008)

The 2008 Cambodia Anthropometrics Survey was carried out by NIS with technical and financial support from UNICEF. The main purpose of the survey was to provide policymakers and planners with updated information on nutrition in light of steep increases in the price of food. Anthropometrical measures were taken on children aged 0-60 months to determine nutritional status as described in Section 1.4. In order to provide a comprehensive view on nutrition in the country, data on micronutrient deficiency, food consumption, disease, coping strategies, infant/young child feeding, and health services were included in the survey. The survey was designed to give data on the nutritional

status of children in the country and in 19 domains defined as either single provinces or small groups of provinces.

CAS2008 sampled 7268 households with children aged 0-60 months from 760 PSUs, corresponding closely to enumeration areas from the 2008 national census, in 19 strata comprising the larger provinces and groups of smaller provinces. Most contributing households had only one eligible child, but 16% had two or more (see Table 3.3).

Table 3.3 Eligible Children (0-4 years) per Household, CAS2008.

No. of children	1	2	3	4	Total
No. of households	6046	1159	59	4	7268

The final dataset used consisted of 8537 children in 760 PSUs. The structure is shown in Table 3.4. Eight of the 193 districts are not included, and of those present some have very few PSUs, so direct estimates at district and sub-district are not possible.

Table 3.4 Structure of CAS2008 dataset at various levels.

	province	district	commune	PSU
Contains	24	185	709	760
Mean children	356	46	12	11.2
Min children	21	9	5	5
Mean PSUs	31.7	4.1	1.1	
Min PSUs	2	1	1	

Key: PSU=primary sampling unit

The target variables for estimating stunting, underweight and wasting are height-for-age, weight-for-age, and weight-for-height (see Section 1.4). These were calculated from the raw height and age measurements using a programme provided by the WHO.

The CAS2008 report (Ministry of Health, 2002) gave the national prevalence of underweight as 28.8%, not significantly different from the 2005 estimate. The estimated prevalence of stunting was reported as 39.5%, which represents an improvement over the 2005 figure. Nutritional status was found to vary with the age and sex of the child, place of residence (urban/rural) and wealth status of the parents. However poor nutrition was found to be a national issue affecting every sector of society.

3.3 Cambodia Demographic and Health Survey, 2010 (CDHS2010)

The 2010 Cambodia Demographic and Health Survey, the third in a series of demographic surveys, was carried out by NIS with technical support from ORC Macro and financial support from USAID. The survey was designed to provide up-to-date information on infant and child mortality, fertility preferences, family planning behaviour, maternal mortality, utilization of maternal and child health services, health expenditures, women's status, and knowledge and behaviour regarding HIV/AIDS and other sexually transmitted infections. Anthropometrical measures were taken on selected children (aged 0-60 months) to determine nutritional status as described in Section 1.4, in addition to detailed information on household demographic characteristics, environmental conditions and child feeding and caring practices.

CDHS2010 sampled 16344 households from 611 PSUs, corresponding to enumeration areas from the 2008 census, in 38 strata formed from the urban and rural parts of 19 provinces or groups of provinces. Our interest is in the nutritional status of children below five years, so households with no eligible children were eliminated. Most contributing households had only one eligible child, but 24% had two or more (see Table 3.5).

Table 3.5 Eligible Children (0-4 years) per Household, CDHS2010.

No. of children	1	2	3	4	5	Total
No. of households	2429	674	64	9	1	3177

The final dataset used consisted of 4010 children in 607 PSUs. The structure is shown in Table 3.6. Six of the 75 districts are not included, and of those present some have very few PSUs, so direct estimates at district and sub-district are not possible.

The target variables of height-for-age, weight-for-age, and weight-for-height (see Section 1.4) were calculated using the WHO's Stata programme. The CDHS 2001 report (NIS, 2011) gave the national prevalence of stunting as 40%, and underweight 28%.

Table 3.6 Structure of CDHS2010 dataset at various levels.

	province	district	commune	PSU
Contains	24	187	526	607
Mean children	167	21.4	7.6	6.6
Min children	7	3	1	1
Mean PSUs	25.3	3.2	1.2	
Min PSUs	1	1	1	

Key: PSU=primary sampling unit

3.3 Census, 2008 (Census2008)

The Cambodian Government has committed to conducting a population census every ten years, commencing in 1998 which saw the first census since Cambodia became a democratic country. The 2008 census was carried out in early March, with a specified census data of 3rd March 2008. Three types of form were administered: Form A enumerated buildings, recording their purpose and construction materials; Form B enumerated households, recording ownership, utilities and appliances; Form C enumerated individual people, recording demographic variables. Since all three contained consistent codes for regional and household identifiers, data from the three forms could be merged to create a single household-level dataset in a format compatible with CSES2009, and a child-level dataset compatible with CAS2008 and CDHS2010.

The census collected information on all residents of Cambodia based on their usual place of residence, but excluding temporary visitors, tourists, resident foreign diplomats and refugees. Households were classified as residential or institutional type (e.g. hostels, hospitals, jails). Since the survey data only covered residential households, it was decided to restrict the census data to only residential households.

The enumerated population on census night was declared to be 13,395,682 in 2,841,897 households, with 19.5% living in urban areas. The structure of each of the two derived census datasets (household- and child-level) is shown in Tables 3.7 and 3.8, in terms of number of households and number of enumeration areas.

Table 3.7 Structure of Census household dataset at various levels.

	province	district	commune	village	ea
Contains	24	193	1621	14073	28455
Mean households	117228	14578	1736	200	99
Min households	7193	850	60	3	3
Mean ea	1186	147	17.6	2	
Min ea	66	10	2	1	

Key: ea= enumeration area

Table 3.8 Structure of Census child (under 5) dataset at various levels.

	province	district	commune	village	ea
Contains	24	193	1621	14073	28448
Mean children	56890	7074	842	97	48
Min children	3800	518	48	2	1
Mean households	45757	5690	677	78	39
Min households	3019	400	32	2	1
Mean ea	1186	147	17.5	2	
Min ea	66	10	2	1	

Key: ea=enumeration area

3.4 Commune Database, 2009 (CDB2009)

The commune database (CDB), also known as the Seila database or Seila commune database, or the National Committee for Decentralisation and Deconcentration (NCDD) database, collects information on the demographic, socio-economic and physical assets of each village and commune in Cambodia. Starting in 2002, the CDB is maintained by the Ministry of Planning, with data collection taking place at the end of the year. Data are collected by Village Chiefs and Commune Clerks, giving some village-level and some commune-level variables.

The data are used by communes for preparation of socio-economic profiles at commune, district and provincial levels, as part of the annual planning exercises. The CDB is also used by the Ministry of Planning to produce a poverty index for the allocation of investment funds for communes.

The CDB was not designed for measuring economic poverty as defined in section 1.4 since it does not directly measure household consumption expenditure. However it does

contain variables that may be useful proxies for village- and commune-level economic poverty incidence. Pinney (2007) and Hou, Ny and Karim, (2010) both use commune level (rather than household level data as for small-area estimation) to develop estimates of poverty incidence at commune level. Hou, Ny and Karim is particularly interesting. Their results will be discussed in Section 5.3 and compared with the small-area estimates via ELL. Hou, Ny and Karim use an IDPoor database (see Section 3.5 below) that at 2009 contained only a limited number of communes, to set up a model predicting poverty incidence, and then they apply their regression model to the Commune Data Base which contains information for most of the communes in Cambodia. In concept then, the idea is similar to ELL, requiring matching of predictor variables in two different data sources. The sample available for IDPoor for 2009 is not random and the model can only be fitted at aggregate commune level so, unlike ELL, their technique is not able to provide particularly sound estimates of accuracy, nor can it be extended (as ELL can) to providing poverty gap or severity. However Hou, Ny and Karim (2010) does have the major advantage over ELL that it can provide annual updates, rather than only being implementable following a near coincident survey and population census. (c.f. Isidro, Haslett and Jones, 2010a & 2010b).

The CDB2009 contained information on 13983 villages, slightly less than the total number in Census2008. Moreover there were some inconsistencies in the area codes, not all of which could be resolved. Ultimately about 200 villages in Census2008 remained unmatched with CDB2009 information.

3.5 Identification of Poor Households (IDPoor –GTZ)

The IDPoor information is collected as part of the German Technical Cooperation Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ) with the Cambodian Government. IDPoor's objective is to "develop an improved procedure for identifying poor households, reach official consensus on the common use of a harmonised procedure, and put it into practice in selected provinces".

The intention is that this "process will in the medium term assist various sector programmes to reach the poor more effectively. It will also be possible to reduce the overall administrative costs of selecting target groups and allow scarce public resources to be redistributed in a way that benefits the poorest target groups. Already-identified potential areas for application of a harmonised procedure are the provision of medical

services through Health Equity Funds, improving access of the poor to education through targeted financial support, the provision of services related to rural development, and allocation of land to the poor" (Cambodia German Cooperation, 2006). The project consists of a national-level and a provincial-level component.

In 2010 and 2011 combined, IDPoor has been further extended to cover all Cambodia communes. IDPoor is nevertheless not intended to be a formal, statistically based methodology for measuring poverty incidence, but is particularly useful as a basis for targeting poor households within communes.

4. Implementation

4.1 Selection of auxiliary data

The auxiliary data X used to predict the target variable Y can be classified into two types: the survey variables, obtainable or derivable from the survey at household or individual level, and area-level variables applying to particular geographic units that can be merged from other sources into the survey data using area codes (e.g. province-district-commune-village-ea codes). The latter includes means of census variables calculated at enumeration area level from the census data.

As noted earlier, it is important that any auxiliary variables used in modelling and predicting should be comparable in the estimation (survey) data set and the prediction (census) data set. In the case of survey variables, we begin by examining the survey and census questionnaires to find out which questions in each elicit equivalent information. In some cases equivalence may be achieved by collapsing some categories of answers. For example in the 2008 census questionnaire there are eight categories for Roof Material, whereas in CSES2009 there are ten such categories, some of which appear to correspond exactly to the census categories and others which do not: a new categorization needs to be defined into which the Census2008 and CSES2009 categories can be mapped. A preliminary identification and matching of common survey and census variables, in consultation with NIS staff, was reported by Haslett et al (2010) for CSES2009 and CAS2008. This process was repeated with CDHS2010, and all three sets of common variables were then subjected to statistical checks to ensure that the corresponding survey and census variables matched statistically as well as conceptually. In the case of categorical data we compare proportions in each category: for numerical data, such as household proportion of females, we compare the means and standard deviations. For this purpose confidence intervals can be calculated for the relevant statistics in the survey data set, taking account of the stratification and clustering in the sample design. The equivalent statistic for the census data should be within the confidence interval for the survey. Failures in statistical matching can sometimes be resolved by further collapsing categorical variables. A list of matching variables for each of the survey datasets is given in Appendices A.1 to A.3.

For modelling purposes the first level of each categorical variable was dropped so that the first category becomes the reference category with which others are compared. We also

created some new variables from this basic list, for example mean-corrected squared household size defined as hhszsq=(hhsize-4.77)^2 , and interactions between basic variables such as region×hhsize which modifies the effect of household size according to whether the household is in the Phnom Penh, Other Urban or Rural region. The variable ezone was also added to allow for differences between Ecological Zones as defined in section 1.2.

For the CAS dataset we faced the difficulty that some of the variables expected to be useful in predicting nutrition status of individual children, for example educational attainment of the mother, were not directly available in the census because there was no explicit link between children in the census dataset and their birth mother. We attempted to circumvent this problem by creating these variables at household level, averaging over potential mothers in the household. This inevitably introduces some measurement error into these variables.

Generally, variables which are in either census dataset, but are either not in the survey or do not match properly, can still be used by forming regional averages and merging them with the survey data using regional indicators. The inclusion of these census means should be straightforward since they can be merged with the survey and census data using indicators for the geographical unit to which each household or individual belongs. This can be problematic in practice however, because of changing boundaries and the creation of new units or codes. Most of these problems were solved in collaboration with NIS, and the few remaining unmatched households should have negligible influence on the final estimates. Appendix A.4 gives a list of all the census means considered in the modelling process. These variables have all been averaged at enumeration-area level. Since the CDHS2010 dataset did not have enumeration-area codes, the census means were recalculated at village level for merging with the CDHS2010.

Poverty estimates from the CDB2009 via IDPoor were not included in the small area modelling, because keeping them out of the small-area estimation modelling enabled later direct comparison at commune level of the CDB based poverty incidence estimates with those from the small-area estimation. These two sets of poverty incidence estimates at commune level are independent, because one uses IDPoor and CDB, and the other the 2008 Census of Population and CSES2009, so any links between them provide evidence of the veracity of both. In Appendix A.5 is a list of CDB variables considered by the Cambodia National Committee for Sub-National Democratic Development (NCDD) in

their modelling.

4.2 First stage regressions

The selection of an approprite model for (2.1) is a difficult problem. We have a large number of possible predictor variables (36 + 52 = 98 for CSES2009: see Appendix A) to choose from, with inevitably a good deal of interrelationship between them in the form of multicollinearity. If we also include two-way interactions there are well over a thousand. (A “two-way interaction” is the product of two basic or “main-effect” variables). Squares or other transformations of numerical variables could also be considered. As noted in Section 2.5, we must be careful not to over-fit, so the number of predictors included in the model should be small compared to the number of observations in the survey, but there is also the problem of selecting a few variables from the large number available which appear to be useful, only to find (or even worse, not find) an apparently strong statistical relationship in the survey data, which does not hold for the population as a whole.

The search for significant relationships over such a large collection of variables must inevitably be automated to a certain extent, but we have chosen not to rely entirely on automatic variable selection methods such as stepwise or best-subsets regression. See Miller (2002) for a general discussion of subset selection. We have generally adopted the principle of hierarchical modelling in which higher-order terms such as two-way interactions are included in the model only if their corresponding main-effects are also included. Thus we begin with main-effects only, and add interaction and nonlinear terms carefully and judiciously. However due to the failure of statistical matching for several demographic variables in Phnom Penh, interactions with region were included as this allowed omission of the variables which had inadequate statistical matching in Phnom Penh, while still allowing these variables to be included for the remaining regions. We look not just for statistical significance but also for a plausible relationship. For example, the effect of household size (hhsize) on log expenditure was investigated by first fitting hhsize as a categorical variable, and then choosing a parsimonious functional form that produces the correct approximate shape. This is shown in Figure 4.1.

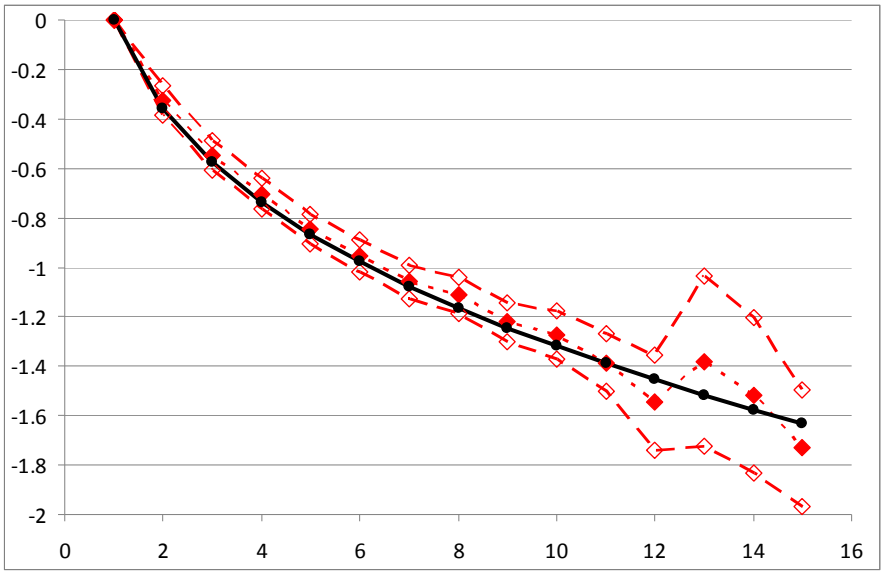


Figure 4.1 Coefficients for effect of hhsize *on log expenditure, fitted as categorical with 95% confidence limits (dashed lines), and fitted linear model* $\beta_0+\beta_1$hhsize$+\beta_2$ln(hhsize).

This process was repeated for all numerical variables (number of rooms, number of cellphones etc.) to give in each case a parsimonious functional representation of the effect of each numerical auxiliary variable on the target variable. For example, the effect of increasing number of rooms on log expenditure in CSES2009 seems to attenuate after numroom=4, so larger values were set to four. Following the initial fit, some categorical variables were collapsed further to give smaller numbers of distinct categories when there was no significant difference between the estimated effects of similar categories. For example, the eight categories of wall in CSES2009 were eventually collapsed to two: "bamboo" and "other".

Other implementations of ELL methodology have fitted separate models for each stratum defined by the survey design. This has the advantage of tailoring the model to account for the different characteristics of each stratum, but it can increase the problem of over-fitting if some strata are small. We chose initially to try for one model across the whole country, and then to use regional interaction terms as necessary to allow for modelling differences between regions. This has the advantage of more stable parameter estimates and a better chance of finding genuine relationships that apply outside of the estimation data. Following this approach for modelling log expenditure in CSES2009, we obtained an R^2 value of 49.3% (see Appendix B.1).

We were less successful at finding good predictive models based on R^2 for the other target variables (height-for-age, weight-for-age, weight-for-height and log per capita kilocalorie consumption). For modelling height-for-age and weight-for-age in CAS2008 and CDHS2010 achieved R^2 values were in the range 15-20%. The resulting models are given in Appendices B.2, B.3, B.4, and B.5. Although the R^2 for each was 20% or less, it is interesting to note that the major component of unexplained variation for each appears to be between children in the same household. This is discussed further in the next Section. We were unable to find predictive models for weight-for-height, so do not provide small-area estimates for this variable.. Regarding log per capita kilocalorie consumption, we achieved an R^2 value of only about 13%. Here there was still significant unexplained variability at cluster level, suggesting that small-area estimation for communes may not have sufficient precision. Hence, neither models nor maps for weight-for-age / wasting or kilocalories per capita have been included in this report.

We departed from the usual ELL implementation in our use of a single-stage, robust regression procedure for estimating model (2.1). This has the advantages of accounting for the survey design and obtaining consistent estimates of the covariance matrices in a single step. These covariance matrices were saved, along with the parameter estimates and both household- and cluster-level residuals (as defined in Section 2.3), for implementation of the prediction step.

4.3 Variance modelling

Like Healy et al (2003) we amended the regression model (2.2) for the household-level variance to prevent very small residuals from becoming too influential. We used a slightly different amendment:

$$L_{ij} \equiv \ln\left(\frac{\hat{e}_{ij}^2 + \delta}{A - \hat{e}_{ij}^2}\right) = Z_{ij}\alpha + r_{ij}$$

where δ is a small positive constant and A is chosen to be just larger than the largest $\hat{e}_{ij}^2$ (e.g. δ = 0.0001, $A = 1.05 \times \max \hat{e}_{ij}^2$). These choices can be justified empirically by graphical examination of the L_{ij}, which should show neither abrupt truncation nor extreme outliers. The predicted value of the household-specific variance, using the delta method, then becomes:

$$\sigma_{e,ij}^2 = \left[\frac{AB_{ij} - \delta}{1 + B_{ij}}\right] + \frac{1}{2}\hat{\sigma}_r^2 \left[\frac{(A+\delta)B_{ij}(1 - B_{ij})}{(1 + B_{ij})^3}\right]$$

where $B = e^{Z\alpha}$.

There was however very little heteroscedasticity in any of our models. For example the heteroscedasticity regressions for log expenditure gave an R^2 value of below 1%. These models for variance essentially control for outliers, by adjusting or shrinking large residuals $\hat{e}_{ij}$ toward zero. They form an explicit part of the ELL methodology. Other forms are possible. Even skipping this step would have been acceptable given the low R^2 values. However, in keeping with the need to maintain international comparison, for example with Cambodia, South Africa, Bangladesh, and the Philippines, heteroscedasticity modelling has been used here for log expenditure, using Region as the only covariate. Despite the negligible R^2, the coefficients for region 2 (other urban) and region 3 (other rural) were statistically significant in the regression, and there is a priori expectation of structural differences between regions.

For modelling height-for-age and weight-for-age we found it necessary to depart from the usual methodology, in order to account for the expected correlation in these measures between children in the same household. We now have a three-level model, in which the regression residuals can be decomposed into three components

$$u_{ijk} = c_i + h_{ij} + e_{ijk} \tag{4.1}$$

for child k in household j of cluster (PSU) i. The variances σ_c^2, σ_h^2, σ_e^2 of the respective components can be estimated by maximum likelihood (ML) or restricted maximum likelihood (REML), and the cluster- and household-level residuals (or random effects) derived as empirical best linear unbiased predictors (EBLUPs). For methodological details see Laird and Ware (1982) and Robinson (1991). The alternative of defining household-level residuals to be the average of the regression residuals for each respective household is not appropriate here, as most households had only one child. Our previous implementation of this method in Nepal (Jones and Haslett, 2006) adjusted the three sets of residuals for shrinkage and used these in a nonparametric bootstrap procedure, as described in the next section. Here we use the much simpler parametric bootstrap approach, sampling from normal distributions with variances set to the estimated variance components. There should be little difference in practice as estimation with this many levels tends to encourage approximate normality in the residuals.

4.4 Simulation of predicted values

Simulated values for the model parameters α and β were obtained by parametric bootstrap, i.e. drawn from their respective sampling distributions as estimated by the survey regressions. Simulation of the cluster-and standardized household-level effects c_i and e^*_{ij} presents several possible choices. A parametric bootstrap could be used by fitting suitable distributions (e.g. Normal, t) to the residuals and drawing randomly from these. For simulating log per capita expenditure, and hence poverty incidence, we chose a non-parametric bootstrap in which we sample with replacement from the residuals, i.e. from the empirical distributions. One can either resample the e^*_{ij} from the full set or only from those within the cluster corresponding to the chosen h_i. We chose the latter, which links the household effects estimated via the bootstrap to households in the same cluster, so when mean-correcting the standardized residuals (see Section 2.3) we used

$$\hat{e}^*_{ij} = \hat{e}_{ij} / \hat{\sigma}_{e,ij} - \frac{1}{n_i}\sum_{j=1}^{n_i} \hat{e}_{ij} / \hat{\sigma}_{e,ij}$$

Note that mean correction when needed can be an indication of the extent of any bias in the bootstrap and hence of an incorrect regression model, so it is encouraging that mean corrections here were small in relative terms.

A total of 100 bootstrap predicted values Y^b_{ij} were produced for each unit in the census and for each target variable, as described in Section 2.4. For the three-level models, height-for-age and weight-for-age, this was amended slightly to

$$Y^b_{ijk} = X_{ijk}\beta^b + c^b_i + h^b_{ij} + e^b_{ijk} \ , \quad b = 1, \ldots B$$

with the residuals at each level $c^b_i, h^b_{ij}, e^b_{ijk}$ drawn independently from normal distributions with mean zero and variances equal to the estimated variance components from the regression analysis.

4.5 Production of final estimates

Since a log transform was applied in modelling expenditure, we first reverse this transformation by exponentiating, i.e. predicted expenditure $E^b_{ij} = e^{Y^b_{ij}}$. The predicted values can then be grouped at the appropriate geographic level. Our main target is commune-level small-area estimates, but we have also considered higher levels of aggregation (region, ecological zone, and province) for comparison with the direct survey

estimates. Once the predicted values have been produced and stored it is easy to investigate alternative levels of aggregation, using the standard errors as a guide to what is an appropriate level.

For expenditure the census units are households and the target variables are household average values, so the aggregation needs to be weighted by household size. Thus for example the formula for P_R^b the *b*th bootstrap estimate of poverty incidence (α = 0 in equation 1.1) in region *R* is amended to:

$$P_R^b = \sum_{ij \in R} n_{ij} \cdot I(E_{ij}^b < z) \Big/ \sum_{ij \in R} n_{ij}$$

where n_{ij} is the size of household *ij* in *R*. The census units for height-for-age and weight-for-age are individual children, so no weighting is required. For example the estimated incidence of stunting for region *R* is:

$$S_R^b = \sum_{ij \in R} I(HAZ_{ij}^b < -2.00) / N_R$$

where N_R is the number of eligible children in *R*.

The 100 bootstrap estimates for each region, e.g. $P_R^1 \ldots P_R^{100}$ were summarized by their mean and standard deviation, giving a point estimate and a standard error for each region. For expenditure we have produced commune-level averages in addition to incidence, gap and severity below each of the total poverty line and the food poverty line. For height-for-age and weight-for-age we only give two measures: incidence below two standard deviations and incidence below three standard deviations.

5. Results for Poverty Measures

5.1 Comparison with CSES2009 Estimates

The results for poverty incidence were first accumulated to high levels of aggregation for comparison with the direct estimates available from the CSES2009 survey. Table 5.1 shows both sets of estimates (P0) together with their standard errors (se). These estimates are all based on the preliminary expenditure data and poverty lines, so are for comparison purposes only. The standard errors for the direct survey estimates have been calculated using a robust variance technique which controls for the survey design. The standard errors for the small-area estimates (SAE) are the standard deviations of the 100 bootstrap estimates. We have added a standardized difference between the two sets of estimates, defined as

$$Z = \frac{\text{Small area estimate - direct estimate}}{\sqrt{(\text{small area se})^2 + (\text{direct estimate se})^2}}$$

If both methods are correctly estimating the same quantities, then Z should approximate a standard normal distribution.

Table 5.1 Comparison of estimates of poverty incidence

	CSES		SAE		Standard
	P0	se	P0	se	Difference
Cambodia	0.229	0.008	0.227	0.005	-0.256
Phnom Penh	0.129	0.017	0.116	0.011	-0.631
Other Urban	0.193	0.021	0.212	0.010	0.820
Rural	0.246	0.009	0.242	0.006	-0.418
Phnom Penh	0.128	0.017	0.116	0.011	-0.615
Plain	0.204	0.011	0.196	0.007	-0.612
Tonlesap	0.274	0.016	0.275	0.010	0.097
Plateau/Mountain	0.299	0.025	0.293	0.016	-0.234
Coastal	0.188	0.024	0.213	0.008	0.986

These Z scores suggest that the small-area estimates are all within two standard errors of the direct estimates, indicating a reasonable level of agreement between the two, especially since there are eight tests of significance, so that it could be expected that one Z score would exceed two even if none were really statistically significant.

We note from Table 5.1 that, although in all cases the SAEs are more precise (i.e. smaller standard errors) than the direct estimates, there is little reduction in standard error from the small-area methodology at the largest levels of aggregation. This is because the uncertainty in the direct estimates due to sampling variability is replaced by uncertainty in the estimated model for the SAEs. At the lower levels however the improvement in precision is much more dramatic.

5.2 Poverty at District and Commune levels

Table 5.2 Summary of district-level poverty measures

	Incidence		Gap		Severity	
	P0	se0	P1	se1	P2	se2
Mean	0.2480	0.0218	0.0690	0.0078	0.0249	0.0036
Standard deviation	0.0787	0.0099	0.0268	0.0043	0.0118	0.0023
Minimum	0.0646	0.0085	0.0125	0.0024	0.0036	0.0008
Maximum	0.5095	0.0571	0.1763	0.0264	0.0761	0.0148

Table 5.2 gives a statistical summary of the estimates for the 193 districts. A complete listing of the estimates is given in Appendix C.1. Poverty incidence at district level ranges from 6.5% (Chamkar Mon in Phnom Penh) to 50.9% (Ta Veaeng in Ratanak Kiri), with a standard deviation of 7.87%. The standard errors of these estimates are acceptably small, being in all cases less than 5.8% and with a mean of 2.18% (about 28% of the variability between districts).

The general pattern of the poverty estimates in Cambodia is that poverty is comparatively low in Phnom Penh, slightly higher in the plains, river valleys, around the shores of Tonle Sap and on the sea coast, but rises to higher levels in the northeast and in the internal area away from the shore to the southwest of Tonle Sap.

Table 5.3 gives a statistical summary of the estimates for the 1621 communes. The standard errors of the incidence estimates have a mean of 4.8%, and most standard errors (1050 out of 1621) are below 5%. Figure 5.1 shows that, as expected, the larger standard errors occur in the smaller communes in terms of population size. For the most part, then, these estimates would seem to be useful in making poverty comparisons at commune level.

Table 5.3 Summary of commune-level poverty measures

	Incidence		Gap		Severity	
	P0	se0	P1	se1	P2	se2
Mean	0.2418	0.0482	0.0667	0.0170	0.0238	0.0077
Standard deviation	0.0875	0.0195	0.0291	0.0084	0.0119	0.0044
Minimum	0.0382	0.0135	0.0070	0.0036	0.0019	0.0013
Maximum	0.6188	0.1482	0.2247	0.0643	0.0997	0.0372

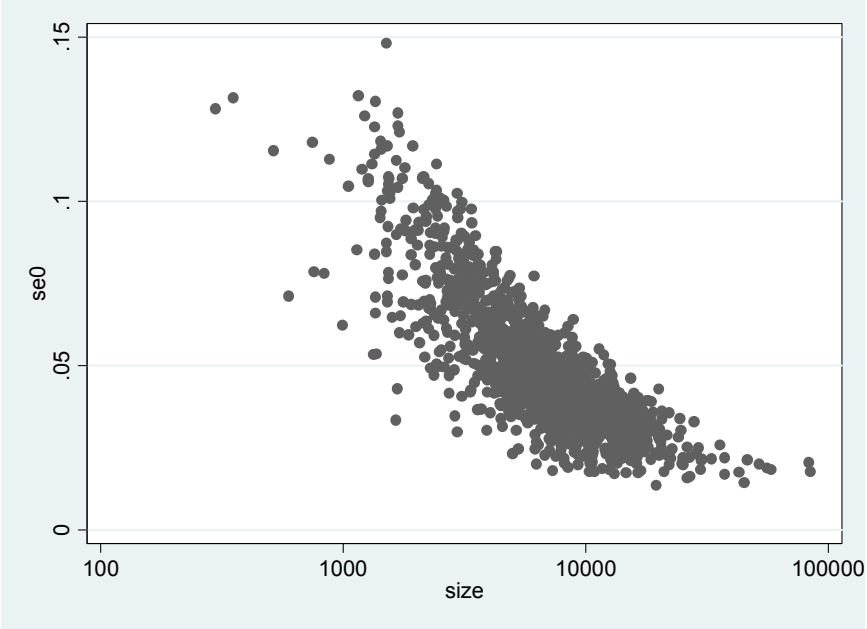


Figure 5.1 Standard error of poverty incidence estimate versus commune size

5.2 Comparison with Commune Database Estimates of Poverty Incidence

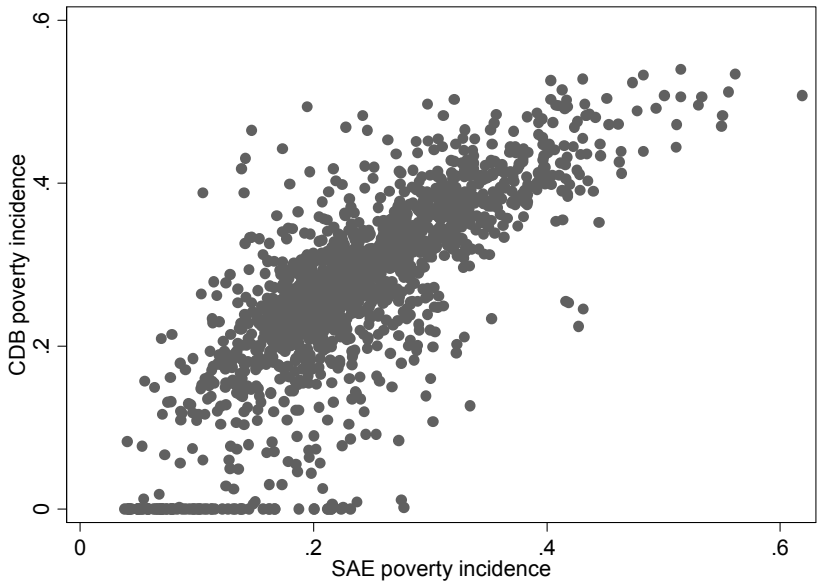


Figure 5.2 CDB versus SAE estimates of poverty incidence at commune level

A comparison at commune level of our provisional small-area estimates with those obtained via the Commune Database shows that, although derived in very different ways using different data and methodologies, they correlate reasonably well (Pearson correlation coefficient 0.782). One noticeable difference is the group of communes estimated to have zero poverty incidence from the CDB. These are almost all within Phnom Penh province; most have been estimated by SAE as having low but non-zero levels of poverty, although a few (particularly in Dangkao and Ruessi Kaev districts) are estimated to have levels of over 20%.

5.4 Poverty Maps

Maps of these poverty estimates, including incidence, gap and severity at commune level, are given in Appendix D.1.

6. Results for Malnutrition Measures

6.1 Results for stunting

Table 6.1 Comparison of estimates of stunting incidence (S2) from CAS2009

	CAS		SAE		Standard
	S2	se	S2	se	Difference
Cambodia	0.372	0.007	0.414	0.006	4.494
Phnom Penh	0.293	0.035	0.316	0.027	0.505
Other Urban	0.296	0.020	0.348	0.015	2.133
Rural	0.389	0.008	0.429	0.006	3.932
Phnom Penh	0.293	0.035	0.316	0.027	0.505
Plain	0.359	0.013	0.410	0.006	3.472
Tonlesap	0.400	0.012	0.426	0.007	1.857
Plateau/Mountain	0.406	0.013	0.446	0.008	2.640
Coastal	0.357	0.018	0.408	0.007	2.652

Key: se=standard error

Table 6.2 Comparison of estimates of stunting incidence (S2 from CDHS)

	CDHS		SAE		Standard
	S2	se	S2	se	Difference
Cambodia	0.399	0.011	0.409	0.007	0.800
Phnom Penh	0.260	0.032	0.309	0.025	1.217
Other Urban	0.313	0.022	0.325	0.012	0.471
Rural	0.419	0.013	0.426	0.008	0.464
Phnom Penh	0.260	0.032	0.309	0.025	1.217
Plain	0.386	0.021	0.406	0.009	0.886
Tonlesap	0.416	0.018	0.417	0.007	0.079
Plateau/Mountain	0.453	0.024	0.448	0.009	-0.220
Coastal	0.426	0.034	0.404	0.010	-0.630

Key: se=standard error

Initially, separate small-area estimates of stunting (S2) were prepared from the CAS2008 and CDHS2010 surveys. Tables .61, 6.2 compare the SAEs with the direct survey estimates from each source. The CDHS-based small-area estimates match reasonably well with the direct estimates, whereas the CAS-based estimates seem to be too high,

particularly in the rural areas. As noted earlier, the first stage regression models for height-for-age were poor in terms of predictive power, with R^2 values of 20%, 17% respectively (see Appendices B.2, B.3). Despite this, it appears from Tables 6.1, 6.2 that the small-area estimates of stunting at high aggregation levels still have smaller standard errors than the direct estimates from the surveys. This is perhaps due to the fact that very little of the residual variation from the model is at PSU-level, so that this unexplained variation, though considerable, is mostly averaged over a large number of individuals.

Although the CAS-based small-area estimates do not match particularly well with the direct estimates, they do match quite well with the DHS-based SAEs. When comparisons are made at province level there is a good degree of correlation between the SAEs and direct estimates for both surveys, with a few anomalous small provinces where the direct estimates are not reliable. We have decided therefore that, despite some reservations about the quality of the CAS-based estimates, that there is still some value in them. We have therefore decided, for our final estimates, to use a weighted average of the CAS- and DHS-based SAEs, the weights being the inverse of the variance of each individual estimate, i.e.

$$S2_{Final} = \frac{w_1 S2_{CAS} + w_2 S2_{CDHS}}{w_1 + w_2}$$

where $w_1 = [\mathrm{se}(S2_{CAS})]^{-2}$ and $w_2 = [\mathrm{se}(S2_{CDHS})]^{-2}$. This gives an optimal weighting of the two estimates assuming independence (which is a reasonable assumption because the two surveys were undertaken separately), resulting in smaller standard errors than either of the individual estimates. A similar approach was used for severe stunting (S3) – the proportion of children at least three standard deviations below average height-for-age.

Turning to the district-level estimates, summarized in Table 6.3, we find that the standard errors are quite small, with an average of only 1.2%. One out of 193 is a little over 3%. The estimates of stunting incidence range from 23% (Prampir Meakkakra) to 54% (Andoung Meas). The standard errors for severe stunting are also quite small, averaging 0.9% in comparison with the standard deviation of 3.6%, so should provide a reasonably accurate comparisons of severe stunting between areas. A complete listing of the estimates is given in Appendix C.2.

Table 6.3 Summary of district-level estimates of stunting incidence (S2, S3)

	Stunting		Severe stunting	
	S2	se2	S3	se3
Mean	0.4207	0.0121	0.1893	0.0088
Standard deviation	0.0544	0.0046	0.0364	0.0036
Minimum	0.2347	0.0065	0.0802	0.0047
Maximum	0.5428	0.0324	0.2823	0.0267

Even at commune level, as shown in Table 6.4, the estimates of both S2 and S3 have reasonably small standard errors in comparison with the variability between the communes. Stunting incidence S2 has an average standard error of 2.3%, and only six out of 1621 have standard errors of over 5%. Estimates at commune level range from 18% (Boeng Keng Kay Muong in Phnom Penh) to 62% (Kaoh Pang in Ratanak Kiri). Standard errors for severe stunting S3 average 1.7%, in comparison with the standard deviation of 4% between the communes. Thus, although the models used to derive the estimates have low predictive power for individual children, they seem to be capturing a considerable amount of variability in malnutrition between communes.

Table 6.4 Summary of commune-level estimates of stunting incidence (S2, S3)

	Stunting		Severe stunting	
	S2	se2	S3	se3
Mean	0.4176	0.0233	0.1860	0.0172
Standard deviation	0.0614	0.0067	0.0404	0.0057
Minimum	0.1847	0.0115	0.0543	0.0081
Maximum	0.6183	0.0637	0.3307	0.0492

Key: se2=standard error of S2
se3=standard error of S3

6.2 Results for underweight

As with stunting, two separate sets of estimates of underweight (U2) were originally prepared, one based on CAS2009 and the other on CDHS2010. Comparison between these and the direct survey-only estimates are presented in Tables 6.5, 6.6. Again the CAS-based estimates seem significantly too high, largely because of over-estimation in the rural areas. The CDHS-based estimates are also a little too high overall, but reproduce quite well the pattern of the direct estimates at ecozone level, with all differences being

less than two standard errors. Both sets of SAEs correlate quite well with their respective direct estimates at province level, except for a few anomalies for smaller provinces. We therefore conclude again that both are capturing regional variation in incidence of underweight, so for our final estimates have use a weighted average of the CAS- and DHS-based SAEs, with inverse-variance weights, i.e.

$$U2_{Final} = \frac{w_1 U2_{CAS} + w_2 U2_{CDHS}}{w_1 + w_2}$$

where $w_1 = [\text{se}(S2_{CAS})]^2$ and $w_2 = [\text{se}(S2_{CDHS})]^2$, with similar approach taken for severe underweight (S3).

Table 6.5 Comparison of estimates of incidence of underweight (U2) from CAS2009

	CAS		SAE		Standard
	U2	se	U2	se	Difference
Cambodia	0.274	0.007	0.308	0.005	4.038
Phnom Penh	0.160	0.022	0.190	0.018	1.072
Other Urban	0.219	0.017	0.234	0.012	0.706
Rural	0.292	0.008	0.325	0.006	3.474
Phnom Penh	0.160	0.022	0.190	0.018	1.072
Plain	0.287	0.013	0.317	0.007	2.049
Tonlesap	0.298	0.010	0.323	0.006	2.053
Plateau/Mountain	0.260	0.013	0.303	0.010	2.672
Coastal	0.262	0.017	0.316	0.007	3.002

Key: se=standard error

Table 6.6 Comparison of estimates of incidence of underweight (U2) from CDHS2010

	CDHS		SAE		Standard
	U2	se	U2	se	Difference
Cambodia	0.284	0.010	0.314	0.007	2.485
Phnom Penh	0.182	0.030	0.174	0.023	-0.197
Other Urban	0.204	0.013	0.228	0.010	1.410
Rural	0.300	0.011	0.334	0.008	2.422
Phnom Penh	0.182	0.030	0.174	0.023	-0.197
Plain	0.282	0.018	0.318	0.008	1.810
Tonlesap	0.289	0.015	0.320	0.008	1.845
Plateau/Mountain	0.332	0.021	0.349	0.010	0.723
Coastal	0.271	0.030	0.314	0.007	1.391

Key: se=standard error

The district-level estimates for underweight, described in Table 6.7, have standard errors similar to those for stunting, having an average of only 1.1%, with the largest just over 3%. The underweight estimates themselves range from 14% (Prampir Meakkakra) to 43% (Siem Pang). The standard errors for severe underweight are also quite small, with a standard error of 0.5% in contrast to the district-level standard deviation of 2%. The estimates of underweight and severe underweight are very strongly correlated (r = 0.991) so would give very similar results if used to discriminate between districts. A complete listing of the estimates is given in Appendix C.3.

Table 6.7 Summary of district-level estimates of underweight incidence (U2, U3)

	Underweight		Severe Underweight	
	U2	se2	U3	se3
Mean	0.3134	0.0112	0.0828	0.0053
Standard deviation	0.0518	0.0045	0.0207	0.0024
Minimum	0.1387	0.0064	0.0198	0.0024
Maximum	0.4308	0.0301	0.1398	0.0157

Key: se2=standard error of U2
se3=standard error of U3

Again at commune level the standard errors for underweight incidence are reasonably small, as shown in Table 6.8. Only 4 out of 1621 are above 5%, with an average of 2.1%. Estimated incidence of underweight ranges from 10% (Tuol Svay Prey Ti Pir in Phnom Penh) to 50% (Anlong Phe in Stung Treng). Thus the models for weight-for-age, although similarly low in predictive power to those of height-for-age, seem to be capturing a considerable amount of variability in incidence of underweight between communes.

Table 6.8 Summary of commune-level estimates of underweight incidence

	Underweight		Severe Underweight	
	U2	se2	U3	se2
Mean	0.3174	0.0207	0.0835	0.0106
Standard deviation	0.0583	0.0062	0.0232	0.0038
Minimum	0.1015	0.0103	0.0124	0.0037
Maximum	0.5035	0.0554	0.1797	0.0425

Key: se2=standard error of U2
se3=standard error of U3

6.4 Malnutrition Maps

Maps of the stunting incidence estimates, including severe stunting are given in Appendix D.2. Maps for underweight and severe underweight are in Appendix D.3.

7. Multivariate Analysis

7.1 Correlations

Table 7.1 Pairwise correlations for all three measures, commune level

	P0	S2	U2
P0	1		
S2	0.7882	1	
U2	0.6921	0.8999	1

Here we investigate the relationships between the three measures of deprivation estimated in this report, namely incidences of: poverty (P0), stunting (S2), and underweight (U2). Table 7.1 shows the correlations between the commune-level estimates of these measures. All show strong positive correlation, which suggests that those areas showing high levels of deprivation on one measure tend also to be high on the other measures.

7.2 Principal Components

Table 7.2 Principal component analysis for poverty, stunting and underweight

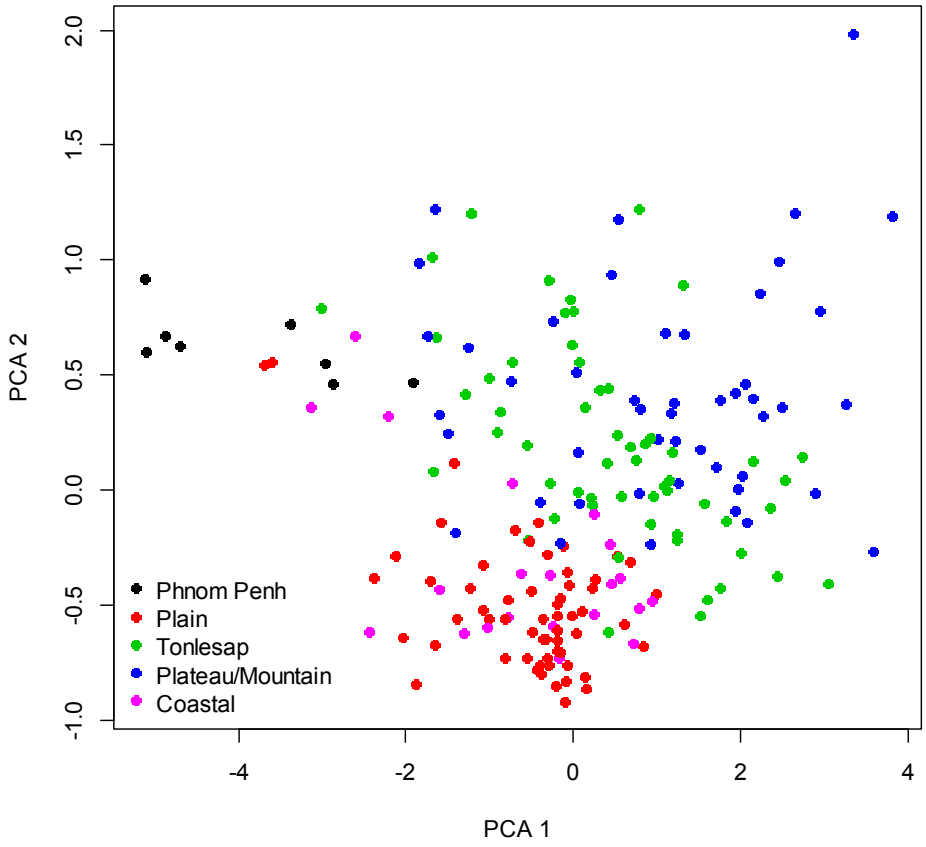
Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	2.5898	2.2649	0.8633	0.8633
Comp2	0.3249	0.2396	0.1083	0.9716
Comp3	0.0853	.	0.0284	1

Variable	Comp1	Comp2	Comp3	Unexplained
P0	0.5503	0.8077	0.2118	0
S2	0.6009	-0.2069	-0.7720	0
U2	0.5797	-0.5522	0.5992	0

Here we consider the possibility of combining all three indicators into a single measure using principal components analysis. The results at commune level, presented in Table 7.2, show that the first principal component, representing the combination giving maximum overall variation, gives approximately equal weighting to all three measures,

and so is close to a simple average of poverty, stunting and underweight incidences. This first component represents 86% of the overall variation in the three indicators, and could be useful if a single measure of overall deprivation is required. The second component contrasts poverty with underweight. These two components together account for 97% of the variation.

Figure 7.1 Plot of first two principal components for districts



The first principal component has positive coeffiocients for P0, S2 and U2, so that it provides a general measure of poverty. The second principal component contrasts P0 with U2 (and to a lesser extent with S2), so that it separates districts that have relatively higher poverty levels but lower underweight from those that have lower poverty but higher underweight. Interestingly, when the values for each district are then plotted, as in Figure 7.1, the districts within ecozones tend to clump together. This indicates that districts within ecozones have similar (though not identical patterns) of poverty and underweight. For example, Phnom Penh tends to have low poverty but mid-range

underweight, the Plain has mid-range poverty but low relative underweight, and the Plateau/Mountain has relatively high values of both poverty and underweight.

The effect of contrasting poverty and underweight, while important, is not however marked. After all, the second principal component explains only 10.8% of the total variance, while the first principal component explains 86.3%, so the overriding consideration for Cambodia as a whole is the positive link between poverty and underweight.

8. Conclusions and Discussion

We have produced small-area estimates of poverty and malnutrition in Cambodia at district and sub-district levels by combining survey data with auxiliary data derived from the 2008 census. A single model was found to be adequate for predicting log average per capita household consumption expenditure and the poverty measures derived from it. The commune-level estimates obtained have acceptably low standard errors.

It is interesting to note that the estimates derived from height-for-age, weight-for-age (but not weight-for-height or calorie intake) also had acceptably small standard errors down to commune level, even though our predictive models for these variables had lower R^2 values than for log average per capita household consumption expenditure. The lower R^2 values for these regression models in part reflect the additional level of variation (children within households) in comparison with the model for log average per capita household consumption expenditure. Smaller R^2 is also more acceptable if the large unexplained variation is truly random across households or individuals, with little or no cluster-level variation. Since the methodology incorporates in the standard errors any remaining cluster-level variation, this would appear to be the case. It is nevertheless likely that some of this variation represents missing variables in the model which would give better prediction if they were available. If important factors are missing then the small-area estimates obtained will not reflect the true variability in these malnutrition indicators, and will tend to have larger standard errors than would otherwise be the case. Calorie intake is inevitably imprecisely measured, so a large part of its unexplained variation (and the main reason it could not be modelled adequately) is due to measurement error. However, this argument does not apply to height-for-age, weight-for-age, or weight-for-height which are measured quite accurately. This suggests that the problems modelling weight-for-height reflect comparatively small differences between parts of Cambodia for this measure. There are other factors, particularly health-related ones, that would be useful predictors of malnutrition, but these variables were not available for the population from the census data and so could not be included in the small-area models.

Geographic Information System (GIS) variables were not used directly in the regression and heteroscedasticity models. GIS variables are necessarily at aggregate level and, as for census means, because they are aggregated are not able to provide household level

information. Like all regressor variables, they are to be included in models only where they explain variation *in addition* to that explained by the other variables in the model. GIS and other variables, even when they are not included directly in the model, can nevertheless be important in their own right. As a consequence, although maps of small-area estimates of the various poverty measures are important, so are various complementary maps of GIS and other variables. What is important is whether such variables have high correlation with the small-area estimates (even if they are not in the regression model itself).

As noted earlier, we have departed from previous implementations of ELL methodology in a few important ways. More detailed discussion can be found in Haslett and Jones (2005b, 2010). For example, the strategy for choosing appropriate regression models for the target variable is not usually made explicit, but Miller (2002) sounds a number of cautions. Using separate survey based models for subgroups such as geographical strata, especially where there are a large number of such subgroups, and selecting variables have from a very large pool of possibilities including all interaction terms cannot be recommended. Model-fitting criteria such as adjusted R^2 or *AIC* will penalize for fitting too many variables, but do not account for the number of variables that are being selected from. Cross-validation (i.e. dividing the sample, fitting a model to one part, and testing its utility on the other) might be useful here. We have tried where possible to fit a single model for the whole population, including interaction terms only when the corresponding main effects are also included and looking carefully at the interpretability of the estimated effects, i.e. whether the model makes sense. This is a time-consuming procedure but can lead to more stable parameter estimation and more reliable prediction. This does not preclude fitting subgroup or area effects in models when required, or combining area based models into an essentially equivalent single model containing appropriate interactions to improve stability of regression parameter estimates. When the effects of most factors on the target variable are similar in all areas, with modulation only between rural and non-rural areas, an urban/rural possibly with some interactions with other variables will suffice. Even a single model can produce very different area based estimates when appropriate as the results in Appendix C attest. Furthermore if there is prior knowledge on which factors are likely to affect the target variable, this can be incorporated into the model selection. A more formal way of doing this would be through a Bayesian analysis, but this is beyond the scope of the present work.

The use of specialised survey regression routines, such as those available in Stata, Sudaan and WesVar, in the initial model fitting to the survey data has distinct advantages, since it incorporates the entire survey design and gives a consistent estimate of the covariance matrix. These specialized routines use a robust estimation methodology, essentially collapsing the covariance matrix within clusters, and such methods are consequently more stable than ones which estimate a covariance within each cluster. A perceived disadvantage is that such robust methods may give poor estimates if used for small subpopulations with few clusters. However this is a real effect, not an artefact of the fitting procedure. Note that such routines require *all* survey data to be included in any analysis (even of a subpopulation) in order to give unbiased standard errors, so that analysis of sub-setted survey data is not recommended, even if different models are being fitted to different subgroups. The weighting of the survey observations is complex not only because of the survey design but also because the target variable is often a per capita average. Alternatively, if individual data are used, these will be correlated when from the same family, although the robust variance estimate is still valid even there because it only assumes independence between clusters, not of observations within clusters.

To allow for non-independence between children in the same household at the prediction stage, we have extended the ELL approach to incorporate three levels of variation. Whilst the estimation of variance components in such a hierarchical model is now well-understood, the use of estimated random effects in a non-parametric bootstrap raises some theoretical issues, such as adjustment for degrees of freedom, which might provide fruitful areas for further research. We have also tested, to the extent possible given many sampled communes contain only one sampled primary sampling unit (PSU), that small area (i.e. commune) level random effects are negligible when estimating standard errors.

The benefits of the ELL methodology accrue when interest is in several nonlinear functions of the same target variable, as in the case here of six poverty measures defined on household per capita expenditure, or in distributional properties. If only a single measure were of interest it might be worthwhile to consider direct modelling of this. For example small-area estimates of poverty incidence could be derived by estimating a logistic regression model for incidence in the survey data. Ghosh and Rao (1994) consider this situation within the framework of generalized linear models. If on the other hand there are several target variables which might be expected to be highly correlated, it might increase efficiency to use a multivariate model rather than separate univariate regressions. The discussion in Section 7.3 and 8.2 does however raise some interesting

issues about the utility of such multivariate models, since such techniques tend to shrink estimates of each component toward one another, and it is sometimes the contrast rather than the combination of variables such as height-for age, weight-for-age, and weight-for-height that is important.

From a theoretical perspective, the best (i.e. most efficient) small-area estimator uses the actual observed Y when these values are known, i.e. for the units sampled in the survey, and the predicted Y values otherwise. The resulting estimator can be thought of as a weighted mean of the direct estimator, from the survey only, and an indirect estimator derived from the auxiliary data, the weights being related to the standard errors of the two estimates. In practice it may be impossible for confidentiality reasons to identify individual households in the survey and match them to the census, but there is a theoretical basis for using a weighted mean of the two estimates and thereby increasing precision. Further it is not necessary to resample unconditionally from the empirical distribution of the cluster-level residuals for those clusters which are present in the survey. An alternative would be to resample each of these parametrically from an estimated conditional distribution, i.e. for clusters present in the survey we would calculate the bootstrap predictions using the known value rather than a draw from a random distribution. This would however not have a major effect where the number of clusters in the sample is small relative to the number of clusters defined over the whole population. See also Valliant, Dorfman and Royall (2000).

The provision of standard errors with the small-area estimates is important because it gives the user an indication of how much accuracy is being claimed, conditional on the model being correct. Ultimately decisions are to be made on which areas should receive the most development assistance, so it is important that this information be given to users in a way that is most useful for this purpose. It is not clear exactly how the standard error information should be incorporated, but this is at least in part because the answer will depend on the parameters of the decision problem. We have explored a possible way of incorporating the standard errors into a poverty map, first calculating standardized departures from a pre-specified incidence level, say 40%, as

$$Z = \frac{\text{estimate} - 0.40}{\text{standard error}}$$

and then transforming this into a probability assuming a normal distribution. This value can then be mapped and interpreted as the probability that the corresponding area has a

poverty incidence at least as high as the pre-chosen level. Thus when targeting assistance we could focus on those areas which we believe have the greatest chance of exceeding a threshold poverty incidence, although as with any single map some caution is required if the population sizes in the areas differ markedly. The probabilities here are calculated on the assumption that the sampling distributions of the small-area estimates of incidence are approximately normal. A nonparametric alternative would be to take the proportion of bootstrap estimates above the cut-off value.

From a technical perspective, the statistical methods used would benefit from further theoretical development and justification. The range of models possible using small-area estimation is very broad, and while the ELL methodology has a number of theoretical and practical advantages, sensitivity of estimates to different small-area estimation models remains an only partially explored issue. This question relates both to the choice of the ELL method, *vis-à-vis* others, and to the choice of explanatory variables within models (e.g. submodels for different areas, cross-validation of variables selected from a large pool including higher level interactions, consistency of sign and magnitude of parameter estimates with likely influence on poverty in the presence of correlated variables). These questions need theoretical work and extend beyond the present study.

Ground truthing or validation of small-area estimates by visits to selected small areas after models have been fitted and small-area estimates derived from them can be a useful exercise. Some cautions are however warranted. The first is that small-area estimation is a technique that works best in aggregate - not every small-area estimate can be expected to give precise information, so that choosing areas to visit on the basis of possible anomalies can give a biased picture of the utility of the estimates as a whole. It is also difficult to ask participants in a validation exercise to differentiate various types of poverty or not to include aspects (such as health or water quality) which because they are not included in the census variables cannot be part of the small-area estimates themselves. Validation exercises are also usually limited by funds, so that formal testing of the accuracy of the small-area estimates is not possible by this method. Nevertheless, validation can provide useful qualitative insights and even more importantly a forum for discussion of results of poverty mapping with local communities.

Small-area models are not perfect, and standard errors derived from them depend on the model being at least approximately correct, or at least correct enough to make sound predictions. Despite these caveats, from a practical point of view the Cambodia small-

area poverty and malnutrition estimates presented in this report are at a much finer geographical level than has previously been possible and consequently should be of considerable benefit when a mechanism for allocation of development assistance is required.

Bibliography

Alderman H., Babita M., Demombynes G., Makhata N. and Ozler B., 2002. "How low can you go? Combining census and survey data for mapping poverty in South Africa", *Journal of African Economics*, **11**, 169-200.

Cambodian-German Cooperation,. 2006. *Identification of Poor Households Project (IDPoor): Project Outline*, April 2006

Chambers R.L and Skinner C.J. (eds.), 2003. *Analysis of Survey Data*. Wiley.

Conway, T. and Goh, C., 2006. *Cambodia: Halving poverty by 2015? Poverty Assessment 2006*, World Bank, xviii+198pages+appended tables, February 2006.

Council for Social Development, Cambodia, 2002. *National Poverty Reduction Strategy 2003-2005*, Council for Social Development, Royal Cambodian Government, xiv+238 pages, December 2002.

Dalén, J., 2005a. *Measurement problems in poverty estimation in Cambodia*, Statistics Sweden International Consulting Office, Camstat Report 2005:12.

Dalén, J., 2005b. *Some comments on price index estimates in the draft Cambodia 2004 poverty report by James C Knowles*, Statistics Sweden International Consulting Office, Camstat Report 2005:13.

Dalén, J., 2006. *Sampling issues in a continuous Cambodia Socio-economic Survey*, Statistics Sweden International Consulting Office, Camstat Report 2006:08.

Davidsson, G. and Pettersson, H., 2003. *Survey design for the Cambodian Socio-economic Survey 2003-4*, Statistics Sweden International Consulting Office, Camstat Report 2003:03.

Efron B. and Tibshirani R.J., 1993. *An Introduction to the Bootstrap*. Chapman and Hall.

Elbers C., Lanjouw J.O. and Lanjouw P., 2001. *Welfare in villages and towns: micro-level estimation of poverty and inequality*, unpublished manuscript, World Bank.

Elbers C., Lanjouw J.O., Lanjouw P. and Leite P.G., 2001. *Poverty and inequality in Brazil: new estimates from combined PPV-PNAD data*, unpublished manuscript, World Bank.

Elbers C., Lanjouw J. and Lanjouw P., 2003. "Micro-level estimation of poverty and inequality", *Econometrica*, **71**, 355-364.

Food and Agriculture Organization/ World Health Organization/ United Nations University, 1985. *Energy and Protein Requirements*, WHO Technical Report Series 724. Geneva, WHO.

Foster J., Greer J. and Thorbeck E., 1984. "A class of decomposable poverty measures", *Econometrica*, **52**, 761-766.

Fujii T., 2004. "Commune-level estimation of poverty measures and its application in Cambodia", In *Spatial Disparities in Human Development: Perspectives from Asia*, edited by Kanbur R., Venables A.J., & G. Wan. United Nations University Press.

Fujii, T. ,2002. *Estimation of Poverty Rates at Commune-Level in Cambodia – using Small-Area Estimation Technique to Obtain Reliable Estimates*, Ministry of Planning and the United Nations World Food Programme, October 2002, vi+62 pages+CD of estimates.

Fujii T., Lanjouw P., Alayou S. and Montana L., 2004. *Micro-level estimation of prevalence of child malnutrition in Cambodia*,. World Bank.

Food and Agriculture Organization / World Health Organization / United Nations University, 1985. *Energy and Protein Requirements*, WHO Technical Report Series 724. Geneva, WHO.

German Technical Co-operation, 2006a. *Identification of Poor Households Project (IDPoor)*, Cambodian – German Co-operation, German Technical Co-operation (GTZ).

German Technical Co-operation, 2006b. *Procedures for identification of Poor Households in Cambodia, Version 1 for 2007 Rounds 1a and 1b*, German Technical Co-operation (GTZ) / Cambodia Ministry of Planning.

Ghosh M. and Rao J.N.K., 1994. "Small area estimation: an appraisal", *Statistical Science*, **9**, 55-93.

Hamill P.V.V., Dridz T.A., Johnson C.Z., Reed R.B. et al., 1979. "Physical growth: National Center for Health Statistics percentile". *American Journal of Clinical Nutrition*, **32**, 607-621.

Haslett, S., 2008 *Potential for Small-area estimation and Poverty Mapping at District and Commune Level in Cambodia: Interim Feasibility Report*, Massey University report prepared for the World Food Programme, Cambodia, May 2008.

Haslett, S. Isidro, M. and Jones, G. 2010. "Comparison of survey regression techniques in the context of small-area estimation of poverty", *Survey Methodology*, **36**, 2, 157-170.

Haslett, S. and Jones, G., 2005a. *Estimation of Local Poverty in the Philippines*, Philippines National Statistics Co-ordination Board / World Bank, November 2005.

Haslett, S. and Jones, G., 2005b. "Small-area estimation using surveys and censuses: some practical and statistical issues", *Statistics in Transition*, **7** (3), 541-555

Haslett. S. and Jones, G., 2008. *Potential for Small-area estimation and Poverty Mapping at District and Commune Level in Cambodia*, report to UN World Food Programme, December 2008.

Haslett, S. and Jones, G., 2010. "Small-area estimation of poverty: the aid industry standard and its alternatives", *Australian and New Zealand Journal of Statistics*, **52**(4), 341-362.

Haslett, S., Jones, G. and Isidro, M., 2010. *Potential for Small-area estimation of Malnutrition at District and Commune Level in Cambodia: Feasibility Report Phases 1 and 2*, Report for the UN World Food Programme.

Healy A.J., Jitsuchon S. and Vajaragupta, Y., 2003. *Spatially disaggregated estimation of poverty and inequality in Thailand*, World Bank.

Hou, T. E,, Ny, B. and Karim, R., 2010. *Poverty and Select CMDGs Maps and Charts 2003-2009: Preliminary Results of CDB-based Research and Analysis Project*, 2010 updated edition, Kingdom of Cambodia Ministry of Planning & UNDP.

Isaksson, A., 2007. *Inference from the Cambodian Socio-economic survey*, Statistics Sweden International Consulting Office, Camstat Report 2007:05.

Isidro. M., 2010. "Intercensal updating of small-area estimates", Thesis in partial fulfilment of PhD degree requirements, Massey University, New Zealand.

Isidro, M., Haslett, S. and Jones, G., 2010a. "Comparison of intercensal updating techniques for local level poverty statistics", *Proceedings of Statistics Canada Symposium 2010*, 10B-4, http://www.statcan.gc.ca/conferences/symposium2010/abs-res-eng.htm

Isidro, M, Haslett, S. and Jones, G., 2010b. "Extended Structure-Preserving Estimation Method for Updating Small-Area Estimates of Poverty", *Joint Statistical Meetings, Proceedings of the American Statistical Association 2010*, Vancouver, Canada, http://www.amstat.org/meetings/jsm/2010/onlineprogram

Johansson, S., 2005. *Recall and diary estimates of food consumption. Understanding the differences in CSES 2004*, Statistics Sweden International Consulting Office, Camstat Report 2005:14.

Johansson, S., 2006. *Designing the continuous Cambodia Socio-economic survey (1-3)*, Statistics Sweden International Consulting Office, Camstat Report 2006:02.

Jones, G. and Haslett, S., 2003. *Local Estimation of Poverty and Malnutrition in Bangladesh*. Bangladesh Bureau of Statistics and UN World Food Programme.

Jones G., and Haslett, S., 2006. *Small-area estimation of poverty, caloric intake and malnutrition in Nepal*, Published: Nepal Central Bureau of Statistics / UN World Food Programme / World Bank, September 2006, 184pp, ISBN 999337018-5.

Knowles, J C., 2006. *A new set of poverty indicators for Cambodia, 1993/4 to 2004*, Report to the EAS Country Units of the World Bank, Washington DC. Original report 2005; revised version 28 November 2006.

Laird, N.M. and Ware, J.H., 1982. "Random-Effects Models for Longitudinal Data," *Biometrics*, **38**, 963 - 974.

Lehtonen, R. and Pakhinen, E., 2004. *Practical Methods for Design and Analysis of Complex Sample Surveys*, 2nd Edition, Wiley.

Longford, N. T., 2005. *Missing Data and Small-area estimation*, Springer Verlag.

Miller, A., 2002. *Subset Selection in Regression*, 2nd edition, Chapman and Hall / CRC

Ministry of Planning , Cambodia, 2005. *National Strategic Development Plan, 2006-2010*, Royal Government of Cambodia, Ministry of Planning, xviii+121pages+appendices, December 2005

National Institute of Statistics, Cambodia, 2004. *A Poverty Mapping Profile of Cambodia, 2004*, National Institute of Statistics, Ministry of Planning, xiv+122pages, February 2006.

National Institute of Statistics, Cambodia, 2007. *Statistical Master Plan for Cambodia*, National Institute of Statistics, Ministry of Planning, v+40pages, September 2007.

National Institute of Statistics, Cambodia / FAO, 2007. *Summary Report on the Food Insecurity Assessment in Cambodia: 2003/04 Cambodia Socio-economic Survey*, National Institute of Statistics, Ministry of Planning; EC-FAO Food Security "Information for Action Programme", September 2007.

National Institute of Statistics, Ministry of Planning and UNICEF Cambodia, 2009. *Cambodia anthropometrics survey 2008*. Phnom Penh, Cambodia: National Institute of Statistics, Ministry of Planning and UNICEF Cambodia.

National Institute of Statistics (Cambodia), Ministry of Health (Cambodia) and ICF Macro, 2011. *Cambodia Demographic and Health Survey 2010-2011*. ICF Macro, Calverton, United States.

Neupert, R., 2005. *New demographic estimates and updated projections for Cambodia*, Statistics Sweden International Consulting Office, Camstat Report 2005:04.

Pettersson, H., 2004. *Calibration of sampling weights for Cambodian Socio-economic surveys*, Statistics Sweden International Consulting Office, Camstat Report 2004:03.

Pettersson, H., 2005. Calibration *of sampling weights for Cambodian Socio-economic surveys (2)*, Statistics Sweden International Consulting Office, Camstat Report 2005:01.

Pinney, A., 2007. *Initial calculation of small-area estimates of poverty using CSES 2003/04 and Commune Database 2002-2006*, draft report: UNDP / National Institute of Statistics. July 2008.

Rao J.N.K., 1999. "Some recent advances in model-based small-area estimation", *Survey Methodology*, **23**, 175-186.

Rao, J.N.K., 2003. *Small-area estimation*. Wiley.

Robinson, G.K., 1991. "That BLUP is a Good Thing: The Estimation of Random Effects," *Statistical Science*, **6**, 15 - 51.

Sandqvist, A., 2007a. *Development of expenditure statistics in Cambodia*, Statistics Sweden International Consulting Office, Camstat Report 2007:03.

Sandqvist, A., 2007b. *Development of expenditure statistics in Cambodia, II*, Statistics Sweden International Consulting Office, Camstat Report 2007:09.

Skinner C.J., Holt D. and Smith T.M.F. (eds), 1989. *Analysis of Complex Survey Data*. Wiley.

Swindale A. and Ohri-Vachaspati, P., 2004. *Measuring Household Food Consumption: A Technical Guide. Washington DC: Food and Nutrition Technical Assistance (FANTA) Project*, Academy for Educational Development (AED).

UNDP, 2011. *Human development report, 2011*, United Nations Development Programme, New York.

UNICEF-Cambodia, 2008. *Cambodia Anthropometric Survey: English Supplement*, UNICEF, Cambodia.

Valliant, R., Dorfman, A. H., and Royall, R. M., 2000. *Finite Population Sampling and Inference: A Prediction Approach*, John Wiley, New York.

WHO Multicentre Growth Reference Study Group, 2006. *WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development*. Geneva: World Health Organization; pp 312. : http://www.who.int/childgrowth/publications/en/

World Bank / IMF, 1999. *Poverty Reduction Strategy papers – Status and Next Steps* (SM/99/267, Rev. 1, (11/19/99), http://www.imf.org/external/np/pdr/prsp/status.htm

World Bank and Cambodia Development Forum 2007. *Cambodia - Sharing Growth: Equity and Development Report 2007*, World Bank and Cambodia Development Forum, xv+201pages+appended tables, June 2007.

World Food Programme, 2007. *Integrated Food Security and Humanitarian Phase Classification (IPC) Pilot in Cambodia***,** Final Report, World Food Programme, 57 pages, April 2007.

You, Y., Rao, J. N. K. and Kovacevic, M., 2003. "Estimating fixed effects and variance components in a random intercept model using survey data", *Statistics Canada International Symposium Series - Proceedings*.

Zhao, Q., 2006. *User Manual for PovMap*, World Bank. http://siteresources.worldbank.org/INTPGI/Resources/342674-1092157888460/Zhao_ManualPovMap.pdf

Appendices

Appendix A. Auxiliary variables

A.1 Obtainable or derivable from CSES2009.

variable name	variable label
region	phnom penh, o. urban & o. rural
rural	house in rural area
hhsize	household size
numroom	number of rooms
rfree	dwelling is rent free
electric	main source of lighting is electricity
charcoal	charcoal for cooking
lpge	lpg/electric for cooking
wall	wall material
roof	roof material
floor	floor material
notoilet	no toilet within premises
radio	number of radios owned
tv	number of tvs owned
phone	number of phones owned
cellphone	number of cellphones owned
computer	number of computers owned
bicycle	number of bicycles owned
motorbike	number of motorbikes owned
car	number of cars owned
boat	number of boats owned
tractor	number of tractors owned
koyaon	number of koyaons owned
bikeonly	household has bicycle(s) but no car/motorbikes
h_school	head of household attended school
h_senior	head of household over 65 years
h_divsep	head of household divorced/separated
h_lit	head of household literate
h_seced	head of household has some secondary education
h_subag	head of household engaged in subsistence agriculture
pkids06	prop of hh aged 0-6
pkids714	prop of hh aged 7-14
psenior	prop of hh aged 65+
plit	prop of hh literate
pseced	prop of hh with secondary education
pemp	prop of hh employed

A.2 Obtainable or derivable from CAS2008.

variable name	variable label
nch	number of eligible children (0-4 years) in household
tmom	number of eligible mothers in household
age	age in completed years
female	female
sub_ag	at least one adult involved in subsistence agriculture
educ_primary	proportion of eligible mothers with highest educational attainment primary
educ_low_second	proportion of eligible mothers with highest educational attainment low_second
educ_uper_sec~d	proportion of eligible mothers with highest educational attainment upper_second
educ_higher	proportion of eligible mothers with higher education
dsw_spouse	proportion of eligible mothers divorced, widowed or separated
school	proportion of eligible mothers attended school
electric	household has electricity
radio	household has radio
tv	household has tv
cellphone	household has cellphone
bicycle	household has bicycle
motorbike	household has motorbike
car	household has car
boat	household has boat
bikeonly	household has bicycle(s) but no car/motorbikes
roof_natural	household has natural roof
roof_finished	household has finished roof
roof_rudim	household has rudimentary roof

A.3 Obtainable or derivable from CDHS2010.

variable name	variable label
rel_head	Relationship to head
age	Age of household members
female	Female
nch	Number of eligible children (0-4 years)
hhsize	Number of household members
electric	Household has electricity
radio	Household has radio
tv	Household has television
bicycle	Household has bicycle/cyclo
motorbike	Household has motorcycle/scooter
cell	Household has a mobile telephone
boat	Household has a boat
h_fem	female head of household
pkids06	prop of hh aged 0-6
pkids714	prop of hh aged 7-14
psenior	prop of hh aged 65+
pwamale	prop of hh males 15-64
pseced	prop of hh with secondary education
ck_fwood	Cook with firewood
ck_charc	Cook with charcoal
notoilet	No toilet
dr_public	Drinking water from pipe of public tap
dr_tubed	Drinking water from pipe of tube well or borehole
h_noschool	Head of household never attended school
h_seced	Head of household attended secondary school
floor	Main floor material
wall	Main wall material
roof	Main roof material
h_age	age of head of household
h_marital	marital status of head of household

A.4 Census means (enumeration-area level) from Census2008

variable name	variable label
rural_e	mean rurality
radio_e	mean number of radios owned
phone_e	mean number of phones owned
cellphone_e	mean number of cellphones owned
computer_e	mean number of computers owned
bicycle_e	mean number of bicycles owned
motorbike_e	mean number of motorbikes owned
car_e	mean number of cars owned
boat_e	mean number of boats owned
koyaon_e	mean number of koyaons owned
nethome_e	propn with internet at home
netout_e	propn accessing internet outside home
deaths_e	death rate
rfree_e	propn rent free
elec_e	propn with electricity
charc_e	propn using charcoal for cooking
lpge_e	pron using lpg/electric for cooking
notoilet_e	propn with no toilet within premises
pipew_e	propn piped/tubed water
withinw_e	propn with water within premises
onerm_e	propn with only one room
bikeonly_e	propn with bicycle(s) but no car/motorbikes
nhh_e	number of households
wallrud_e	propn with rudimentary wall
roofrud_e	propn with rudimentary roof
floorrud_e	propn with rudimentary floor
resplus_e	propn residential+shop/business
mtnonkh_e	propn mother-tongue non-khmer
regnonb_e	propn religion non-buddhist
bovill_e	propn born outside current village
hmfovill_e	propn hhead moved from outside current village
hmov2_e	propn hhead less 2yrs in current village
hmovins_e	propn hhead moved for insecurity/calamity/unemployment
hlitfe_e	propn hhead literate in French or English
disabled_e	propn disabled
hunemp_e	propn hhead unemployed
hemp06_e	propn hhead employed 6 months or less in past year
subag_e	propn hhead in subsistence agriculture/hunting
hownacc_e	propn hhead own-account workers
wkabroad_e	propn employed working abroad
h_school_e	propn head of household attended school
h_senior_e	propn head of household over 65 years
h_divsep_e	propn head of household divorced/separated
h_lit_e	propn head of household literate
h_seced_e	propn head of household has some secondary education
h_subag_e	propn head of household engaged in subsistence agriculture
pkids06_e	prop of ea aged 0-6
pkids714_e	prop of ea aged 7-14
psenior_e	prop of ea aged 65+
plit_e	prop of ea literate
pseced_e	prop of ea with secondary education
pemp_e	prop of ea employed

A.5 Village (_v) and commune (_c) variables from CDB2009

variable name	variable label
totfam_v	number of families in village
totpop_v	population of village
kmpsch_v	distance (km) to primary school
kmjsch_v	distance (km) to junior sec. school
kmssch_v	distance (km) to senior sec. school
kmmkt_v	distance (km) to nearest market
kmroad_v	distance (km) to year-round road
hrsroad_v	time to year-round road by motor vehicle
hfem_v	propn hh with female head
hfu5_v	% hh headed by mother w/ u5s
rland1_v	% hh w/ less than 1ha rice land
fland1_v	% hh w/ less than 1ha farm land
dispoll_v	% hh suffering disater/pollution
toilet_v	% hh with toilet
publand_v	% hh living on public land
ucjob_v	% aged 18-60 with uncertain/irregular jobs
fucjob_v	% females aged 18-60 with uncertain/irregular jobs
immigrant_v	% immigrants
emigrant_v	% emigrants
hmless_v	% homeless
totpop_c	population of commune
totfam_c	number of families in commune
psch100_c	primary schools per 100 people
pcrm100_c	primary school classrooms per 100 people
pcls100_c	primary school classes per 100 people
lscls100_c	lower secondary school classes per 100 people
uscrm100_c	upper secondary school classrooms per 100 people
ptch100_c	primary school teachers per 100 people
ricepc_c	per capita rice production
ricepha_c	rice production per hectare
mkt1000_c	markets per 1000 people
clin1000_c	health clinics per 1000 people
rural_c	rurality of commune
flood_c	prone to flooding
drought_c	prone to drought

Appendix B. Regression results

B.1 Model for loge(expenditure) in CSES2009

n	p	R^2	σ^2_u	σ^2_h/σ^2_u
11949	32	0.4934	0.173	0.245

where n = sample size, p = number of variables, R^2 = coefficient of determination
σ_u^2 = residual variance, σ_h^2/σ_u^2 = ratio of cluster to total residual variation

Variable	Coef.	Std. Err.	t	P>t	Label
pkids06	-0.1092	0.0288	-3.79	0.000	prop of hh aged 0-6
plit	0.1036	0.0215	4.82	0.000	prop of hh literate
pseced	0.0961	0.0228	4.22	0.000	prop of hh with secondary education
car	0.1928	0.0262	7.36	0.000	number of cars owned
computer	0.0698	0.0282	2.47	0.014	number of computers owned
electric	0.0513	0.0246	2.08	0.038	main source of lighting is electricity
phone	0.1620	0.0636	2.55	0.011	number of phones owned
tv	0.0516	0.0105	4.90	0.000	number of tvs owned
floor_t	0.0761	0.0295	2.58	0.010	floor of tiles
floor_c	0.0224	0.0072	3.14	0.002	floor of cement,parquet
roof_t	0.0617	0.0188	3.29	0.001	roof of tiles
roof_m	0.0369	0.0152	2.43	0.015	roof of metal
wall_b	-0.0519	0.0134	-3.86	0.000	walls of bamboo/mixed type
boat_e	0.1439	0.0426	3.38	0.001	mean number of boats owned
cellphone_e	0.1067	0.0292	3.65	0.000	mean number of cellphones owned
h_lit_e	0.2889	0.1236	2.34	0.020	propn hhead literate
plit_e	-0.4268	0.1445	-2.95	0.003	prop of ea literate
resplus_e	0.3287	0.0977	3.37	0.001	propn hh residential+shop/business
reg3	0.1898	0.0411	4.62	0.000	rural (outside Phnom Penh)
tonlesap	-0.0570	0.0191	-2.98	0.003	Tonlesap ecological zone
plnmount	-0.0722	0.0260	-2.77	0.006	Plains/Mountains ecological zone
hhsizeXs23	-0.0441	0.0088	-4.99	0.000	interaction of household size and regions excluding Phnom Penh
lnhhszXs23	-0.2881	0.0430	-6.69	0.000	interaction of log household size and regions excluding Phnom Penh
notolietXs23	-0.0321	0.0154	-2.08	0.038	interaction of no toilet within premises and regions excluding Phnom Penh
cellphone~23	0.1389	0.0090	15.47	0.000	interaction of cellphone and regions excluding Phnom Penh
rfreeXs23	-0.0709	0.0293	-2.42	0.016	interaction of rfree and regions excluding Phnom Penh
floor_sXs23	0.3363	0.1032	3.26	0.001	interaction of floor made out of stone and regions excluding Phnom Penh

roof_cXS3	0.2328	0.0432	5.38	0.000	interaction of roof made out of concrete, other and regions excluding Phnom Penh
numrooomXS2	0.1412	0.0157	9.00	0.000	interaction of number of rooms and rural
numroomXS3	0.0561	0.0114	4.94	0.000	interaction of number of rooms and rural
motorbikeXS2	0.0920	0.0180	5.11	0.000	interaction of motorbike and urban
motorbikeXS3	0.1067	0.0102	10.46	0.000	interaction of motorbike and rural
_cons	8.7825	0.0666	131.96	0.000	constant term

B.2 Model for height-for-age in CAS2008

n	p	R^2	σ_c^2	σ_h^2	σ_e^2
7965	26	0.203	0.0691	0.3937	1.3937

where n = sample size, p = number of variables, R^2 = coefficient of determination
σ_c^2 = cluster-level variance, σ_h^2 = household-level variance, σ_e^2 = residual variance

Variable	Coef.	Std. Err.	t	P>t	Label
age_1	-0.9881	0.0590	-16.74	0.000	aged 1 year
age_2	-1.3365	0.0556	-24.03	0.000	aged 2 years
age_3	-1.4311	0.0649	-22.06	0.000	aged 3 years
age_4	-1.5346	0.0595	-25.77	0.000	aged 4 years
female	0.1330	0.0368	3.62	0.000	female
cellphone	0.2026	0.0503	4.03	0.000	household has cellphone
radio	0.3471	0.0909	3.82	0.000	household has radio
motorbike	0.1792	0.0476	3.77	0.000	household has motorbike
roof_natural	-0.1155	0.0498	-2.32	0.021	household has natural roof
school	0.3447	0.1235	2.79	0.005	prop of eligible mothers attended school
educ_primary	-0.2995	0.1209	-2.48	0.014	prop of eligible mothers with highest educ attainment primary
educ_low_sec	-0.1724	0.1166	-1.48	0.140	prop of eligible mothers with highest educ attainment low_second
dsw_spouse	0.2033	0.0939	2.17	0.031	prop of eligible mothers divorced, widowed or separated
pkids06_e	-2.2573	0.8483	-2.66	0.008	prop of ea aged 0-6
cellphone_e	0.3099	0.1000	3.1	0.002	mean number of cellphones owned
motorbike_e	-0.3922	0.1703	-2.3	0.022	mean number of motorbikes owned
boat_e	-0.2721	0.1106	-2.46	0.014	mean number of boats owned
psenior_e	3.3011	1.3265	2.49	0.013	prop of ea aged 65+
bovill_e	0.2024	0.1290	1.57	0.117	propn born outside current village
region_3	0.3734	0.1536	2.43	0.015	rural (outside Phnom Penh)
schoolXS2	0.4859	0.1600	3.04	0.002	interaction of school and other urban
cellphoneXS2	-0.2850	0.1304	-2.19	0.029	interaction of cellphone and other urban
radioXS3	-0.2584	0.1009	-2.56	0.011	interaction of radio and rural
_cons	-1.0001	0.2049	-4.88	0.000	constant term

B.3 Model for height-for-age in CDHS2010

n	p	R^2	σ_c^2	σ_h^2	σ_e^2
3986	15	0.1680	0.0562	0.4177	1.2527

where n = sample size, p = number of variables, R^2 = coefficient of determination
σ_c^2 = cluster-level variance, σ_h^2 = household-level variance, σ_e^2 = residual variance

Variable	Coef.	Std. Err.	t	P>t	Label
age_1	-0.9009	0.0832	-10.83	0.000	aged 1 year
age_2	-1.1973	0.0795	-15.06	0.000	aged 2 years
age_3	-1.1971	0.0731	-16.37	0.000	aged 3 years
age_4	-1.2412	0.0695	-17.86	0.000	aged 4 years
hhsize	-0.0426	0.0136	-3.12	0.002	number of household members
radio	0.1179	0.0525	2.25	0.025	household has a radio
cell	0.1248	0.0544	2.29	0.022	household has a mobile telephone
h_seced	0.1484	0.0532	2.79	0.005	head of household attended secondary school
floor_t	0.3162	0.1113	2.84	0.005	tiled floor
plit_v	0.9045	0.2690	3.36	0.001	prop of village literate
regnonb_v	0.5208	0.2602	2.00	0.046	propn in village non-buddhist
boat_v	0.2738	0.0836	3.27	0.001	mean number of boats per hh in village
koyaon_v	0.7981	0.3855	2.07	0.039	mean number of koyaons per hh in village
phone_v	3.3405	1.5933	2.10	0.036	mean number of phones per hh in village
pseced_v	0.5252	0.3153	1.67	0.096	prop of village with secondary education
_cons	-1.5334	0.1884	-8.14	0.000	constant tern

B.4 Model for weight-for-age in CAS2008

n	p	R^2	σ_c^2	σ_h^2	σ_e^2
					0.8101
7965	16	0.155	0.0642	0.2791	

where n = sample size, p = number of variables, R^2 = coefficient of determination
σ_c^2 = cluster-level variance, σ_h^2 = household-level variance, σ_e^2 = residual variance

Variable	Coef.	Std. Err.	t	P>t	Label
age1	-0.6095	0.0442	-13.80	0.000	aged 1 year
age2	-0.8022	0.0385	-20.81	0.000	aged 2 years
age3	-0.8960	0.0504	-17.79	0.000	aged 3 years
age4	-1.0083	0.0444	-22.69	0.000	aged 4 years
motorbike	0.1350	0.0373	3.62	0.000	household has motorbike
radio	0.0874	0.0355	2.46	0.014	household has radio
cellphone	0.0804	0.0359	2.24	0.026	household has cellphone
educ_usec	0.2399	0.0943	2.54	0.011	prop of eligible mothers with upper sec education
school	0.0797	0.0393	2.03	0.043	prop of eligible mothers attended school
car	0.1628	0.0888	1.83	0.067	household has car
pkids06_e	-1.1856	0.5711	-2.08	0.038	prop of ea aged 0-6
hmovins_e	0.2029	0.0643	3.15	0.002	prop hhead moved for insecurity/calamity/unemployment
mtnonkh_e	0.2739	0.0955	2.87	0.004	prop mother-tongue non-khmer
h_lit_e	0.2860	0.1194	2.40	0.017	propn head of household literate
motorbike_e	-0.2871	0.1289	-2.23	0.026	mean number of motorbikes owned
cellphone_e	0.2699	0.0632	4.27	0.000	mean number of cellphones owned
_cons	-0.9870	0.1577	-6.26	0.000	constant term

B.5 Model for weight-for-age in CDHS2010

n	p	R^2	σ_c^2	σ_h^2	σ_e^2
3985	20	0.1410	0.0111	0.2760	0.7413

where n = sample size, p = number of variables, R^2 = coefficient of determination
σ_c^2 = cluster-level variance, σ_h^2 = household-level variance, σ_e^2 = residual variance

Variable	Coef.	Std. Err.	t	P>t	Label
age_1	-0.4360	0.0622	-7.01	0.000	aged 1 year
age_2	-0.5939	0.0560	-10.61	0.000	aged 2 years
age_3	-0.6767	0.0598	-11.32	0.000	aged 3 years
age_4	-0.7799	0.0577	-13.51	0.000	aged 4 years
hhsize	-0.0193	0.0096	-2.01	0.045	number of household members
h_seced	0.1513	0.0447	3.39	0.001	head of household attended secondary school
boat	0.1855	0.0725	2.56	0.011	household has a boat
cell	0.1387	0.0455	3.05	0.002	household has a mobile telephone
dr_public	0.2008	0.0820	2.45	0.015	drinking water from pipe or public tap
ck_charc	0.1465	0.0757	1.94	0.053	cook with charcoal
floor_t	0.2537	0.0876	2.90	0.004	floor type tile
floor_o	-0.8250	0.2520	-3.27	0.001	floor type other
roof_4	0.3647	0.1594	2.29	0.023	roof type concrete
region_2	0.1565	0.1053	1.49	0.138	urban (outside Phnom Penh)
region_3	0.1686	0.1142	1.48	0.141	rural (outside Phnom Penh)
plit_v	0.6242	0.1586	3.94	0.000	prop of village literate
koyaon_v	0.9078	0.3286	2.76	0.006	mean number of koyaons per hh in village
bovill_v	0.3944	0.0958	4.11	0.000	prop born outside current village
h_senior_v	1.4560	0.5040	2.89	0.004	prop head of hh in village over 65 years
regnonb_v	0.2574	0.1439	1.79	0.074	propn in village non-buddhist
_cons	-1.8614	0.1994	-9.34	0.000	constant term

Appendix C. Small-area estimates

C.1 District-level poverty measures

P0 = poverty incidence, se0 = standard error of P0, P1 = poverty gap, etc.

pcode	province	dcode	district	P0	se0	P1	se1	P2	se2
1	Banteay Meanchey	2	Mongkol Borei	0.2404	0.0141	0.0656	0.0047	0.0233	0.0020
		3	Phnum Srok	0.2941	0.0237	0.0814	0.0084	0.0288	0.0038
		4	Preah Netr Preah	0.2798	0.0192	0.0772	0.0067	0.0273	0.0029
		5	Ou Chrov	0.2419	0.0183	0.0649	0.0060	0.0224	0.0026
		6	Serei Saophoan	0.2264	0.0157	0.0620	0.0058	0.0229	0.0027
		7	Thma Puok	0.2557	0.0207	0.0694	0.0072	0.0242	0.0032
		8	Svay Chek	0.2606	0.0183	0.0697	0.0063	0.0240	0.0028
		9	Malai	0.2889	0.0201	0.0812	0.0074	0.0296	0.0035
		10	Ou Chrov	0.2677	0.0158	0.0747	0.0060	0.0282	0.0029
2	Battambang	1	Banan	0.2773	0.0154	0.0782	0.0056	0.0283	0.0025
		2	Thma Koul	0.3099	0.0153	0.0894	0.0058	0.0339	0.0027
		3	Bat Dambang	0.1747	0.0126	0.0460	0.0043	0.0166	0.0019
		4	Bavel	0.2921	0.0166	0.0829	0.0059	0.0301	0.0026
		5	Aek Phnum	0.2100	0.0183	0.0564	0.0059	0.0195	0.0025
		6	Moung Ruessei	0.3032	0.0138	0.0872	0.0053	0.0321	0.0025
		7	Rotonak Mondol	0.2901	0.0245	0.0835	0.0095	0.0307	0.0045
		8	Sangkae	0.2967	0.0168	0.0848	0.0061	0.0315	0.0028
		9	Samlout	0.2836	0.0230	0.0806	0.0084	0.0294	0.0038

		10	Sampov Lun	0.2152	0.0227	0.0592	0.0076	0.0210	0.0033
		11	Phnom Proek	0.2238	0.0216	0.0618	0.0072	0.0220	0.0030
		12	Kamrieng	0.2429	0.0206	0.0678	0.0073	0.0243	0.0033
		13	Koas Krala	0.3566	0.0279	0.1052	0.0112	0.0396	0.0055
		14	Moung Ruessei	0.3497	0.0252	0.1019	0.0098	0.0378	0.0045
3	Kampong Cham	1	Batheay	0.2443	0.0137	0.0655	0.0048	0.0226	0.0021
		2	Chamkar Leu	0.2143	0.0109	0.0586	0.0038	0.0207	0.0017
		3	Cheung Prey	0.2621	0.0169	0.0708	0.0058	0.0249	0.0025
		4	Dambae	0.2567	0.0188	0.0704	0.0066	0.0248	0.0029
		5	Kampong Cham	0.1326	0.0163	0.0329	0.0052	0.0115	0.0023
		6	Kampong Siem	0.1887	0.0132	0.0489	0.0042	0.0164	0.0017
		7	Kang Meas	0.2169	0.0140	0.0574	0.0048	0.0196	0.0021
		8	Kaoh Soutin	0.1848	0.0148	0.0478	0.0048	0.0160	0.0020
		9	Krouch Chhmar	0.2221	0.0130	0.0593	0.0046	0.0204	0.0020
		10	Memot	0.2289	0.0139	0.0622	0.0049	0.0218	0.0022
		11	Ou Reang Ov	0.1894	0.0124	0.0491	0.0041	0.0164	0.0017
		12	Ponhea Kraek	0.2084	0.0115	0.0552	0.0037	0.0189	0.0016
		13	Prey Chhor	0.2350	0.0124	0.0627	0.0043	0.0217	0.0019
		14	Srei Santhor	0.2011	0.0126	0.0526	0.0041	0.0178	0.0017
		15	Stueng Trang	0.2281	0.0139	0.0619	0.0047	0.0216	0.0021
		16	Tboung Khmum	0.2174	0.0124	0.0578	0.0041	0.0200	0.0018
		17	Tboung Khmum	0.1933	0.0207	0.0506	0.0074	0.0180	0.0034
4	Kampong Chhnang	1	Baribour	0.2583	0.0203	0.0718	0.0070	0.0256	0.0030
		2	Chol Kiri	0.2970	0.0385	0.0835	0.0133	0.0301	0.0057
		3	Kampong Chhnang	0.2423	0.0258	0.0652	0.0093	0.0242	0.0042
		4	Kampong Leaeng	0.3179	0.0309	0.0905	0.0115	0.0330	0.0052
		5	Kampong Tralach	0.2809	0.0173	0.0787	0.0064	0.0283	0.0029

		6	Rolea B'ier	0.2381	0.0170	0.0647	0.0058	0.0227	0.0025
		7	Sameakki Mean Chey	0.3026	0.0190	0.0847	0.0070	0.0304	0.0032
		8	Tuek Phos	0.3185	0.0238	0.0913	0.0089	0.0335	0.0040
5	Kampong Speu	1	Basedth	0.2812	0.0200	0.0779	0.0071	0.0277	0.0032
		2	Chbar Mon	0.2371	0.0200	0.0638	0.0074	0.0237	0.0035
		3	Kong Pisei	0.2332	0.0194	0.0626	0.0064	0.0216	0.0026
		4	Aoral	0.3345	0.0284	0.0967	0.0109	0.0356	0.0051
		5	Odongk	0.2522	0.0185	0.0692	0.0063	0.0245	0.0027
		6	Phnum Sruoch	0.2855	0.0182	0.0810	0.0063	0.0294	0.0028
		7	Samraong Tong	0.2341	0.0180	0.0635	0.0062	0.0222	0.0026
		8	Thpong	0.3058	0.0259	0.0861	0.0095	0.0311	0.0043
6	Kampong Thom	1	Baray	0.2617	0.0138	0.0725	0.0047	0.0258	0.0021
		2	Kampong Svay	0.3064	0.0191	0.0879	0.0074	0.0323	0.0035
		3	Stueng Saen	0.2243	0.0180	0.0609	0.0068	0.0221	0.0032
		4	Prasat Ballangk	0.3856	0.0263	0.1149	0.0108	0.0434	0.0052
		5	Prasat Sambour	0.2964	0.0244	0.0833	0.0085	0.0300	0.0038
		6	Sandan	0.3233	0.0234	0.0931	0.0089	0.0342	0.0041
		7	Santuk	0.3022	0.0175	0.0859	0.0065	0.0312	0.0029
		8	Stoung	0.2856	0.0167	0.0806	0.0062	0.0291	0.0028
7	Kampot	1	Angkor Chey	0.1863	0.0138	0.0479	0.0045	0.0159	0.0019
		2	Banteay Meas	0.2069	0.0138	0.0541	0.0045	0.0183	0.0019
		3	Chhuk	0.2259	0.0151	0.0605	0.0051	0.0209	0.0021
		4	Chum Kiri	0.2563	0.0213	0.0689	0.0072	0.0238	0.0031
		5	Dang Tong	0.2490	0.0220	0.0671	0.0072	0.0233	0.0030
		6	Kampong Trach	0.2541	0.0158	0.0689	0.0055	0.0242	0.0024
		7	Kampot	0.2580	0.0149	0.0712	0.0051	0.0254	0.0023
		8	Kampong Bay	0.1848	0.0196	0.0497	0.0074	0.0182	0.0035

8	Kandal	1	Kandal Stueng	0.1373	0.0092	0.0345	0.0028	0.0113	0.0011
		2	Kien Svay	0.1727	0.0122	0.0446	0.0039	0.0154	0.0016
		3	Khsach Kandal	0.1816	0.0116	0.0472	0.0038	0.0158	0.0016
		4	Kaoh Thum	0.1917	0.0121	0.0500	0.0039	0.0169	0.0016
		5	Leuk Daek	0.2157	0.0184	0.0578	0.0060	0.0200	0.0025
		6	Lvea Aem	0.1672	0.0151	0.0426	0.0047	0.0141	0.0019
		7	Mukh Kampul	0.1537	0.0123	0.0398	0.0040	0.0133	0.0017
		8	Angk Snuol	0.1208	0.0085	0.0295	0.0025	0.0096	0.0010
		9	Popnhea Lueu	0.1677	0.0121	0.0433	0.0038	0.0146	0.0015
		10	S'ang	0.1686	0.0099	0.0440	0.0032	0.0149	0.0013
		11	Ta Khmau	0.1206	0.0129	0.0295	0.0041	0.0102	0.0018
9	Koh Kong	1	Botum Sakor	0.1948	0.0265	0.0515	0.0088	0.0175	0.0037
		2	Kiri Sakor	0.1482	0.0400	0.0370	0.0127	0.0119	0.0052
		3	Kaoh Kong	0.1502	0.0393	0.0383	0.0122	0.0127	0.0050
		4	Smach Mean Chey	0.1636	0.0187	0.0416	0.0061	0.0147	0.0028
		5	Mondol Seima	0.0893	0.0209	0.0216	0.0062	0.0067	0.0024
		6	Srae Ambel	0.2600	0.0234	0.0719	0.0087	0.0262	0.0041
		7	Thma Bang	0.2225	0.0399	0.0585	0.0137	0.0198	0.0060
10	Kratie	1	Chhloung	0.2801	0.0223	0.0795	0.0082	0.0292	0.0037
		2	Kracheh	0.2617	0.0304	0.0740	0.0118	0.0285	0.0057
		3	Preaek Prasab	0.2655	0.0229	0.0736	0.0085	0.0262	0.0039
		4	Sambour	0.3692	0.0249	0.1120	0.0098	0.0432	0.0047
		5	Snuol	0.2952	0.0234	0.0855	0.0088	0.0316	0.0040
		6	Kracheh	0.3536	0.0248	0.1056	0.0098	0.0400	0.0047
11	Mondul Kiri	1	Kaev Seima	0.3151	0.0328	0.0934	0.0129	0.0355	0.0063
		2	Kaoh Nheaek	0.3581	0.0392	0.1062	0.0164	0.0402	0.0082
		3	Ou Reang	0.2826	0.0571	0.0807	0.0210	0.0297	0.0098
		4	Pech Chreada	0.2767	0.0452	0.0770	0.0154	0.0275	0.0067

		5	Saen Monourom	0.1550	0.0315	0.0414	0.0106	0.0147	0.0047
12	Phnom Penh	1	Chamkar Mon	0.0646	0.0107	0.0125	0.0024	0.0036	0.0008
		2	Doun Penh	0.0828	0.0130	0.0165	0.0033	0.0049	0.0012
		3	Prampir Meakkakra	0.0855	0.0152	0.0169	0.0037	0.0050	0.0013
		4	Tuol Kouk	0.0737	0.0132	0.0145	0.0031	0.0042	0.0011
		5	Dangkao	0.1773	0.0167	0.0386	0.0046	0.0122	0.0018
		6	Mean Chey	0.1377	0.0137	0.0287	0.0036	0.0088	0.0013
		7	Ruessei Kaev	0.1425	0.0162	0.0297	0.0045	0.0091	0.0017
		8	Ruessei Kaev	0.1299	0.0172	0.0272	0.0045	0.0084	0.0016
13	Preah Vihear	1	Chey Saen	0.4344	0.0402	0.1320	0.0167	0.0507	0.0083
		2	Chhaeb	0.4315	0.0421	0.1348	0.0176	0.0531	0.0089
		3	Choam Khsant	0.3171	0.0298	0.0918	0.0114	0.0339	0.0053
		4	Kuleaen	0.3347	0.0386	0.0977	0.0146	0.0363	0.0067
		5	Rovieng	0.3712	0.0276	0.1102	0.0111	0.0416	0.0055
		6	Sangkom Thmei	0.3679	0.0368	0.1074	0.0144	0.0400	0.0068
		7	Tbaeng Mean Chey	0.3984	0.0536	0.1204	0.0222	0.0462	0.0111
		8	Tbaeng Mean Chey	0.2217	0.0254	0.0608	0.0093	0.0222	0.0044
14	Prey Veng	1	Ba Phnum	0.1962	0.0137	0.0505	0.0044	0.0168	0.0018
		2	Kamchay Mear	0.1989	0.0138	0.0510	0.0045	0.0169	0.0019
		3	Kampong Trabaek	0.2212	0.0115	0.0583	0.0039	0.0198	0.0017
		4	Kanhchriech	0.1852	0.0144	0.0476	0.0046	0.0158	0.0019
		5	Me Sang	0.2036	0.0123	0.0523	0.0039	0.0173	0.0016
		6	Peam Chor	0.2298	0.0185	0.0604	0.0058	0.0205	0.0024
		7	Peam Ro	0.2093	0.0154	0.0556	0.0052	0.0194	0.0023
		8	Pea Reang	0.1907	0.0129	0.0486	0.0040	0.0160	0.0016
		9	Preah Sdach	0.2612	0.0134	0.0711	0.0048	0.0249	0.0022
		10	Kampong Leav	0.1398	0.0188	0.0347	0.0058	0.0116	0.0024

		11	Kampong Leav	0.1753	0.0190	0.0441	0.0059	0.0143	0.0025
		12	Sithor Kandal	0.1916	0.0168	0.0490	0.0051	0.0162	0.0020
		13	Prey Veaeng	0.1950	0.0132	0.0499	0.0042	0.0165	0.0017
15	Pursat	1	Bakan	0.2948	0.0162	0.0829	0.0060	0.0299	0.0027
		2	Kandieng	0.3064	0.0195	0.0875	0.0076	0.0321	0.0036
		3	Krakor	0.2949	0.0197	0.0843	0.0073	0.0309	0.0034
		4	Phnum Kravanh	0.3030	0.0200	0.0863	0.0075	0.0315	0.0035
		5	Sampov Meas	0.2405	0.0160	0.0678	0.0056	0.0251	0.0026
		6	Veal Veaeng	0.2715	0.0429	0.0761	0.0151	0.0273	0.0067
16	Ratanak Kiri	1	Andoung Meas	0.4846	0.0534	0.1667	0.0264	0.0717	0.0148
		2	Ban Lung	0.1759	0.0232	0.0455	0.0074	0.0159	0.0032
		3	Bar Kaev	0.3107	0.0341	0.0918	0.0132	0.0346	0.0062
		4	Koun Mom	0.3117	0.0375	0.0894	0.0138	0.0328	0.0063
		5	Lumphat	0.3131	0.0385	0.0905	0.0146	0.0334	0.0068
		6	Ou Chum	0.3993	0.0432	0.1267	0.0182	0.0510	0.0092
		7	Ou Ya Dav	0.3624	0.0392	0.1106	0.0172	0.0431	0.0090
		8	Ta Veaeng	0.5095	0.0570	0.1763	0.0263	0.0761	0.0143
		9	Veun Sai	0.4073	0.0421	0.1302	0.0194	0.0528	0.0103
17	Siemreap	1	Angkor Chum	0.3466	0.0242	0.1000	0.0092	0.0368	0.0042
		2	Angkor Thum	0.3727	0.0373	0.1109	0.0152	0.0420	0.0074
		3	Banteay Srei	0.2916	0.0248	0.0830	0.0095	0.0303	0.0045
		4	Chi Kraeng	0.3272	0.0175	0.0945	0.0065	0.0348	0.0030
		6	Kralanh	0.3084	0.0230	0.0875	0.0083	0.0320	0.0037
		7	Puok	0.3046	0.0173	0.0882	0.0067	0.0326	0.0032
		9	Prasat Bakong	0.2331	0.0215	0.0637	0.0071	0.0224	0.0030
		10	Siem Reab	0.1804	0.0113	0.0490	0.0039	0.0181	0.0018
		11	Soutr Nikom	0.2861	0.0182	0.0809	0.0070	0.0294	0.0032
		12	Srei Snam	0.3763	0.0288	0.1109	0.0120	0.0416	0.0060

		13	Svay Leu	0.3481	0.0329	0.1019	0.0125	0.0380	0.0059
		14	Varin	0.3500	0.0296	0.1020	0.0112	0.0379	0.0052
18	Sihanoukville	1	Mittapheap	0.1309	0.0141	0.0319	0.0045	0.0110	0.0019
		2	Prey Nob	0.2174	0.0155	0.0590	0.0053	0.0206	0.0023
		3	Stueng Hav	0.1395	0.0215	0.0364	0.0071	0.0123	0.0030
		4	Kampong Seila	0.1948	0.0283	0.0517	0.0093	0.0177	0.0039
19	Stung Treng	1	Sesan	0.3160	0.0389	0.0902	0.0139	0.0329	0.0063
		2	Siem Bouk	0.3067	0.0381	0.0859	0.0141	0.0308	0.0065
		3	Siem Pang	0.3927	0.0421	0.1197	0.0167	0.0462	0.0081
		4	Stueng Traeng	0.1957	0.0233	0.0525	0.0084	0.0189	0.0039
		5	Thala Barivat	0.3784	0.0321	0.1137	0.0127	0.0434	0.0062
20	Svay Rieng	1	Chantrea	0.2085	0.0213	0.0548	0.0074	0.0186	0.0033
		2	Kampong Rou	0.1901	0.0140	0.0487	0.0046	0.0161	0.0019
		3	Rumduol	0.1911	0.0145	0.0487	0.0046	0.0160	0.0019
		4	Romeas Haek	0.2048	0.0123	0.0526	0.0040	0.0174	0.0016
		5	Svay Chrum	0.1981	0.0121	0.0512	0.0037	0.0171	0.0015
		6	Svay Rieng	0.1207	0.0120	0.0298	0.0038	0.0098	0.0016
		7	Svay Teab	0.1840	0.0181	0.0470	0.0057	0.0155	0.0024
		8	Chantrea	0.1106	0.0160	0.0276	0.0050	0.0089	0.0020
21	Takeo	1	Angkor Borei	0.2010	0.0214	0.0519	0.0067	0.0173	0.0027
		2	Bati	0.1813	0.0101	0.0471	0.0033	0.0158	0.0014
		3	Borei Cholsar	0.2163	0.0248	0.0567	0.0082	0.0192	0.0035
		4	Kiri Vong	0.1973	0.0127	0.0517	0.0041	0.0175	0.0017
		5	Kaoh Andaet	0.2273	0.0173	0.0605	0.0059	0.0208	0.0026
		6	Prey Kabbas	0.1688	0.0128	0.0430	0.0038	0.0142	0.0015
		7	Samraong	0.2028	0.0118	0.0527	0.0040	0.0177	0.0017
		8	Doun Kaev	0.1543	0.0153	0.0395	0.0051	0.0134	0.0022
		9	Tram Kak	0.1923	0.0101	0.0500	0.0033	0.0168	0.0014

		10	Treang	0.2075	0.0129	0.0540	0.0043	0.0182	0.0018
22	Oddar Meanchey	1	Anlong Veaeng	0.2713	0.0261	0.0778	0.0092	0.0287	0.0042
		2	Banteay Ampil	0.3204	0.0274	0.0893	0.0096	0.0318	0.0043
		3	Chong Kal	0.3465	0.0342	0.0996	0.0129	0.0365	0.0059
		4	Samraong	0.3435	0.0226	0.1034	0.0091	0.0405	0.0045
		5	Trapeang Prasat	0.3062	0.0272	0.0876	0.0099	0.0321	0.0045
23	Kep	1	Damnak Chang'aeur	0.2574	0.0341	0.0705	0.0113	0.0249	0.0048
		2	Kaeb	0.2221	0.0358	0.0602	0.0121	0.0216	0.0053
24	Pailin	1	Pailin	0.2017	0.0197	0.0561	0.0071	0.0207	0.0033
		2	Sala Krau	0.2275	0.0236	0.0632	0.0081	0.0226	0.0035

C.2 District-level stunting measures

S2 = incidence of stunting, se2 = standard error of S2,
S3 = incidence of severe stunting, se3 = standard error of S3

pcode	province	dcode	district	S2	seS2	S3	seS3
1	Banteay Meanchey	2	Mongkol Borei	0.3925	0.0072	0.1694	0.0053
		3	Phnum Srok	0.4217	0.0115	0.1884	0.0080
		4	Preah Netr Preah	0.4266	0.0090	0.1925	0.0065
		5	Ou Chrov	0.4049	0.0105	0.1767	0.0074
		6	Serei Saophoan	0.3470	0.0094	0.1398	0.0055
		7	Thma Puok	0.4139	0.0121	0.1851	0.0079
		8	Svay Chek	0.4186	0.0102	0.1862	0.0069
		9	Malai	0.4072	0.0126	0.1782	0.0093
		10	Ou Chrov	0.3724	0.0151	0.1593	0.0101
2	Battambang	1	Banan	0.4202	0.0085	0.1881	0.0061
		2	Thma Koul	0.3925	0.0089	0.1697	0.0064
		3	Bat Dambang	0.2990	0.0097	0.1123	0.0054
		4	Bavel	0.4079	0.0093	0.1795	0.0069
		5	Aek Phnum	0.4069	0.0131	0.1711	0.0100
		6	Moung Ruessei	0.4160	0.0084	0.1844	0.0057
		7	Rotonak Mondol	0.4234	0.0109	0.1912	0.0079
		8	Sangkae	0.3988	0.0087	0.1719	0.0061
		9	Samlout	0.4231	0.0117	0.1855	0.0084
		10	Sampov Lun	0.4020	0.0158	0.1771	0.0109
		11	Phnom Proek	0.3997	0.0144	0.1752	0.0101
		12	Kamrieng	0.4040	0.0131	0.1763	0.0089
		13	Koas Krala	0.4373	0.0116	0.1963	0.0089
		14	Moung Ruessei	0.4736	0.0124	0.2247	0.0095
3	Kampong Cham	1	Batheay	0.4371	0.0083	0.1997	0.0061
		2	Chamkar Leu	0.4236	0.0080	0.1900	0.0058
		3	Cheung Prey	0.4268	0.0079	0.1922	0.0060

	4 Dambae	0.4673	0.0107	0.2208	0.0084
	5 Kampong Cham	0.2838	0.0139	0.1065	0.0080
	6 Kampong Siem	0.3909	0.0080	0.1679	0.0055
	7 Kang Meas	0.4116	0.0090	0.1803	0.0062
	8 Kaoh Soutin	0.3816	0.0110	0.1599	0.0071
	9 Krouch Chhmar	0.4068	0.0098	0.1787	0.0070
	10 Memot	0.4748	0.0109	0.2252	0.0087
	11 Ou Reang Ov	0.4080	0.0087	0.1794	0.0064
	12 Ponhea Kraek	0.4309	0.0101	0.1964	0.0076
	13 Prey Chhor	0.4072	0.0069	0.1775	0.0050
	14 Srei Santhor	0.4074	0.0092	0.1773	0.0065
	15 Stueng Trang	0.4381	0.0076	0.2016	0.0059
	16 Tboung Khmum	0.4121	0.0073	0.1836	0.0053
	17 Tboung Khmum	0.3705	0.0120	0.1601	0.0091
4 Kampong Chhnang	1 Baribour	0.4206	0.0114	0.1858	0.0083
	2 Chol Kiri	0.4389	0.0196	0.1888	0.0142
	3 Kampong Chhnang	0.3462	0.0135	0.1369	0.0087
	4 Kampong Leaeng	0.4656	0.0143	0.2121	0.0104
	5 Kampong Tralach	0.4200	0.0088	0.1866	0.0063
	6 Rolea B'ier	0.4021	0.0068	0.1742	0.0050
	7 Sameakki Mean Chey	0.4462	0.0081	0.2041	0.0065
	8 Tuek Phos	0.4450	0.0089	0.2055	0.0072
5 Kampong Speu	1 Basedth	0.4415	0.0072	0.2028	0.0053
	2 Chbar Mon	0.3368	0.0136	0.1363	0.0084
	3 Kong Pisei	0.4068	0.0080	0.1771	0.0057
	4 Aoral	0.4953	0.0133	0.2418	0.0109
	5 Odongk	0.4157	0.0076	0.1833	0.0054
	6 Phnum Sruoch	0.4409	0.0075	0.2035	0.0059
	7 Samraong Tong	0.3981	0.0077	0.1729	0.0052
	8 Thpong	0.4611	0.0103	0.2156	0.0078
6 Kampong Thom	1 Baray	0.4280	0.0067	0.1942	0.0051
	2 Kampong Svay	0.4502	0.0083	0.2071	0.0062
	3 Stueng Saen	0.3633	0.0098	0.1516	0.0069
	4 Prasat Ballangk	0.4967	0.0122	0.2431	0.0095

	5 Prasat Sambour	0.4511	0.0103	0.2100	0.0074
	6 Sandan	0.4814	0.0101	0.2346	0.0087
	7 Santuk	0.4495	0.0085	0.2091	0.0066
	8 Stoung	0.4684	0.0089	0.2221	0.0071
7 Kampot	1 Angkor Chey	0.3889	0.0094	0.1658	0.0060
	2 Banteay Meas	0.4075	0.0087	0.1781	0.0057
	3 Chhuk	0.4331	0.0083	0.1966	0.0064
	4 Chum Kiri	0.4411	0.0103	0.2017	0.0076
	5 Dang Tong	0.4293	0.0088	0.1924	0.0066
	6 Kampong Trach	0.4342	0.0079	0.1959	0.0058
	7 Kampot	0.4204	0.0076	0.1885	0.0059
	8 Kampong Bay	0.3137	0.0144	0.1231	0.0083
8 Kandal	1 Kandal Stueng	0.3667	0.0084	0.1524	0.0054
	2 Kien Svay	0.3644	0.0083	0.1530	0.0055
	3 Khsach Kandal	0.3925	0.0070	0.1682	0.0050
	4 Kaoh Thum	0.4182	0.0105	0.1863	0.0076
	5 Leuk Daek	0.3985	0.0110	0.1726	0.0077
	6 Lvea Aem	0.3915	0.0106	0.1657	0.0072
	7 Mukh Kampul	0.3765	0.0093	0.1605	0.0061
	8 Angk Snuol	0.3560	0.0086	0.1444	0.0056
	9 Popnhea Lueu	0.3838	0.0078	0.1637	0.0051
	10 S'ang	0.3923	0.0081	0.1707	0.0055
	11 Ta Khmau	0.2932	0.0133	0.1105	0.0074
9 Koh Kong	1 Botum Sakor	0.4213	0.0179	0.1863	0.0122
	2 Kiri Sakor	0.3893	0.0276	0.1657	0.0182
	3 Kaoh Kong	0.4158	0.0274	0.1788	0.0191
	4 Smach Mean Chey	0.3570	0.0153	0.1485	0.0093
	5 Mondol Seima	0.3663	0.0231	0.1543	0.0158
	6 Srae Ambel	0.4258	0.0103	0.1963	0.0077
	7 Thma Bang	0.4736	0.0179	0.2267	0.0162
10 Kratie	1 Chhloung	0.4184	0.0095	0.1898	0.0068
	2 Kracheh	0.3381	0.0138	0.1377	0.0089
	3 Preaek Prasab	0.4131	0.0087	0.1821	0.0068
	4 Sambour	0.4834	0.0116	0.2352	0.0089

	5 Snuol	0.4690	0.0109	0.2227	0.0087
	6 Kracheh	0.4694	0.0093	0.2240	0.0076
11 Mondul Kiri	1 Kaev Seima	0.4871	0.0165	0.2359	0.0139
	2 Kaoh Nheaek	0.4979	0.0161	0.2412	0.0126
	3 Ou Reang	0.4756	0.0324	0.2239	0.0267
	4 Pech Chreada	0.4894	0.0236	0.2378	0.0173
	5 Saen Monourom	0.4065	0.0176	0.1828	0.0143
12 Phnom Penh	1 Chamkar Mon	0.2412	0.0187	0.0838	0.0094
	2 Doun Penh	0.2559	0.0183	0.0920	0.0094
	3 Prampir Meakkakra	0.2347	0.0182	0.0802	0.0094
	4 Tuol Kouk	0.2568	0.0178	0.0899	0.0090
	5 Dangkao	0.3787	0.0212	0.1590	0.0139
	6 Mean Chey	0.3414	0.0207	0.1376	0.0129
	7 Ruessei Kaev	0.3348	0.0200	0.1334	0.0121
	8 Ruessei Kaev	0.3262	0.0202	0.1283	0.0124
13 Preah Vihear	1 Chey Saen	0.5034	0.0166	0.2500	0.0131
	2 Chhaeb	0.5011	0.0170	0.2471	0.0139
	3 Choam Khsant	0.4597	0.0157	0.2188	0.0121
	4 Kuleaen	0.4859	0.0148	0.2357	0.0124
	5 Rovieng	0.4869	0.0111	0.2385	0.0095
	6 Sangkom Thmei	0.4670	0.0137	0.2234	0.0111
	7 Tbaeng Mean Chey	0.4867	0.0190	0.2367	0.0155
	8 Tbaeng Mean Chey	0.3888	0.0133	0.1679	0.0103
14 Prey Veng	1 Ba Phnum	0.4104	0.0081	0.1802	0.0054
	2 Kamchay Mear	0.4174	0.0080	0.1838	0.0059
	3 Kampong Trabaek	0.4272	0.0070	0.1910	0.0052
	4 Kanhchriech	0.4103	0.0091	0.1783	0.0065
	5 Me Sang	0.4319	0.0076	0.1929	0.0056
	6 Peam Chor	0.4409	0.0091	0.1960	0.0066
	7 Peam Ro	0.3949	0.0089	0.1701	0.0060
	8 Pea Reang	0.4151	0.0076	0.1832	0.0054
	9 Preah Sdach	0.4414	0.0070	0.2016	0.0051
	10 Kampong Leav	0.3475	0.0124	0.1435	0.0096
	11 Kampong Leav	0.4249	0.0121	0.1849	0.0090

	12 Sithor Kandal	0.4147	0.0103	0.1817	0.0074
	13 Prey Veaeng	0.4214	0.0070	0.1871	0.0054
15 Pursat	1 Bakan	0.4327	0.0065	0.1957	0.0049
	2 Kandieng	0.4435	0.0092	0.2022	0.0069
	3 Krakor	0.4352	0.0093	0.1968	0.0068
	4 Phnum Kravanh	0.4395	0.0094	0.2014	0.0070
	5 Sampov Meas	0.3715	0.0090	0.1589	0.0057
	6 Veal Veaeng	0.4544	0.0169	0.2104	0.0128
16 Ratanak Kiri	1 Andoung Meas	0.5428	0.0233	0.2823	0.0201
	2 Ban Lung	0.4035	0.0154	0.1797	0.0107
	3 Bar Kaev	0.5053	0.0192	0.2536	0.0165
	4 Koun Mom	0.4901	0.0173	0.2399	0.0140
	5 Lumphat	0.5179	0.0167	0.2590	0.0143
	6 Ou Chum	0.5212	0.0227	0.2626	0.0186
	7 Ou Ya Dav	0.5265	0.0226	0.2685	0.0193
	8 Ta Veaeng	0.5266	0.0244	0.2658	0.0203
	9 Veun Sai	0.5358	0.0179	0.2714	0.0157
17 Siemreap	1 Angkor Chum	0.4804	0.0119	0.2308	0.0088
	2 Angkor Thum	0.5002	0.0136	0.2486	0.0120
	3 Banteay Srei	0.4729	0.0128	0.2258	0.0106
	4 Chi Kraeng	0.4700	0.0089	0.2239	0.0071
	6 Kralanh	0.4681	0.0096	0.2212	0.0076
	7 Puok	0.4415	0.0081	0.2046	0.0061
	9 Prasat Bakong	0.4329	0.0098	0.1952	0.0072
	10 Siem Reab	0.3642	0.0095	0.1519	0.0062
	11 Soutr Nikom	0.4424	0.0084	0.2034	0.0062
	12 Srei Snam	0.4798	0.0120	0.2297	0.0095
	13 Svay Leu	0.4938	0.0144	0.2430	0.0118
	14 Varin	0.4923	0.0144	0.2389	0.0120
18 Sihanoukville	1 Mittapheap	0.3166	0.0142	0.1224	0.0083
	2 Prey Nob	0.4076	0.0106	0.1782	0.0073
	3 Stueng Hav	0.3888	0.0188	0.1682	0.0130
	4 Kampong Seila	0.4077	0.0146	0.1801	0.0108
19 Stung Treng	1 Sesan	0.5033	0.0168	0.2509	0.0136

	2 Siem Bouk	0.4720	0.0151	0.2186	0.0119
	3 Siem Pang	0.5356	0.0165	0.2745	0.0141
	4 Stueng Traeng	0.3696	0.0134	0.1572	0.0090
	5 Thala Barivat	0.5088	0.0131	0.2544	0.0111
20 Svay Rieng	1 Chantrea	0.4180	0.0133	0.1779	0.0092
	2 Kampong Rou	0.4144	0.0101	0.1804	0.0071
	3 Rumduol	0.4183	0.0103	0.1828	0.0076
	4 Romeas Haek	0.4283	0.0085	0.1900	0.0060
	5 Svay Chrum	0.4247	0.0076	0.1891	0.0053
	6 Svay Rieng	0.3441	0.0123	0.1396	0.0082
	7 Svay Teab	0.4072	0.0098	0.1756	0.0072
	8 Chantrea	0.3696	0.0139	0.1519	0.0092
21 Takeo	1 Angkor Borei	0.4258	0.0114	0.1889	0.0084
	2 Bati	0.4040	0.0078	0.1754	0.0055
	3 Borei Cholsar	0.4283	0.0149	0.1874	0.0108
	4 Kiri Vong	0.4229	0.0088	0.1881	0.0063
	5 Kaoh Andaet	0.4147	0.0102	0.1816	0.0073
	6 Prey Kabbas	0.3973	0.0076	0.1703	0.0055
	7 Samraong	0.4136	0.0072	0.1831	0.0049
	8 Doun Kaev	0.3583	0.0099	0.1496	0.0068
	9 Tram Kak	0.4104	0.0067	0.1802	0.0047
	10 Treang	0.4112	0.0080	0.1798	0.0056
22 Oddar Meanchey	1 Anlong Veaeng	0.4238	0.0124	0.1928	0.0091
	2 Banteay Ampil	0.4631	0.0134	0.2170	0.0100
	3 Chong Kal	0.4615	0.0129	0.2181	0.0106
	4 Samraong	0.4291	0.0101	0.1983	0.0079
	5 Trapeang Prasat	0.4578	0.0141	0.2142	0.0108
23 Kep	1 Damnak Chang'aeur	0.4586	0.0138	0.2132	0.0106
	2 Kaeb	0.3895	0.0166	0.1667	0.0125
24 Pailin	1 Pailin	0.3745	0.0117	0.1597	0.0082
	2 Sala Krau	0.4184	0.0128	0.1857	0.0092

C.3 District-level underweight measures

U2 = incidence of underweight, se2 = standard error of U2,
U3 = incidence of severe underweight, se3 = standard error of U3

pcode	province	dcode	district	U2	seU2	U3	seU3
1	Banteay Meanchey	2	Mongkol Borei	0.2951	0.0079	0.0753	0.0034
		3	Phnum Srok	0.3289	0.0118	0.0896	0.0055
		4	Preah Netr Preah	0.3385	0.0079	0.0941	0.0041
		5	Ou Chrov	0.3077	0.0108	0.0783	0.0050
		6	Serei Saophoan	0.2527	0.0082	0.0579	0.0033
		7	Thma Puok	0.3202	0.0110	0.0867	0.0051
		8	Svay Chek	0.3254	0.0106	0.0890	0.0049
		9	Malai	0.2735	0.0113	0.0658	0.0048
		10	Ou Chrov	0.2382	0.0115	0.0529	0.0044
2	Battambang	1	Banan	0.3070	0.0080	0.0793	0.0036
		2	Thma Koul	0.3017	0.0083	0.0778	0.0038
		3	Bat Dambang	0.2091	0.0072	0.0426	0.0024
		4	Bavel	0.3040	0.0083	0.0780	0.0037
		5	Aek Phnum	0.3029	0.0079	0.0767	0.0038
		6	Moung Ruessei	0.3206	0.0079	0.0832	0.0037
		7	Rotonak Mondol	0.2792	0.0122	0.0670	0.0051
		8	Sangkae	0.3023	0.0077	0.0772	0.0035
		9	Samlout	0.2788	0.0135	0.0684	0.0054
		10	Sampov Lun	0.2704	0.0135	0.0642	0.0056
		11	Phnom Proek	0.2703	0.0119	0.0639	0.0050
		12	Kamrieng	0.2672	0.0115	0.0627	0.0049
		13	Koas Krala	0.2969	0.0123	0.0730	0.0052
		14	Moung Ruessei	0.3133	0.0142	0.0811	0.0065
3	Kampong Cham	1	Batheay	0.3546	0.0083	0.0995	0.0043
		2	Chamkar Leu	0.3305	0.0070	0.0879	0.0034
		3	Cheung Prey	0.3459	0.0089	0.0949	0.0044

	4 Dambae	0.3501	0.0102	0.0989	0.0051
	5 Kampong Cham	0.1973	0.0108	0.0398	0.0037
	6 Kampong Siem	0.3096	0.0073	0.0789	0.0036
	7 Kang Meas	0.3275	0.0075	0.0862	0.0037
	8 Kaoh Soutin	0.2956	0.0089	0.0738	0.0041
	9 Krouch Chhmar	0.3060	0.0098	0.0790	0.0041
	10 Memot	0.3464	0.0073	0.0966	0.0039
	11 Ou Reang Ov	0.3190	0.0076	0.0838	0.0036
	12 Ponhea Kraek	0.3314	0.0077	0.0896	0.0037
	13 Prey Chhor	0.3241	0.0068	0.0857	0.0035
	14 Srei Santhor	0.3234	0.0082	0.0855	0.0039
	15 Stueng Trang	0.3297	0.0069	0.0897	0.0034
	16 Tboung Khmum	0.3129	0.0078	0.0817	0.0034
	17 Tboung Khmum	0.2681	0.0096	0.0651	0.0042
4 Kampong Chhnang	1 Baribour	0.3263	0.0081	0.0854	0.0044
	2 Chol Kiri	0.3502	0.0128	0.0923	0.0062
	3 Kampong Chhnang	0.2372	0.0122	0.0529	0.0048
	4 Kampong Leaeng	0.3656	0.0102	0.1040	0.0051
	5 Kampong Tralach	0.3286	0.0079	0.0886	0.0037
	6 Rolea B'ier	0.3172	0.0074	0.0810	0.0035
	7 Sameakki Mean Chey	0.3498	0.0082	0.0967	0.0041
	8 Tuek Phos	0.3509	0.0097	0.0977	0.0049
5 Kampong Speu	1 Basedth	0.3543	0.0085	0.0987	0.0040
	2 Chbar Mon	0.2363	0.0105	0.0519	0.0042
	3 Kong Pisei	0.3217	0.0079	0.0845	0.0037
	4 Aoral	0.3652	0.0133	0.1051	0.0072
	5 Odongk	0.3215	0.0072	0.0839	0.0033
	6 Phnum Sruoch	0.3402	0.0085	0.0928	0.0042
	7 Samraong Tong	0.3090	0.0070	0.0788	0.0033
	8 Thpong	0.3498	0.0098	0.0963	0.0050
6 Kampong Thom	1 Baray	0.3456	0.0069	0.0960	0.0035
	2 Kampong Svay	0.3506	0.0078	0.0976	0.0040
	3 Stueng Saen	0.2687	0.0080	0.0654	0.0039
	4 Prasat Ballangk	0.3969	0.0129	0.1214	0.0075

	5 Prasat Sambour	0.3634	0.0095	0.1057	0.0051
	6 Sandan	0.3850	0.0105	0.1158	0.0062
	7 Santuk	0.3507	0.0083	0.0982	0.0043
	8 Stoung	0.3792	0.0089	0.1116	0.0049
7 Kampot	1 Angkor Chey	0.3135	0.0086	0.0803	0.0043
	2 Banteay Meas	0.3303	0.0079	0.0889	0.0036
	3 Chhuk	0.3376	0.0081	0.0922	0.0040
	4 Chum Kiri	0.3578	0.0096	0.1019	0.0052
	5 Dang Tong	0.3452	0.0088	0.0954	0.0045
	6 Kampong Trach	0.3463	0.0077	0.0962	0.0042
	7 Kampot	0.3288	0.0075	0.0875	0.0039
	8 Kampong Bay	0.2243	0.0103	0.0499	0.0043
8 Kandal	1 Kandal Stueng	0.2888	0.0071	0.0701	0.0034
	2 Kien Svay	0.2736	0.0072	0.0659	0.0028
	3 Khsach Kandal	0.3102	0.0076	0.0794	0.0033
	4 Kaoh Thum	0.3056	0.0076	0.0772	0.0035
	5 Leuk Daek	0.3090	0.0094	0.0797	0.0042
	6 Lvea Aem	0.2968	0.0076	0.0726	0.0034
	7 Mukh Kampul	0.2909	0.0072	0.0716	0.0031
	8 Angk Snuol	0.2752	0.0065	0.0659	0.0029
	9 Popnhea Lueu	0.2902	0.0064	0.0714	0.0029
	10 S'ang	0.3007	0.0064	0.0756	0.0028
	11 Ta Khmau	0.1882	0.0084	0.0364	0.0028
9 Koh Kong	1 Botum Sakor	0.3305	0.0131	0.0923	0.0074
	2 Kiri Sakor	0.2904	0.0182	0.0702	0.0087
	3 Kaoh Kong	0.2870	0.0180	0.0729	0.0084
	4 Smach Mean Chey	0.2311	0.0125	0.0504	0.0052
	5 Mondol Seima	0.2477	0.0164	0.0560	0.0067
	6 Srae Ambel	0.3402	0.0098	0.0944	0.0053
	7 Thma Bang	0.3406	0.0193	0.0965	0.0108
10 Kratie	1 Chhloung	0.2972	0.0105	0.0767	0.0046
	2 Kracheh	0.2364	0.0119	0.0525	0.0046
	3 Preaek Prasab	0.3145	0.0095	0.0805	0.0043
	4 Sambour	0.3664	0.0108	0.1036	0.0056

	5 Snuol	0.3260	0.0103	0.0863	0.0049
	6 Kracheh	0.3565	0.0099	0.0992	0.0051
11 Mondul Kiri	1 Kaev Seima	0.3023	0.0165	0.0785	0.0081
	2 Kaoh Nheaek	0.3508	0.0187	0.0964	0.0090
	3 Ou Reang	0.3018	0.0283	0.0763	0.0134
	4 Pech Chreada	0.2988	0.0235	0.0760	0.0111
	5 Saen Monourom	0.2587	0.0172	0.0620	0.0079
12 Phnom Penh	1 Chamkar Mon	0.1473	0.0121	0.0261	0.0037
	2 Doun Penh	0.1575	0.0141	0.0277	0.0046
	3 Prampir Meakkakra	0.1387	0.0139	0.0198	0.0039
	4 Tuol Kouk	0.1467	0.0125	0.0252	0.0038
	5 Dangkao	0.2272	0.0173	0.0530	0.0069
	6 Mean Chey	0.1953	0.0154	0.0404	0.0055
	7 Ruessei Kaev	0.2065	0.0163	0.0434	0.0061
	8 Ruessei Kaev	0.1774	0.0135	0.0361	0.0047
13 Preah Vihear	1 Chey Saen	0.3495	0.0187	0.0999	0.0096
	2 Chhaeb	0.3813	0.0168	0.1130	0.0090
	3 Choam Khsant	0.3345	0.0166	0.0948	0.0083
	4 Kuleaen	0.3709	0.0154	0.1095	0.0092
	5 Rovieng	0.3731	0.0121	0.1065	0.0067
	6 Sangkom Thmei	0.3650	0.0142	0.1043	0.0073
	7 Tbaeng Mean Chey	0.3516	0.0201	0.0963	0.0100
	8 Tbaeng Mean Chey	0.2478	0.0118	0.0560	0.0051
14 Prey Veng	1 Ba Phnum	0.3264	0.0090	0.0852	0.0044
	2 Kamchay Mear	0.3313	0.0087	0.0891	0.0043
	3 Kampong Trabaek	0.3340	0.0080	0.0894	0.0040
	4 Kanhchriech	0.3270	0.0093	0.0858	0.0043
	5 Me Sang	0.3426	0.0081	0.0938	0.0040
	6 Peam Chor	0.3278	0.0081	0.0871	0.0038
	7 Peam Ro	0.3003	0.0074	0.0754	0.0039
	8 Pea Reang	0.3279	0.0079	0.0868	0.0039
	9 Preah Sdach	0.3453	0.0076	0.0952	0.0038
	10 Kampong Leav	0.2625	0.0108	0.0628	0.0052
	11 Kampong Leav	0.3210	0.0117	0.0832	0.0059

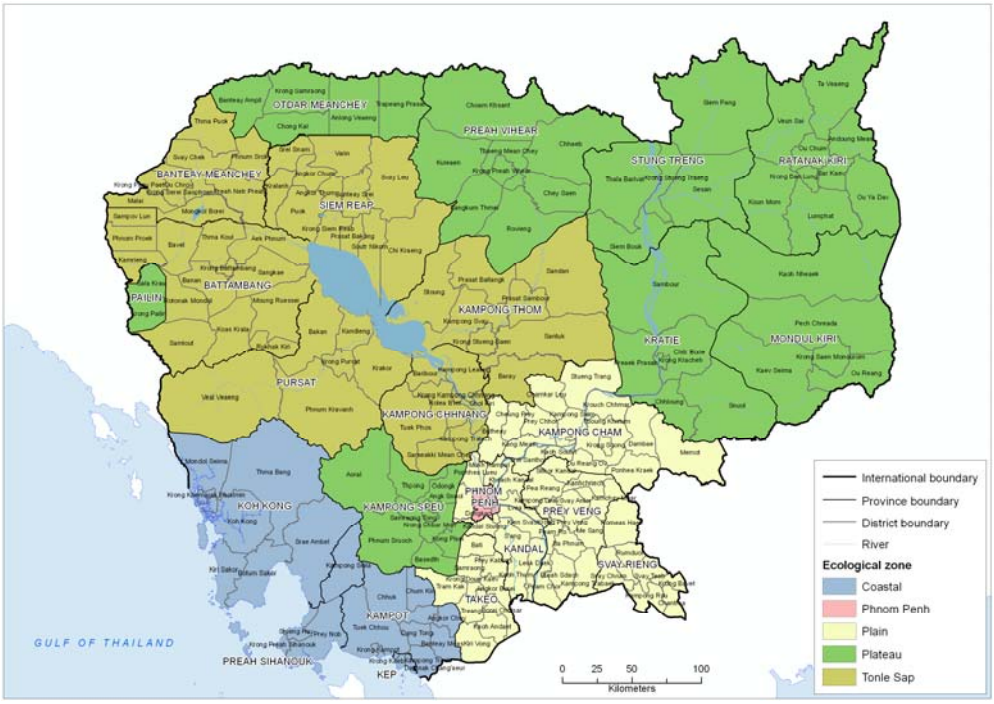
	12 Sithor Kandal	0.3225	0.0107	0.0838	0.0049
	13 Prey Veaeng	0.3390	0.0082	0.0926	0.0040
15 Pursat	1 Bakan	0.3324	0.0070	0.0889	0.0034
	2 Kandieng	0.3366	0.0081	0.0912	0.0038
	3 Krakor	0.3367	0.0083	0.0914	0.0041
	4 Phnum Kravanh	0.3366	0.0088	0.0911	0.0046
	5 Sampov Meas	0.2746	0.0077	0.0674	0.0036
	6 Veal Veaeng	0.2926	0.0165	0.0718	0.0076
16 Ratanak Kiri	1 Andoung Meas	0.3837	0.0286	0.1117	0.0152
	2 Ban Lung	0.2407	0.0117	0.0546	0.0051
	3 Bar Kaev	0.3428	0.0237	0.0963	0.0121
	4 Koun Mom	0.3395	0.0163	0.0934	0.0083
	5 Lumphat	0.3623	0.0179	0.0981	0.0093
	6 Ou Chum	0.3374	0.0277	0.0923	0.0131
	7 Ou Ya Dav	0.3593	0.0261	0.1017	0.0132
	8 Ta Veaeng	0.3417	0.0301	0.0922	0.0157
	9 Veun Sai	0.3911	0.0228	0.1128	0.0117
17 Siemreap	1 Angkor Chum	0.4106	0.0122	0.1287	0.0072
	2 Angkor Thum	0.4305	0.0154	0.1392	0.0089
	3 Banteay Srei	0.3776	0.0128	0.1120	0.0069
	4 Chi Kraeng	0.3780	0.0081	0.1116	0.0044
	6 Kralanh	0.3870	0.0100	0.1181	0.0058
	7 Puok	0.3685	0.0087	0.1087	0.0048
	9 Prasat Bakong	0.3491	0.0091	0.0968	0.0045
	10 Siem Reab	0.2619	0.0075	0.0628	0.0030
	11 Soutr Nikom	0.3510	0.0076	0.0987	0.0040
	12 Srei Snam	0.3999	0.0121	0.1230	0.0066
	13 Svay Leu	0.3703	0.0134	0.1075	0.0074
	14 Varin	0.3883	0.0126	0.1179	0.0066
18 Sihanoukville	1 Mittapheap	0.2080	0.0097	0.0423	0.0033
	2 Prey Nob	0.3208	0.0077	0.0829	0.0036
	3 Stueng Hav	0.2686	0.0145	0.0619	0.0063
	4 Kampong Seila	0.3030	0.0141	0.0786	0.0069
19 Stung Treng	1 Sesan	0.3621	0.0158	0.1019	0.0081

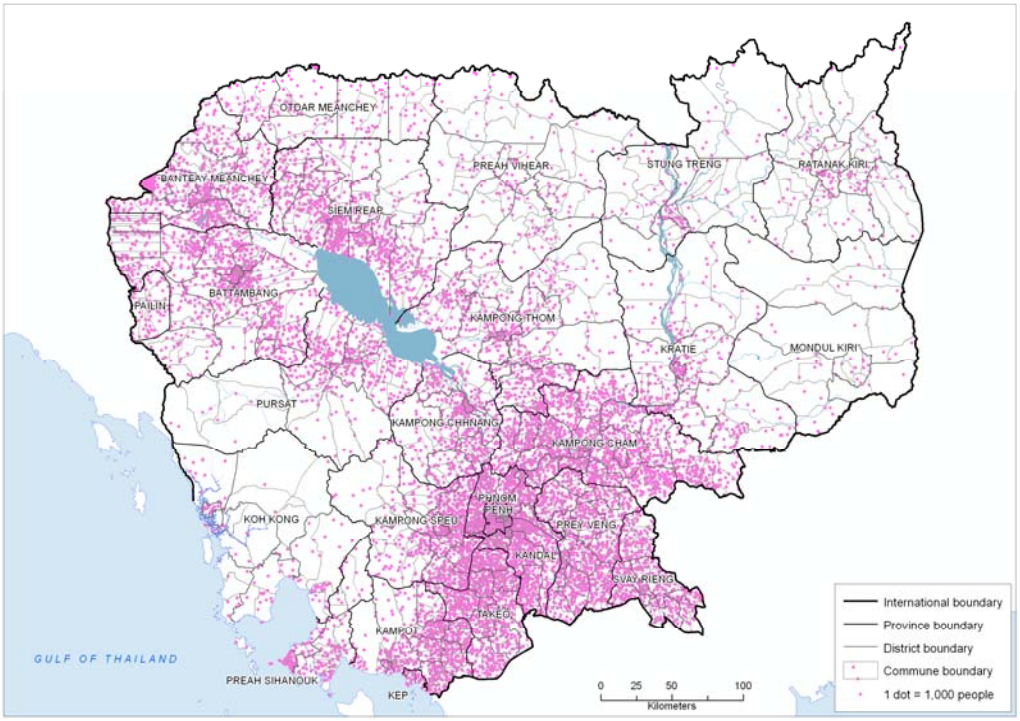
	2 Siem Bouk	0.3349	0.0136	0.0888	0.0064
	3 Siem Pang	0.4308	0.0184	0.1398	0.0106
	4 Stueng Traeng	0.2532	0.0109	0.0584	0.0049
	5 Thala Barivat	0.4027	0.0124	0.1255	0.0071
20 Svay Rieng	1 Chantrea	0.3229	0.0115	0.0833	0.0056
	2 Kampong Rou	0.3374	0.0104	0.0892	0.0052
	3 Rumduol	0.3432	0.0108	0.0922	0.0053
	4 Romeas Haek	0.3479	0.0078	0.0962	0.0039
	5 Svay Chrum	0.3354	0.0070	0.0902	0.0034
	6 Svay Rieng	0.2559	0.0089	0.0603	0.0041
	7 Svay Teab	0.3280	0.0099	0.0853	0.0051
	8 Chantrea	0.2818	0.0104	0.0677	0.0046
21 Takeo	1 Angkor Borei	0.3204	0.0099	0.0824	0.0045
	2 Bati	0.3222	0.0071	0.0851	0.0033
	3 Borei Cholsar	0.3209	0.0116	0.0812	0.0055
	4 Kiri Vong	0.3300	0.0081	0.0873	0.0038
	5 Kaoh Andaet	0.3264	0.0095	0.0872	0.0045
	6 Prey Kabbas	0.3135	0.0071	0.0811	0.0036
	7 Samraong	0.3315	0.0072	0.0888	0.0037
	8 Doun Kaev	0.2805	0.0085	0.0683	0.0042
	9 Tram Kak	0.3321	0.0070	0.0888	0.0032
	10 Treang	0.3319	0.0077	0.0877	0.0036
22 Oddar Meanchey	1 Anlong Veaeng	0.2724	0.0119	0.0664	0.0050
	2 Banteay Ampil	0.3329	0.0141	0.0899	0.0066
	3 Chong Kal	0.3293	0.0152	0.0890	0.0074
	4 Samraong	0.2914	0.0105	0.0749	0.0051
	5 Trapeang Prasat	0.2785	0.0149	0.0673	0.0061
23 Kep	1 Damnak Chang'aeur	0.3542	0.0130	0.0977	0.0066
	2 Kaeb	0.2945	0.0160	0.0737	0.0082
24 Pailin	1 Pailin	0.2303	0.0111	0.0504	0.0043
	2 Sala Krau	0.2603	0.0133	0.0589	0.0053

Appendix D. Maps

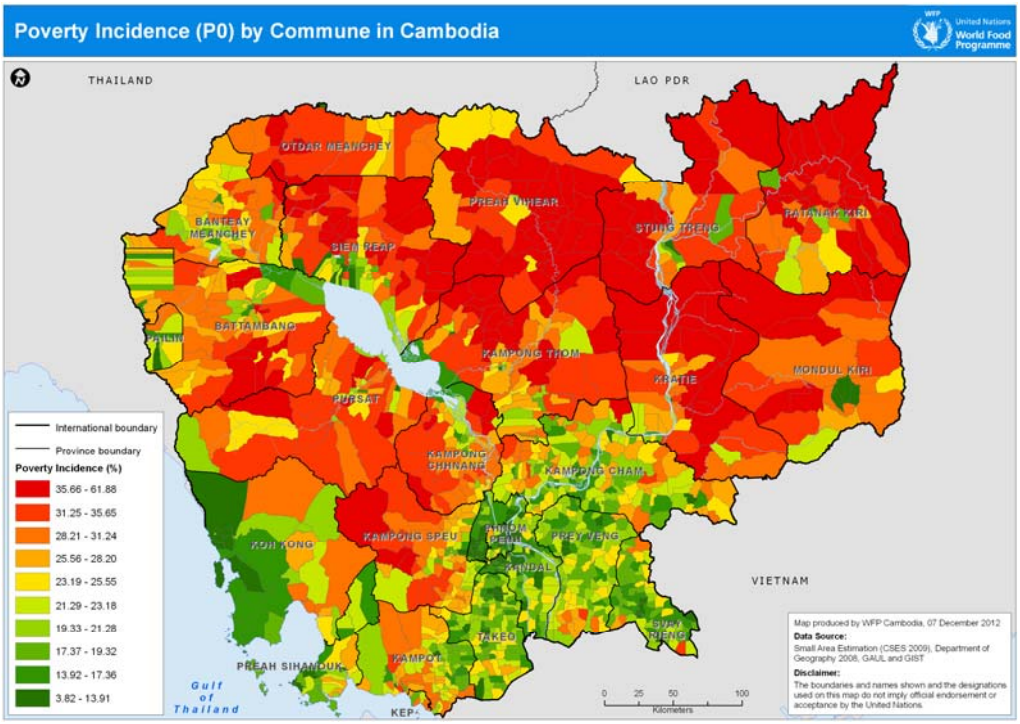
Appendix D.0.

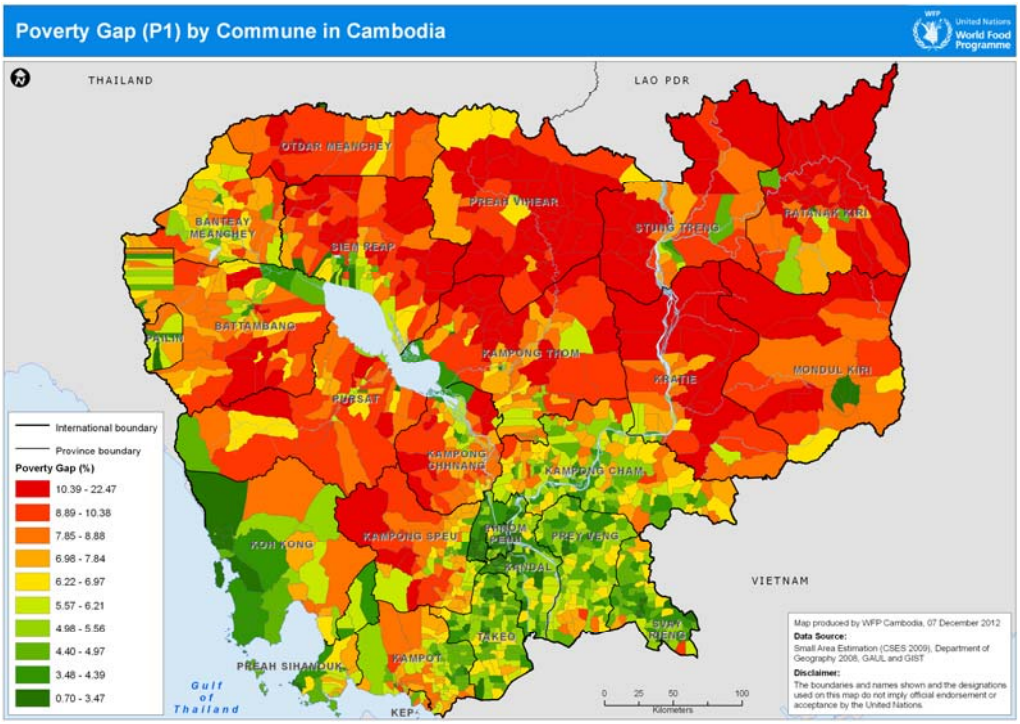
Maps of the administrative units including commune boundaries, ecological zones, and population density.

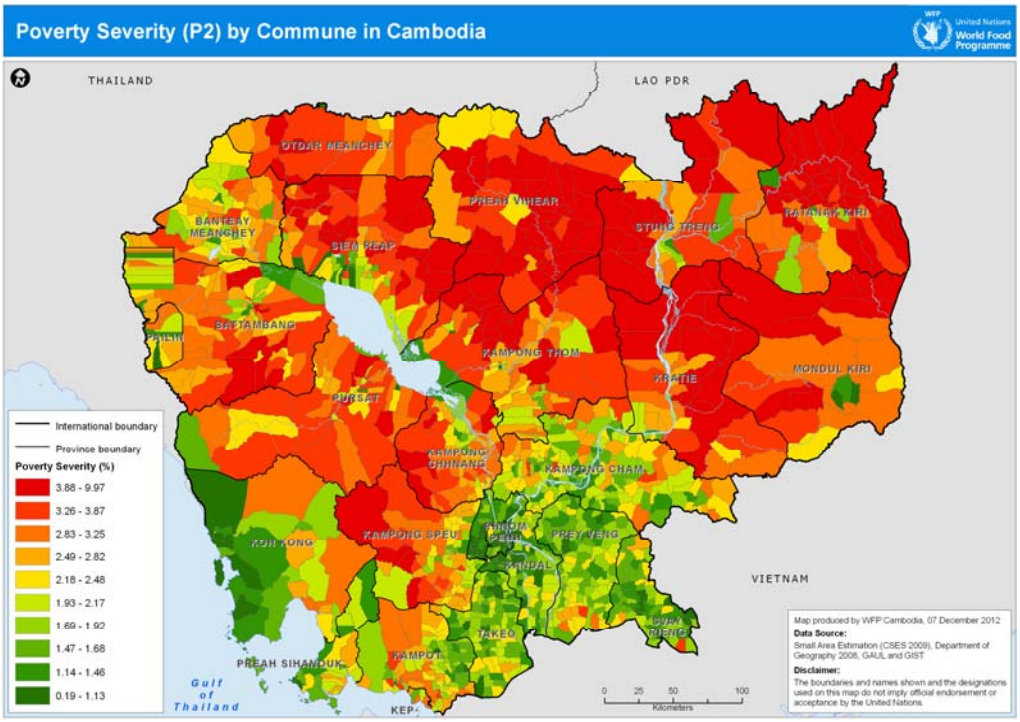



OTDAR MEANCHEY
BANTEAY MEANCHEY
SIEM REAP
PREAH VIHEAR
STUNG TRENG
RATANAK KIRI
BATTAMBANG
PAILIN
KAMPONG THOM
KRATIE
MONDUL KIRI
PURSAT
KAMPONG CHHNANG
KAMPONG CHAM
PHNOM PENH
KOH KONG
KAMPONG SPEU
PREY VENG
KANDAL
SVAY RIENG
TAKEO
KAMPOT
GULF OF THAILAND
PREAH SIHANOUK
KEP
0 25 50 100
Kilometers
International boundary
Province boundary
District boundary
Commune boundary
1 dot = 1,000 people

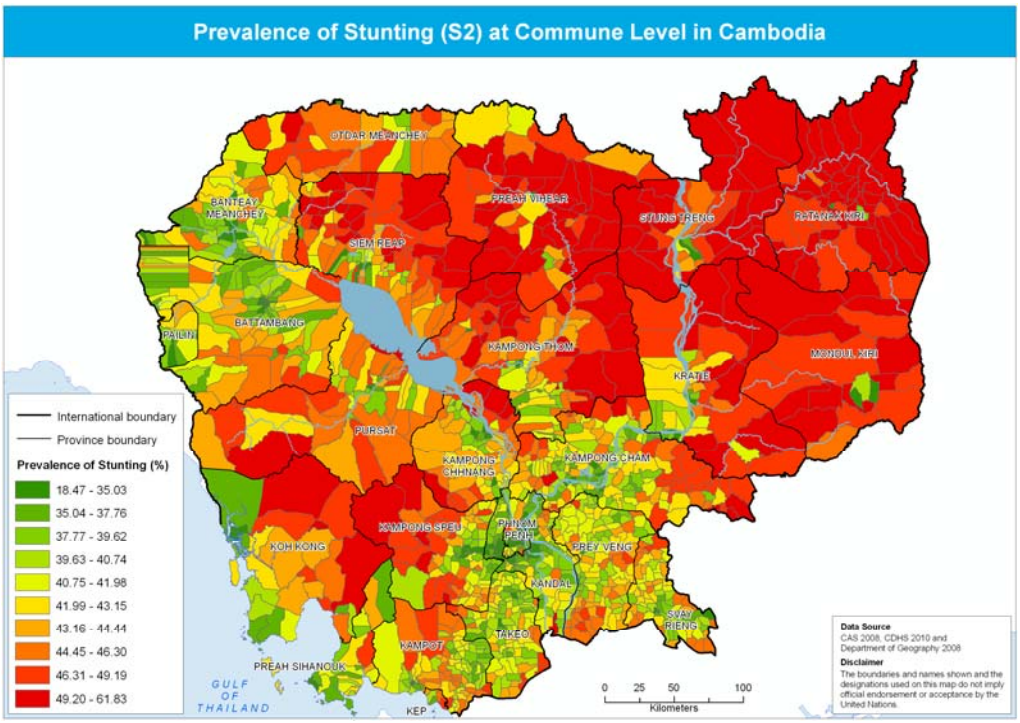
Appendix D.1.
Poverty incidence, gap and severity at commune level.

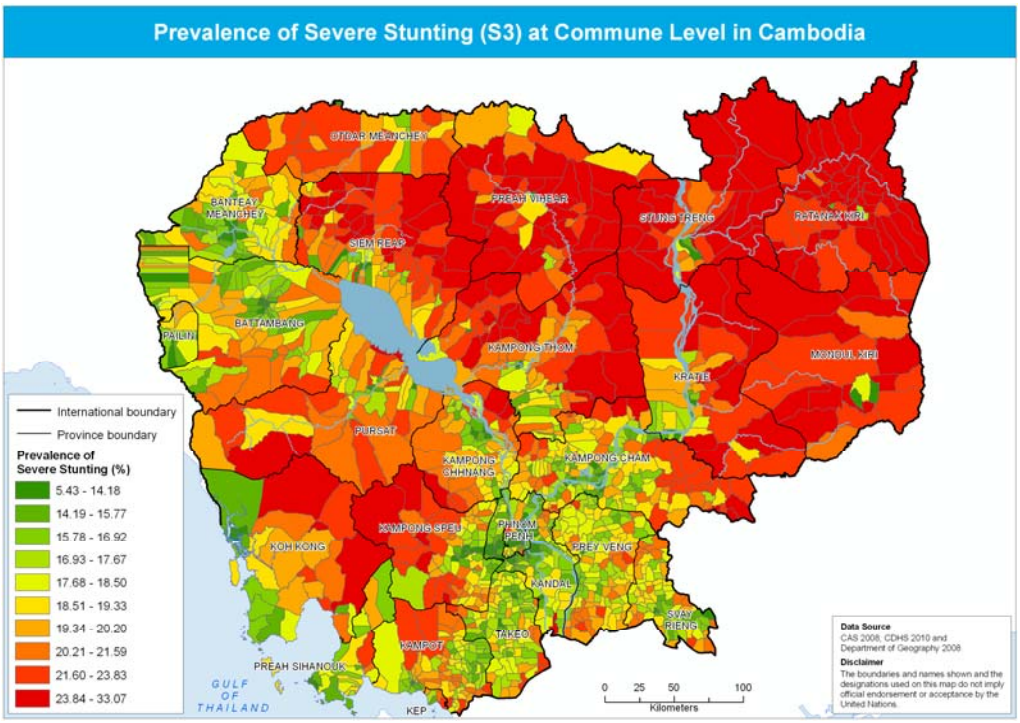



Poverty Gap (P1) by Commune in Cambodia
United Nations World Food Programme
THAILAND
LAO PDR
VIETNAM
OTDAR MEANCHEY
BANTEAY MEANCHEY
SIEM REAP
PREAH VIHEAR
STUNG TRENG
RATANAK KIRI
BATTAMBANG
PAILIN
KAMPONG THOM
KRATIE
MONDUL KIRI
PURSAT
KAMPONG CHHNANG
KAMPONG CHAM
KOH KONG
KAMPONG SPEU
PHNOM PENH
PREY VENG
KANDAL
SVAY RIENG
TAKEO
KAMPOT
PREAH SIHANOUK
KEP
Gulf of Thailand
International boundary
Province boundary
Poverty Gap (%)
10.39 - 22.47
8.89 - 10.38
7.85 - 8.88
6.98 - 7.84
6.22 - 6.97
5.57 - 6.21
4.98 - 5.56
4.40 - 4.97
3.48 - 4.39
0.70 - 3.47
0 25 50 100
Kilometers
Map produced by WFP Cambodia, 07 December 2012
Data Source:
Small Area Estimation (CSES 2009), Department of Geography 2008, GAUL and GIST
Disclaimer:
The boundaries and names shown and the designations used on this map do not imply official endorsement or acceptance by the United Nations.


Poverty Severity (P2) by Commune in Cambodia
United Nations World Food Programme
THAILAND
LAO PDR
VIETNAM
OTDAR MEANCHEY
BANTEAY MEANCHEY
SIEM REAP
PREAH VIHEAR
STUNG TRENG
RATANAK KIRI
BATTAMBANG
PAILIN
KAMPONG THOM
KRATIE
MONDUL KIRI
PURSAT
KAMPONG CHHNANG
KAMPONG CHAM
KOH KONG
KAMPONG SPEU
PHNOM PENH
PREY VENG
KANDAL
SVAY RIENG
TAKEO
KAMPOT
PREAH SIHANOUK
KEP
Gulf of Thailand
International boundary
Province boundary
Poverty Severity (%)
3.88 - 9.97
3.26 - 3.87
2.83 - 3.25
2.49 - 2.82
2.18 - 2.48
1.93 - 2.17
1.69 - 1.92
1.47 - 1.68
1.14 - 1.46
0.19 - 1.13
0 25 50 100
Kilometers
Map produced by WFP Cambodia, 07 December 2012
Data Source:
Small Area Estimation (CSES 2009), Department of Geography 2008, GAUL and GIST
Disclaimer:
The boundaries and names shown and the designations used on this map do not imply official endorsement or acceptance by the United Nations.

Appendix D.2.
Maps of stunting incidence and severe stunting.





Appendix D.3.
Maps of underweight and severe underweight.