Potential for Small Area Estimation and Poverty Mapping at District and Commune Level in Cambodia

Feasibility Report
Phases 1, 2 and 3

Professor Stephen Haslett
Dr Geoffrey Jones
Massey University
New Zealand

20 October 2008

Summary

The general conclusion of this report is that updated small area estimation of poverty incidence, gap, and severity for Cambodia seems feasible at commune level.

For kilocalories, the information at household level necessary for assessment was not available by the time statistical analysis was completed, so this report is not able to comment directly on feasibility of small area estimation for kilocalories.

Should small area estimation of poverty proceed beyond the current feasibility stage however, and kilocalories information become available, assessment of (and if possible estimates for) small area estimation of kilocalories should be included along with small area estimation of poverty incidence, gap and severity as part of the extended study.

Executive Summary

1. Small area estimation is a mathematical technique to extract more detailed information from existing data sources by statistical modelling. The methodology is important because it produces finer level information than is possible for a sample survey analysed by standard methods, for poverty related variables that are not collected in the census. The cost of small area studies can be saved many times over by having this better poverty information at a finer level for use in aid allocation.

2. This report provides a brief summary of the training phase (Phase 1) and details of the two non-training phases (Phases 2 and 3) undertaken by staff from Massey University, New Zealand, to assess feasibility for small area estimation of poverty and poverty mapping in Cambodia at district or commune level. This report includes and hence supersedes the earlier Phase 1 and 2 report.

3. The particular aspects of poverty that are considered are poverty incidence, gap and severity relative to the national poverty line, with an additional focus on kilocalorie consumption. While the last of these four strongly reflects the food security interests and concerns of the sponsor of the feasibility study, which is the World Food Programme, this feasibility report also recognises the importance to a wide range of international aid agencies of all four measures, and the centrality of the first three to sound economic measurement of poverty.

4. For Cambodia, the data sources considered are the population census (which has been conducted in 1998 and 2008) and the Cambodia Socio-economic Survey (CSES) which was conducted in 2003/4, on a smaller scale in 2007, and which will be conducted again as a full scale survey in 2009. Use of the Seila commune database, more recently known as the National Committee for Decentralisation and Deconcentration (NCDD) database, is also discussed.

5. From a statistical modelling point of view, optimal timing of any full scale, small area estimation of poverty and/or poverty mapping would be after the 2008 census information and the 2009 CSES have both been processed, since the 2009 CSES will have a larger sample size than the 2007 CSES (approximately 15,000 versus 3,600). The additional sample size in 2009 will mean more accurate small area estimates, and use of the 2009 rather than the 2007 CSES is more likely to provide sound small area estimates at commune rather than district level.

6. From a policy and planning point of view however, the situation may be reversed. The current Cambodia *National Strategic Development Plan, 2006-2010* (Ministry of Planning, 2005) is scheduled for revision before 2010. On this timetable, small area estimates, even if at district rather than commune level, based on the 2007 CSES and the 2008 census would be better for timely planning purposes, since even the full-scale 2009 CSES is unlikely to provide sound poverty estimates below regional level without the use of small area estimation techniques.

7. A poverty mapping exercise in 2009, using the 2008 census (after the detailed household and personal level data are computerised and edited) and 2007 CSES data would provide a sound basis for a more extensive and detailed poverty mapping once the 2009 CSES data was available. Statistical modelling using the

2007 CSES would be a very useful starting point for further poverty mapping using the 2009 CSES and the 2008 census.

8. Some additional care is needed if using the 2007 CSES (rather than the 2009 CSES) and the 2008 census data for small area estimates, since the 2007 CSES and the 2008 census will use commune and district boundaries and coding that differ more than those for the 2009 CSES and the 2008 census. The boundaries will be revised with the 2008 census but not retrospectively. Tight control and recording of all such boundary changes, and care in coding region, district, commune and preferably village for every household in the census is essential.

9. Despite these cautions, on balance, running the statistical modelling in 2008/9 using the 2008 census and the 2007 CSES to get small area estimates of poverty and poverty maps would be a useful and timely planning option.

10. This feasibility report includes preliminary statistical model fitting using the 2003/4 CSES, which is Phase 3 of the WFP Massey University contract. Had the 2007 CSES data been available before the completion of the analysis for this study, statistical models for the 2003/4 CSES could have been cross-validated by fitting models to the 2003/4 data to check whether their optimality properties are stable over time using the 2007 CSES data. However even without this cross-validation, the analyses in this report provide a detailed statistical assessment of poverty mapping feasibility.

11. The general conclusion of this report is that small area estimation of poverty incidence, gap and severity for Cambodia seems feasible at commune level. For kilocalories, the information necessary for assessment of feasibility is not yet available, so this report is not able to provide detailed comment about feasibility of small area estimation for kilocalories. Should small area estimation of poverty proceed beyond feasibility stage however, and kilocalories information be available, assessment of (and if possible estimates for) small area estimation of kilocalories should be included as part of the extended study.

12. No poverty estimates are produced as part of this feasibility study. Such estimates require further funding beyond this feasibility phase, and available clean 2008 census and 2007 (and/or 2009) CSES data including detailed expenditure information, in addition to the 2003/4 CSES data currently available. Food cost and kilocalorie equivalence information for all food types, so that kilocalorie consumption can be estimated for every surveyed household, will also be required as a prerequisite to any small area estimation of kilocalories per capita.

13. The completion of this report 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, and World Bank, and World Food Programme (WFP) - which commissioned this research - in Phnom Penh 21-25 April and 14-16 October 2008. The authors are grateful for these extensive contributions. Viewpoints and opinions expressed in this report do not however necessarily reflect those of all or any of the organisations consulted.

1. Introduction

This assessment is in three phases, which together provide a feasibility assessment of the potential for small area estimation of poverty incidence, gap, and severity, plus kilocalorie consumption in Cambodia. None of the three phases includes provision of small area estimates, however.

Phase One, completed in February / March 2008, involved lecturing at the Geoinformatics Centre, Asian Institute of Technology in Bangkok, Thailand, "Second Training Session on Poverty Mapping". While the lecturing given focused on "finalizing the candidate variables" and "modelling for small area estimation", it also provided general support on methods suitable for use by workshop participants for statistical analysis of county survey and census data. Included was extended discussion with individuals and groups at the workshop, particularly the four people who attended from the Cambodian National Institute of Statistics (NIS) Poverty Mapping Group: Ing Sokun, Nguon Sovann, Saint Lundy, and Sin Serey Vuth.

Phase Two, completed in April and May 2008 involved:

- Analysis of existing research and information on food security necessary for statistical assessment of kilocalorie consumption for small area estimation in Cambodia, based on reference material supplied by the World Food Programme and existing knowledge of small area estimation methods. Note that the nutritional aspects of methods of estimating kilocalories from expenditure data did not form part of this phase.
- Analysis of relevant Cambodia questionnaires (in English) from the population censuses 1998 and 2008), the national household income and expenditure survey (Socio-economic Survey of Cambodia - CSES), and Comprehensive Food Security and Vulnerability Analysis information, supplied by the World Food Programme.
- Identification and listing of questions asked in census(es), and the CSES survey that *prima facie* are similar enough to be used for small area estimation of poverty incidence, gap and severity, and kilocalorie consumption. This investigation was based on the English versions of questionnaires only, but also involved extensive discussion with Cambodian National Institute of Statistics staff about the original versions of the questionnaires which were administered in Khmer.
- Preparation of an interim report, "Potential for Small Area Estimation and Poverty Mapping at District and Commune Level in Cambodia: Interim Feasibility Report".

Phase Three, completed May to October 2008, involved:

- Identification of variables in both survey(s) and census(es) that are potentially useful for small area estimation of poverty incidence, gap, severity, and kilocalorie consumption, in conjunction with the Cambodian National Institute of Statistics.
- Developing and testing of preliminary statistical regression models (including estimation of variance components) used for poverty incidence, gap and severity, based only on CSES 2003/4 survey data supplied by NIS. Note that assessment for kilocalories was not possible as the required kilocalorie information at household level was not available from CSES 2003/4 for each surveyed household at analysis stage of this feasibility study.
- Identification, in conjunction with the Cambodian National Institute of Statistics, of any Administrative Unit (area) code changes that may complicate any later analysis of the statistical relationship between survey and census data (outside the scope of this study). This was based on survey and census information supplied by NIS through WFP.
- Comment on the potential impact of these statistical analyses on small area estimation of poverty incidence, gap and severity.
- Preparation of this second and final feasibility report.

Completion of Phases 2 and 3 has included consultation and discussion with the following people and organisations:

National Institute of Statistics
H. E. San Sy Than, Director General, 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
Saint Lundy, Deputy Director, General Statistics, 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
Dr 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

United Nations Development Programme / International Fund for Agricultural Development
Ung Dara Rat Moni, IFAD/UNDP Advisor

United Nations World Food Programme
Thomas Keusters, Country Director, Cambodia
Bradley Busetto, Officer-in-Charge, Cambodia
Coco Ushiyama, Deputy Country Director, Cambodia
Kim Ratha, Senior Programme Assistant, Vulnerability Analysis and Mapping (VAM), Cambodia
Michael Sheinkman, Senior Regional Programme Advisor, Vulnerability Analysis and Mapping (VAM), WFP Bureau for Asia, Thailand

World Bank
Dr James Knowles, International Consultant.
Tim Conway, Senior Poverty Specialist
Neak Samsen, Poverty Specialist

Note that, while additional people were approached, the limited period of the visits to Phnom Penh during April and October 2008 meant not all were available. Nevertheless, every effort has been made to ensure a full range of opinion was canvassed and considered in the preparation of both this and the preliminary report. This effort was aided by the uniformity of opinion that there is need for updated small area estimates of poverty in Cambodia based on CSES 2007 and the 2008 census once the census data is available in 2009, and on CSES 2009 and the census once CSES 2009 data is available in 2010.

2. Background

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 poverty 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. 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 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, Haslett and Parajuli, 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. Pinney's method necessarily uses the logarithm of average village consumption, which differs from the variable used at household level by Fujii (2002) since it does not equal average logarithm of household consumption within village. For Pinney's model, higher R^2 than for Fujii's is expected, given village rather than household data are being modelled. Nevertheless, Pinney's model has a remarkably small number of explanatory variables for an $R^2>0.7$, which is very encouraging for developing new models for current CSES 2007 (and CSES 2009) with 2008 census data that are less complicated than Fujii's, especially since Pinney fits a single model for the whole country with regional variation covered by using zone as a fixed effect. Pinney seems to have used zone (which is categorical) ordered by average consumption, and fitted it as a linear effect. Despite some of these aspects limiting comparison and possibly utility, Pinney's analysis raises the interesting possibility that the annual commune database, which collects information from every commune, could be used to update any small area estimates derived from the 2008 census and concurrent (eg 2007 or 2009) survey data from the CSES. After any such 2007/8/9 small area estimates (which are the topic of this feasibility study) have been calculated and their accuracy assessed, use of the NCDD database for annual updating may require bias adjustment, given differences between NCDD data collection procedures and those for the census. However, there is a possible method available for assessing the extent of any such bias, by comparing NCDD database and census information from 2008, when both data sources are collected concurrently.

There is an important series of reports for any small area estimation of poverty project in Cambodia that are published by Statistics Sweden via its International Capacity Building Project at NIS. Perhaps the most relevant of these are Davidsson and Pettersson (2003), Pettersson (2004, 2005), Neupert (2005), Dalén (2005a, 2005b), Johansson (2005, 2006), Dalén (2006), Sandqvist (2007a, 2007b) and Isaksson (2007). The majority of these reports focus on the design and analysis of the CSES, which is crucial information when fitting regression models to this survey data, but others, for example, relate to definitions of expenditure - also a critical issue for expenditure modelling using CSES data.

This sequence of reports is also relevant to estimation of kilocalorie consumption at small area level through definitions and measurement of expenditure, consumer price index (CPI) measurement, and regional variation in food prices, eg Dalén (2005a, 2005b), Johansson. (2005), Sandqvist (2007a, 2007b). Johansson (2005) also raises the important question of differences in estimates of food consumption using diary and recall methods in the CSES, and concludes that for Cambodia, while estimates for some food types are generally similar, others clearly differ. The report by Dalén (2005b) raises issues about prices, including food prices, and provides an assessment of the research of Knowles (2006) who developed a new, alternative set of poverty indicators for Cambodia. Note that the information collected by NIS for the CPI is collected only in Phnom Penh and five other of the twenty provinces, namely: Kampong Cham, Battambang, Siem Reap, Kandal, Sihanoukville. This limits provincial comparison of prices, and will have an influence on the setting of poverty lines, as will current inflationary pressures. Kilocalorie consumption per household was not available for all sampled households from the CSES for 2003/4 or any consequent period at the time of completion of the statistical analysis for this feasibility report. Having such kilocalorie information remains an essential prerequisite to small area estimates of kilocalorie consumption, whether measured on "per person" or "per adult equivalent" basis. The former is measure is simpler since it divides household kilocalories by number of people in each household. The latter would require a more detailed adult equivalence scale by age and sex.

There are other publications related to poverty assessment in Cambodia, but generally these are not relevant to small area poverty estimation methodology, even though their analyses provide useful information on poverty itself. The German Technical Co-operation (GTZ) 2006 publication *Identification of Poor Households Project (IDPoor)* (GTZ, 2006a) outlines a study for at least three provinces, together with questionnaires, to identify poor households. A second GTZ 2006 publication (GTZ 2006b), in conjunction with the Cambodian Ministry of Planning, *Procedures for identification of Poor Households in Cambodia, Version 1 for 2007 Rounds 1a and 1b*, .details a useful but non-statistical mechanism for scoring poverty of households based on questionnaire and other information related to study operations.

There is also useful general background and maps in the National Institute of Statistics, Ministry of Planning; EC-FAO Food Security publication, *Summary Report on the Food Insecurity Assessment in Cambodia: 2003/04 Cambodia Socio-economic Survey*. The figures given there on p6 "Level of food deprivation in Cambodia based on food data from the 2003/4 Cambodian Socio-economic Survey" indicate the South West as the area of greatest food insecurity. This can be compared with Fujii's (2002) report for 1997/8 which indicates that the highest level of expenditure poverty is in an approximately north-south band through the middle of the country, and Conway and Goh (2006) which indicates that the North East is where certain poverty problems are concentrated. Of these only Conway and Goh, and Fujii have apparently contradictory information, although this may reflect the time difference since Fujii used 1998 census and 1997 CSES, and Conway and Goh the 2003/4 CSES. These differences do however suggest revision of Fujii's 1997/8 small area estimates of poverty would be timely. This NIS/FAO report document also uses food tables and expenditure of food items (as will be required for 2007/8/9 small area kilocalorie estimates) and comments on p43, "The ASEAN food composition table supplemented by USDA were used to build the nutrient conversion table of the CSES 2003/4 food items. However, it is important that such a table be reviewed with relevant experts from the health and nutrition institutions for future use". Also of relevance to food security internationally is the development of the Integrated Humanitarian and Food Security Phase Classification (IPC), which is an innovation for improving analysis and decision making on food security outcomes and providing a common approach internationally for food security analysis and response.

The National Institute of Statistics 2006 publication, *A Poverty Mapping Profile of Cambodia, 2004* provides a useful general summary of poverty issues derived in the main from the 2003/4 CSES. It includes maps of poverty incidence by geographical zones, and poverty incidence and gap by province (see pages 49-51, Figs 10.4-10.6).

There is also a useful sequence of National Institute of Statistics publications, not referenced explicitly in this report, but listed and available through NIS, on the Inter-censal Population Survey 2004, reclassification of urban areas 2004, population projections, Cambodian Demographic and Health Survey 2004 and 2005, plus various CD-ROMs and 1998 population and housing census information.

Conway and Goh (2006) is a general World Bank report on poverty in Cambodia. It contains the map mentioned in the previous paragraph, showing that the North East has the most severe consumption poverty – p 37, Fig 3.3 "Standards of living are lowest in the Mountain/Plateau and Tonle Lap Regions." – this figure has a percentage basis, and the report comments in Fig 3.4 "The majority of the poor are found in the Plains and Tonle Sap Provinces". The report gives the 2004 total poverty line "Local definition" in current riel per capita per day as 2351 riel for Phnom Penh, 1952 for other urban, and 1753 for rural (Table 6, p193). P189 discusses the updated food poverty line which is set at 2,100 calories per day. Ch 3 p35 gives a useful summary of the extent and "Nature of poverty in Cambodia, 2004".

There is a further World Bank and Cambodia Development Forum report, *Cambodia - Sharing Growth: Equity and Development Report 2007*. Page 194, Table 1, "Estimation of

wage equations (paid employees)" has a similar regression to that required for small area estimation except using log wages rather than log expenditure. This regression uses age, age squared, head of household, whether married, whether immigrant, work hours in last 7 days, highest educational grade (<class 3, 4-6, 7-9, 10-12, 12 or above), and industry. The report has a strong economic focus; there is little on food *per se* except via agriculture, but the regression model may nevertheless provide clues for variables to include in statistical models for small area estimation of poverty.

The 2002 report, *National Poverty Reduction Strategy 2003-2005*, from the Council for Social Development, Royal Cambodian Government, is a further general document. It discusses Food Insecurity on p26 as part of section 3.1.5 and Nutrition on p94 section 4.3.2 (which contains a useful list of cost-effective interventions). There is comment p156, section 6.3 Poverty Impact Evaluation that, "…this initial NPRS does not include a formal programme of quantitative impact evaluation", so the report does not provide methodological detail for small area estimation *per se.*

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".

The Ministry of Health, Cambodia / World Food Programme / Measure DHS+ - ORC Macro 2003 report *Micro-level Estimation of the Prevalence of Stunting and Underweight Among Children in Cambodia* 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 importance of these two WFP-linked reports also warrants 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, Haslett and Parajuli, 2006), provides a detailed perspective on the spatial distribution of poverty. Other variables are also important however (eg health information, rainfall, and other Geographical Information System (GIS) data), even if these cannot be produced as such a fine level. For most users of this information, an atlas of maps is much more useful than a detailed technical report on small area estimation methodology, even if it 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 utility of the 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.

The December 2005**,** *National Strategic Development Plan, 2006-2010*, published by the Cambodian Ministry of Planning, is an overarching document outlining a national strategy for development in Cambodia. It contains a section on Monitoring and Evaluation – see Chapter 6. The principal importance of this document to small area estimation of poverty is however timing. The next *National Strategic Development Plan* is scheduled to be developed and published by 2011, and, as noted strongly by Ministry of Planning staff during our joint MoP/WFP meeting in April 2008, would best be informed by a small area estimation of poverty study. To meet this timetable, poverty estimates using small area techniques would need to be completed in 2009. (Since the 2008 census data are to be available in late 2008, but the next, large scale CSES is to be undertaken in 2009, and the data from CSES 2009 will not be available in time for the next scheduled *National Strategic Development Plan.*.) Although it has a smaller sample size so that small area estimates will not be so accurate, and additional work will be required to link survey and census area codes, a viable alternative is to use the 2007 CSES rather than the 2009 CSES for a small area estimation project that could be begun once 2008 census and detailed expenditure information from the 2007 CSES are released. Both data sources should be available in late 2008, so that it is imperative on the present planning schedule that any small area estimation of poverty programme for Cambodia be underway as soon as possible after this date. For this reason, establishing a suitable and available funding source for small area estimation beyond the current feasibility phase must be given high priority now, if results are to be ready in time for the next scheduled National Strategic Development planning round.

Even without this planning imperative however, using the 2007 CSES and the 2008 census for small area estimates of poverty would have other benefits. It would provide updated poverty maps for 2007/8, it would later provide the first of two snapshots in time at a fine geographical level for a range of poverty indicators when the 2009 CSES was used for small area estimation later, and (from a technical point of view) it would strengthen the basis for that later small area analysis using the 2009 CSES.

3. Requirements for and feasibility of small area estimation of poverty in Cambodia

This report not only provides the background material in the previous two sections. In this and consequent sections it also covers all the assessment phases undertaken by the professional statisticians from Massey University, New Zealand to assess feasibility for updated small area estimation of poverty and poverty mapping in Cambodia at district or commune level. It includes preliminary statistical models fitted to Cambodia Socio-economic Survey (CSES) data.

The particular aspects of poverty considered in both this final feasibility report and in the earlier preliminary report are poverty incidence, gap and severity relative to the national poverty line, with an additional focus on kilocalorie consumption. While the last of these four strongly reflects the food security interests and concerns of the sponsor of the feasibility study, which is the World Food Programme, the feasibility reports also recognise the importance to a wide range international aid agencies of all four measures, and the centrality of the first three to sound economic measurement of poverty.

No poverty estimates are produced or were intended to be produced in this feasibility study. Such estimates would require further research funding, available clean 2008 census data, and 2007 (and/or 2009) CSES data including detailed expenditure information, in addition to the 2003/4 CSES data currently available. Complete linking of area codes used in survey and census would be essential. Food cost and kilocalorie equivalence information by food type will also be required. The aim of the feasibility study is instead to find the variables that *prima face* are sufficiently similar that they have clear potential to be used as the explanatory part of a statistical model for the logarithm of average per capita household expenditure based on CSES data. Such a model is required after the feasibility phase to predict household expenditure for each and every household in the census based on that household's measured characteristics, leading to small area estimates of poverty incidence, gap and severity, plus kilocalorie consumption given the necessary kilocalorie information. The 2003/4 CSES data has been used for this initial modelling. The 2007 CSES data including expenditure was not available before completion of the analysis for this feasibility study, but in principle this can be used to cross-validate the model fitted to the 2003/4 data to assess the extent of any model changes over time. Note that the modelling used in the feasibility study focuses on logarithm of per capita expenditure, since it is this model from which poverty incidence, gap and severity are derived, and it is food consumption data and kilocalorie content for each food item that potentially provides the required household average kilocalorie consumption per capita.

Explicit details of this statistical modelling are included in Sections 4 to 6 below. In this section, the focus is limited to establishing whether variables collected in the census are also collected in the CSES, and if so, exactly how the relevant questions are structured in both data sources. Ideally, exactly the same question with exactly the same categories is asked in the survey and census, but in practice internationally there are usually differences. For Cambodia these differences are given in detail in the material detailed for each potential matching variable (ie question) in the Appendix at the end of this report in Tables A1-A3 and B1-B12.

The feasibility study does not complete the matching of variables between census and survey, since this would also require the 2008 census data and the 2007 and/or 2009 CSES data. What is provided in these "matching tables" A1-A3 and B1-B12 however, taken in conjunction with the preliminary modelling based on CSES data alone, is the best available indication of

whether small area estimation is likely to work for Cambodia. Given that the conclusion is positive, funding should be sought immediately from a variety of international donors for a full scale small area estimation study in Cambodia based on the 2008 census and contemporaneous CSES data.

For Cambodia, the data sources considered in the feasibility study are the population census (which has been conducted in 1998 and 2008) and the Cambodia Socio-economic Survey (CSES) which was conducted in 2003/4, on a smaller scale in 2007, and which will be conducted again as a full scale survey in 2009. The CSES questionnaires in 2003/4 and 2009 will be very similar. The 2007 CSES questionnaire is shorter, and the sample smaller, than for 2003/4 or 2009. The comparison table A1-A3 and B1-B12 are based on the CSES and 2008 census questionnaires.

An additional possible data source for small area estimation is the Seila commune database (more recently known as the National Committee for Decentralisation and Deconcentration (NCDD) database). Possible uses for this village-level data have been outlined in the background section above.

From a statistical modelling point of view, optimal timing of any full scale, small area estimation of poverty and/or poverty mapping would be after both the 2008 census information and the 2009 CSES have both been processed, since the 2009 CSES will have a larger sample size than the 2007 CSES (approximately 15,000 versus 3,600). The additional sample size in 2009 will provide the more accurate small area estimates, and use of the 2009 rather than the 2007 CSES is more likely to provide sound small area estimates at commune rather than district level.

From a policy and planning point of view however, the situation may be reversed. The current Cambodia *National Strategic Development Plan, 2006-2010* (Ministry of Planning, 2005) is scheduled for revision before 2010, so on this timetable small area estimates, even if at district rather than commune level, would still be better for timely planning purposes. Note that even the full-scale 2009 CSES is unlikely to provide sound poverty estimates below regional level without the use of small area estimation techniques.

The use of the smaller 2007 CSES dataset does however raise the concern that this feasibility study, based as it is on the larger 2003/4 CSES dataset, may overstate the likely accuracy of small area estimates based on 2007 CSES and 2008 census data. This is a possibility since the statistical analysis in the feasibility study can only provide a guideline. A balancing consideration however is that if the poverty mapping exercise were to be undertaken in 2009 , using the 2008 census and 2007 CSES data,(once the detailed household and personal level census data are computerised and edited), this would provide an earlier update of Fujii's 1998 small area estimates, two snapshots in time rather than one, and, a sound basis for a more extensive and detailed poverty mapping, once the 2009 CSES data was available. The statistical modelling using the 2007 CSES would be a very useful starting point for poverty mapping using the 2009 CSES and the 2008 census.

In the longer term, there would be considerable policy planning benefit from aligning the year in which census and the larger CSES are carried out so that data collection for these both occur two to three years before the following planning round is to be implemented.

Using the 2007 CSES and the 2008 population and housing census for small area estimation, with additional international donor funding and based on this feasibility study, requires some caution however. The 2007 CSES and the 2008 census use commune and district boundaries and coding that differ more than these will for the 2009 CSES and the 2008 census, since the boundaries will be revised with the 2008 census but not retrospectively. Tight control and recording of all such boundary changes, and care in coding region, district and commune for

every household in the census by NIS, and the necessary resources during data input and processing to make this possible, are essential.

Despite these cautions, on balance, running the poverty mapping in 2008/9 using the 2008 census and the 2007 CSES would seem to be a useful and timely planning option.

The resources required for the full poverty mapping exercise after the feasibility phases, whether based on CSES 2007 and/or CSES 2009, plus the 2008 census, are beyond the scope of this study. As noted previously, additional international funding beyond the feasibility stage outlined in this report is required. These additional resources are intended primarily to fund time and expertise. Both NIS staff and international experts in small area estimation are needed. NIS expertise remains critical in the questionnaire matching exercise and when considering area code changes, as well as for interpreting local details within a variety of small area models. The international expertise required consists of staff with a strong theoretical understanding of small area estimation methods, including but not limited to the World Bank method, and with experience applying such methods in a range of developing countries. The additional component is equipment. Small area estimation is computer intensive, since even full scale checking a single statistical model can take three days of computation time using a fast PC. Long term sole access to a machine with a minimum memory size of 4GB, and preferably rather more, will be required. (Computers currently available to the NIS Poverty Mapping Group have a maximum of 2GB memory.) Software is not so much an issue at NIS, since ASEAN have very recently supplied the NIS Poverty Mapping Group with licenced copies of Stata 10 for statistical analysis and of ArcView 9.2 for mapping.

4. Introduction – Phase 3: Analysis

The third and final phase of this feasibility study involved statistical modelling of Cambodian sample survey data (CSES 2003/4). The modelling used the variables identified in the second phase of the feasibility study, which had examined data sources and identified variables contained in both the survey and the 2008 census which were suitable for small-area estimation of poverty.

The variables used in the statistical modelling comprise:

- a target variable, here denoted *y*, recorded for each household in the survey and from which household-level poverty status can be inferred; this variable is log-transformed household per capita consumption expenditure (from which poverty incidence, gap and severity are derived). If household-level kilocalorie information had been available to assess undernourishment, household-based per capita daily calorie intake would have been used;

- a set of auxiliary variables, denoted x_1, x_2, ..., x_p, either recorded at household level in both survey and census or else missing from the survey but available from the census or another external source as an average over a geographic area; for example household size (`hhsize`) is a household-level variable available in both CSES 2003/4 and the 2008 Census, whereas number of schools (`SCs`) is a commune-level variable imported from a geographic information systems (GIS) database;

- a set of regional indicators common to the survey, census and other external data sources; these are required so that area-level information can be merged appropriately with the survey and census databases, and so that small-area estimates can be produced at appropriate geographic levels; for Cambodia these are indicators for province, district, commune and village.

The ELL method uses the auxiliary data to infer the value of the target variable *y*, and hence poverty or malnutrition status, for every household in the census via a statistical model

$$y_{vh} = \beta_0 + \beta_1 x_{1vh} + \beta_2 x_{2vh} + \cdots + \beta_p x_{pvh} + \eta_v + \varepsilon_{vh} \qquad (1)$$

where y_{vh} denotes the value of the target variable in the *h*th household of the *v*th village. The unknown parameters $\beta_0, \beta_1, \cdots, \beta_p$ are estimated using the survey data. The disturbance terms η_v, ε_{vh} represent unexplained variation at village and household level respectively; these are treated in the modelling as random effects with mean zero and variances $\sigma_\eta^2, \sigma_\varepsilon^2$. These variance components are also estimated from the survey data.

The quality of the small-area estimates produced by the ELL method, measured in terms of the standard errors of the estimates, depends crucially on the quality of the model linking the target and auxiliary variables. In this part of the report we examine this issue in detail by identifying, from existing datasets, useful auxiliary variables for the model (1), fitting the model to the CSES 2003/4 data and summarizing the performance of the model statistically.

5. Measures of Model Performance

There are two aspects of model performance commonly used to evaluate linear models like (1). The first measures the proportion of the variability in the target variable (y) explained by the predictors (x); this is commonly denoted R^2. The second, the mean squared error (MSE), measures the overall size of the unexplained variation. In the ELL method however there is no direct link between these and the precision of the final small-area estimates; this is partly because the variable of interest at small-area level is not y itself but a nonlinear function of y (for example *poverty incidence* is the proportion of households in an area for which y is below a poverty line). Nevertheless, previous experience with the ELL method, combined with theoretical considerations, enable some statements linking the two to be made.

In successful applications of the ELL method to poverty estimation, the R^2 value of the model for log-transformed per capita expenditure tends to be about 50% or higher. It should be noted that R^2 always increases as more x variables are added to the model, but that a point of "diminishing returns" sets in after which an increase in model complexity gives only a negligible improvement in R^2. Furthermore such apparent improvements in the predictive power of the model may be spurious, holding for the estimation data but not for future predictions. The target for R^2 of around 50% for consumption poverty should be achieved with a modest number p of auxiliary variables in relation to the size of the estimation dataset. Calorie intake tends to be harder to predict than consumption expenditure (perhaps, in part, because of measurement errors); there is less experience here, but reasonably successful small-area estimates have been produced using models with an R^2 around 30%.

In (1) for Cambodia, the unexplained variation is decomposed into village-level (η_v) and household-level (ε_{vh}) effects. When the model is used to predict y for the census households, simulated values of these random effects are included to incorporate the uncertainty in these predictions. When the household-level predictions are amalgamated to small-area (commune) level to produce the final estimates, these effects will tend to "average out", reducing the variability at small-area level; the extent to which this happens depends on the number of households, and the number of villages, in the small areas. Since the number of households is generally much larger than the number of villages, the crucial factor affecting precision is usually the size of the village-level effects. Thus while both σ_ε^2 and σ_η^2 should preferably be small, it is particularly important that the latter should be as small as possible.

Table 1. Summary measures for expenditure models from CSES1997.

	p	R^2	σ_η^2	σ_ε^2
Phnom Penh	38	0.481	0.099	0.359
Other urban	40	0.700	0.034	0.354
Rural	58	0.538	0.141	0.236

Table 1 gives the values of these summary measures for the models used by Fujii (2002) for poverty estimation in Cambodia based on CSES1997. Here different models have been fitted in each stratum. The resulting commune-level poverty incidence estimates were felt to be of sufficient precision on the whole, with standard errors averaging around 5%.

6. Modelling Strategy

There is typically a large set of auxiliary variables available for inclusion in the model (1). It is not good practice to include all of these *x* variables as the complete set would be highly multicollinear, leading to low precision in the estimates, and possibly many spurious relationships leading to bias. A model selection procedure is needed to decide which variables to include and to avoid this "over-fitting". We want to achieve a reasonably high R^2 and low $\sigma_\eta^2, \sigma_\varepsilon^2$ with only a moderate number *p* of variables.

We first distinguish between two kinds of variables: numerical (eg household size with numerical values 1, 2, 3, ...) and categorical (eg cooking fuel with values electricity, lpg, wood, ...). Numerical *x* variables contribute one to the total model size *p*; however if the effect is nonlinear we might want to add quadratic (x^2), cubic (x^3) or other terms to capture the relationship. Categorical variables increase the model size by the number of categories in the variable; here it may be possible to simplify the model by combining categories with a similar effect. One "default" category is omitted, so effects are measured relative to this default.

The model may also include "interactions", allowing the effect of one variable to change with different values of another variable. For example the effect of household size may vary in the different strata (Phnom Penh, Other urban, Rural); to accommodate this we would include `hhsize×strata` interaction terms.

Some approaches to the model selection problem have first created a maximal pool of potential *x* variables by adding quadratic and cubic terms for all numeric variables and all possible two-way (sometimes even three-way) interactions, then running an automatic model selection method (stepwise regression) that lets the computer choose the model. The result can be over-complicated and hard to interpret; it also increases the possibility of spurious relationships in the model. We tend to prefer a more cautious approach, starting with a relatively simple and interpretable model and judiciously adding nonlinear and interaction terms where they seem necessary and where the results seem to be plausible.

Some implementations of the ELL method have used different models in different strata, these strata being defined by the survey sample design or, in the case of a survey design with many small strata, as accumulations of geographically contiguous strata. Our approach is to first investigate the possibility of fitting a single country-wide model, with differences between strata accommodated by interaction terms. This allows for the possibility that the effect of some variables is consistent across the different strata, as well as allowing for differences between strata.

Since the village-level error component plays the largest part in the precision of the final small-area estimates, we try to get σ_η^2 as small as possible. This is often aided by using area-level variables such as GIS data and census means. There is also the possibility of including random effects at other levels in the model. Although it was not included in the original ELL specification, there are sound arguments for including an additional random component at small area level in order to reduce the risk of underestimating standard errors for the small area estimates themselves.

7. Estimation Methods

The data used to estimate the model (1) typically comes from a two-stage survey design: in the first stage "primary sampling units" (PSUs) are randomly sampled within each stratum, after which individual households are selected at random within each sampled PSU. The PSUs usually represent natural clusters of households within the population; in CSES 2003/4 the PSUs are villages. One consequence of this design is that individual households can have different representational values within the population; this is reflected in the different "survey weights" given to each household. Statistical methods exist for taking into account the survey weights and the two-stage structure of the sample.

Early implementations of ELL used ordinary least squares (OLS) regression (ie ignoring the clustering and the weights) when selecting the variables. The chosen model was then re-estimated using a form of generalized least squares (GLS) in which estimated variance components (for $\sigma_\eta^2, \sigma_\varepsilon^2$) were used to re-weight households within clusters. This GLS estimation is included in the PovMap software provided by the World Bank. An alternative is to perform the model selection using a survey regression method specifically designed for the purpose and available in some common statistical programs such as Stata. It is sometimes found that variables which seem to have a significant effect in OLS estimation lose their significance when the survey design is accounted for. There are other, more sophisticated technologies available (eg Rao, 2003) for estimating the final model and its variance components. The differences obtained from these various methods do not seem to be great. Because the present exercise is exploratory, we consider here only Stata's survey regression method to investigate model choice and evaluation.

8. Results

Table 2 gives the results for a model fitted to the CSES 2003/4 survey data with log-transformed per capita consumption expenditure as the target variable. The **Coef** column gives the estimated effect of each variable on log expenditure. For example the model estimates that those who cook with electricity or lpg have on average 0.2611 higher log expenditure than those who do not. Household size has both quadratic and interaction effects – these are best understood graphically, as shown in Figure 1 below.

Table 2. Model for log-transformed per capita consumption expenditure

Variable	Coef.	Std. Err.	t	P>t	Meaning
hhsize	-0.1308	0.0078	-16.77	0.000	# person in hh
hhsq	0.0092	0.0009	10.31	0.000	quadratic effect: (hhsize-5)^2
hhsize_ou	0.0107	0.0070	1.53	0.126	interaction of hhsize with stratum 2 (other urban)
hhsize_r	0.0170	0.0079	2.15	0.032	interaction of hhsize with stratum 3 (rural)
room	0.1723	0.0179	9.65	0.000	# rooms in house
rmsq	-0.0118	0.0049	-2.39	0.017	quadratic effect: (room-1.36)^2
room_r	-0.0359	0.0178	-2.02	0.044	interaction of room with stratum 3 (rural)
rmsq_ou	0.0079	0.0049	1.6	0.110	interaction of room with stratum 3 (rural)
skids6	-0.3995	0.0323	-12.37	0.000	proportion in hh of kids under 7
skids714	-0.1019	0.0315	-3.23	0.001	proportion in hh of kids 7-14
hfem_npp	-0.0854	0.0222	-3.85	0.000	female head of household, outside phnom penh
light_b	-0.2048	0.0280	-7.3	0.000	light = battery (default = electric)
light_k	-0.3996	0.0284	-14.07	0.000	light = kerosene, other or none (default = electric)
ck_elpg	0.2611	0.0283	9.21	0.000	cooking by electricity or lpg
married	0.0506	0.0213	2.38	0.018	head of household married
rent	-0.2497	0.0418	-5.98	0.000	house rented
rent_r	0.1202	0.0508	2.37	0.018	interaction of rent with stratum 3 (rural)
hedulev4	-0.0360	0.0167	-2.16	0.031	head of household educated to primary level
medulev2	0.1735	0.0248	6.99	0.000	proportion of family members with non-formal education
medulev3	0.2939	0.0281	10.45	0.000	proportion of family members with elementary education
medulev4	0.5964	0.0544	10.97	0.000	proportion of family members with primary education
medulev5	0.6697	0.0602	11.12	0.000	proportion of family members with secondary education
medulev4_r	-0.1002	0.0603	-1.66	0.097	interaction of medulev4 with stratum 3 (rural)
migrated	0.0334	0.0148	2.25	0.025	head of household has migrated
pnowork	0.1054	0.0279	3.77	0.000	proportion of family members who do not work
pservice	0.1840	0.0237	7.77	0.000	proportion of family members employed in service sector
floortype6	-0.3333	0.1219	-2.74	0.006	proportion of hh in commune with floortype "other"
walltype3	0.2598	0.0604	4.3	0.000	proportion of hh in commune with walltype "concrete, brick or stone"
rooftype5	0.3180	0.1040	3.06	0.002	proportion of hh in commune with rooftype "asbestos cement sheets"
HCsurb	-0.0831	0.0278	-2.99	0.003	# health centers in commune (urban only)
SCs	0.0093	0.0053	1.77	0.077	# schools in commune
_cons	8.4152	0.0436	192.87	0.000	constant term

Note that there are a number of interaction terms which differentiate between the three strata: Phnom Penh, Other Urban and Rural. The overall affect of these is complex and depends on the distribution of the affected variables in the three strata.

Education level has, as expected, a positive effect on expenditure, with average expenditure getting progressively higher as the education level increases (`medulev2-medulev5`). There is an interaction with stratum, so that the positive effect of primary education is less in the rural areas. The negative coefficient of `hedulev4` is perhaps surprising, but it needs to be interpreted in conjunction with the other education variables.

The **Std Err** column in Table 2 measures the uncertainty in the estimated effects. Roughly speaking we would expect the true values to be within two standard errors of the estimates. The **t** and **P>t** columns are tests of whether the true effect size might be zero. Often a threshhold of 0.05 is used for **P>t**, the "p-value", to decide whether a particular variable should be included in the model, so that variables with p-values above 0.05 would be considered for elimination. We have relaxed this somewhat in the present situation, as we are interested in screening for potentially useful variables, particularly where they represent possible strata differences or where they are commune-level averages that might usefully reduce the commune-level component of unexplained variation.

This model has 31 variables and achieves an R^2 of 0.499, with estimated variance components of 0.163 for σ_ε^2 and 0.045 for σ_η^2. These results seem to be comparable with those of Fujii, so we can be reasonably optimistic of sufficiently precise modelling of expenditure for the production of useful small-area poverty estimates.

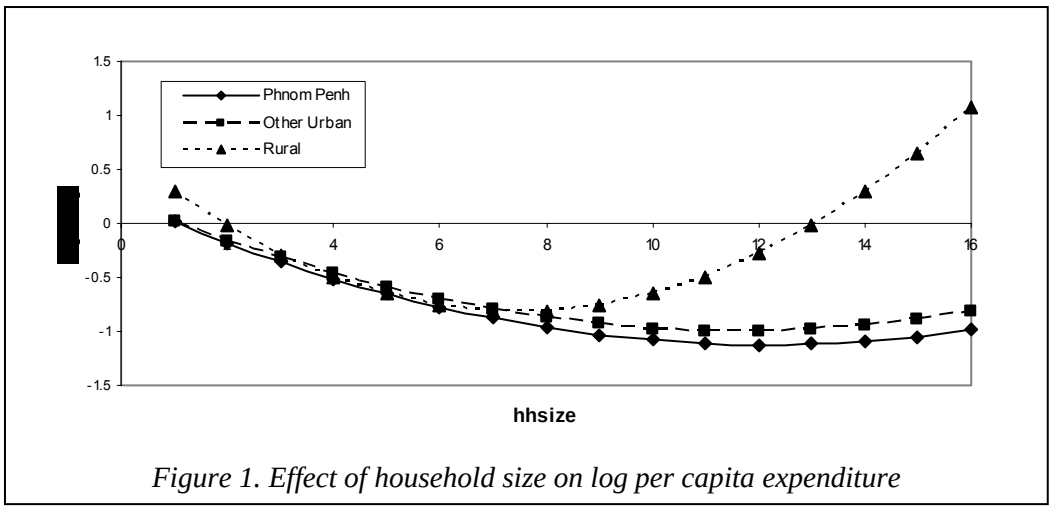

Figure 1. Effect of household size on log per capita expenditure

6. Conclusions

Our modelling exercise, based on the currently available data, indicates that for predicting log-transformed per capita consumption expenditure an R^2 of around 0.5 should be achievable, and that the unexplained variation at commune-level should be reasonably small, or at least comparable with its value in other such analyses. We can therefore be cautiously optimistic about the performance of a future poverty-mapping exercise, assuming similar relationships in the new survey data to be used in such an analysis.

One important factor governing the precision of the small-area estimates is the number of PSUs (villages) within each small area (commune). From previous poverty-mapping exercises in other countries in which we have been involved, the average has been 20 or more: for Cambodia it appears to be 8.6 based on the 1998 census. This means that unexplained commune-level random effects will typically be averaged over only a modest number of values, leading possibly to a lack of precision in at least some of the commune-level estimates. However the fact that our modelling results are comparable to those of Fujii (2002), and that his estimates were for the most part deemed acceptably precise, suggests that this problem may not be too serious. We might nevertheless want to combine geographically adjacent communes in some cases in order to achieve sufficient precision. Furthermore, we would expect in a future modelling exercise to have available census means at commune level, which were not available to us in the present exercise, and which should further reduce the standard errors of the small area estimates of poverty.

Cautious optimism is warranted that updated small area estimation of poverty at commune level in Cambodia is feasible. However, at feasibility stage it is not possible guarantee a similar performance when the new data becomes available. We appear to be entering a period of extreme volatility in both oil and food prices internationally; the effects of these, particularly on the poorest members of a society, are sure to be considerable. This situation will cause the statistical relationships examined in this study to change to some extent, so that given new data the statistical model fitting in this report will need to be repeated, reassessed and confirmed.

The completion of this report 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, World Bank, and World Food Programme (WFP) - which commissioned this research - in Phnom Penh 21-25 April 2008, with WFP in Siem Reap the following week, and again in Phnom Penh 14-16 October 2008. The authors are grateful for these extensive contributions. Viewpoints and opinions expressed in this report do not however necessarily reflect those of all or any of the organisations consulted.

Appendix

Links between Cambodian Census of Population and Housing 2008 and Cambodian Socio-economic Survey (CSES) 2003/4

General Notes for Tables A1-A3 and B1-B12:

1. Tables A1-A3, B1-B12 provide the detail necessary for preliminary matching of variables in the 2008 census (for which the relevant questions are almost identical to the 1998 census) and the Cambodian Socio-economic Survey (CSES). For the CSES, the 2003/4 questionnaire is detailed below. As for the census, CSES questionnaires are very similar, although in years between 2003/4 and 2009, the CSES (when implemented, as in 2007) uses a smaller sample (n=3600 versus n=15.000 approximately) and a reduced bank of questions.

2. Census Forms A and B were not completed simultaneously. Form A was collected several days previous to Form B (ref: 2008 Census Enumerators' Manual p3). This can lead to some mismatching of census households between Form A and Form B.

3. Questionnaire details have been compared in these tables via their English versions; for final matching decisions (between survey and census) local Khmer knowledge of the questions actually asked is essential. The tables below list agreement or otherwise in principle. Even where questions are identical in English, they may not be in Khmer. Even if identical in Khmer, further statistical checking that similar proportions of people in survey and census respond to each apparently equivalent category will be required after the small area estimation feasibility assessment, and before fitting the final small area models to be applied to the 2008 census data.

4. The "Other (specify)" coding used in some census and CSES questions will need to be clarified to ascertain if/where this code has been used to create new codes using specific categories.

5. Standardising by using the same categories and identical questions in census and survey is highly recommended in future, if further small area estimation using a combination of both data sources is planned.

6. The following parts of the census are not directly relevant to matching with the CSES survey questionnaire:
   Form B, Household Questionnaire Part 1
   Charge Register
   List of Defence Establishments
   List of Police Headquarters
   List of Major Institutions
   List of Villages with Boat Population
   List of Remote Villages difficult to Reach
   Training Centres at District Headquarters
   Number of Trainers / trainees
   List of Census Officers
   Enumerator Summary Account for Questionnaires Received by Enumerator
   Supervisors' Summary
   Receipt of materials Given
   List of Filled-in Records Submitted by Enumerator
   Commune Population Statement (Provisional)
   District Population Statement (Provisional)
   Appointment Order for Enumerators / Supervisors
   Appointment Order for the Census Officer

7. The majority of the CSES 2003/4 survey questions are not collected in the census. With the exception of expenditure related information, such non-matching questions are irrelevant to small area estimation using variants of the World Bank methodology for small area poverty estimation, which model expenditure using survey data and use variables that match between survey and census for prediction of expenditure at household level, and which are then combined to small area level.

8. Religion, Physical/mental disability, basis of house occupancy (eg own, rent), internet access at home or elsewhere, and ownership of radio / transistor, television, telephone (fixed), cell phone, personal computer, bicycle, motorcycle, car / van, boat, tractor are collected for the 2008 census, but not for the CSES 2003/4.

9. There are migration related questions in both 2008 census (Form B Household Questionnaire Part 2, Individual particulars Col 9-12) and 2003/4 CSES (Section C) but the questions are different and coding match would be difficult without the extensive local knowledge at the Cambodian National Institute of Statistics (NIS).

10. Tables A1-A3 and B1-B12 below are organised in the same order as the questions appear in the 2008 census questionnaires

Table A1:

Links between Cambodian Census of Population and Housing 2008 and Cambodian Socio-economic Survey (CSES) 2003/4

Wall material – Household level

Topic	***2008 Census Question***	***2008 Census Categories***	***2003/4 CSES Question***	***2003/4 CSES Categories***	***Notes***
Wall material	Form A: Houselist, p1, column 3.	1. Bamboo / thatch / grass / reeds 2. Earth 3. Wood / plywood 4. Concrete / brick / stone 5. Galvanised iron / aluminium / other metal sheets 6. Asbestos cement sheets 7. Salvaged / improvised materials 8. Other (specify)	Form 3, Section3, Q4, p9: "outer wall" and Form 3, Section3, Q5, p9 "inner wall"	1. Bamboo / thatch 2. Wood or logs 3. Plywood 4. Concrete / brick / stone 5. Galvanised iron or aluminium 6. Fibrous cement 7. Makeshift, salvaged or improvised materials 8. Other (specify)	Some but not all categories are identical; amalgamation of categories likely to be required to match census and survey information CSES has "first material" and "second material" as separate sub-questions using the same CSES codes 1-8 CSES 2003/4 has separate "inner wall" and "outer wall" questions using same CSES codes 1-8

Table A2:

Links between Cambodian Census of Population and Housing 2008 and Cambodian Socio-economic Survey (CSES) 2003/4

Roof material – Household level

Topic	***2008 Census Question***	***2008 Census Categories***	***2003/4 CSES Question***	***2003/4 CSES Categories***	***Notes***
Roof material	Form A: Houselist, p1, column 4	1. Bamboo / thatch / grass 2. Tiles 3. Wood / plywood 4. Concrete / brick / stone 5. Galvanised iron / aluminium / other metal sheets 6. Asbestos cement sheets 7. Plastic / synthetic material sheets 8. Other (specify)	Form 3, Section3, Q6, p9	1. Thatch 2. Tiles 3. Fibrous cement 4. Galvanised iron or aluminium 5. Salvaged materials 6. Mixed but predominantly made of iron / aluminium, tiles or fibrous cement 7. Mixed, but predominantly made of thatch or salvaged materials 8. Concrete 9. Plastic sheet 10. Other (specify)	Some but not all categories are identical; amalgamation of categories likely to be required to match census and survey information CSES has “first material” and “second material” as separate sub-questions using the same CSES codes 1-10

Table A3:

Links between Cambodian Census of Population and Housing 2008 and Cambodian Socio-economic Survey (CSES) 2003/4

Floor material – Household level

Topic	***2008 Census Question***	***2008 Census Categories***	***2003/4 CSES Question***	***2003/4 CSES Categories***	***Notes***
Floor material	Form A: Houselist, p1, column 5	1. Earth / clay 2. Wood / bamboo planks 3. Concrete / brick / stone 4. Polished stone 5. Parquet / polished wood 6. Mosaic / ceramic tiles 7. Other (specify)	Form 3, Section3, Q7, p9	1. Earth / clay 2. Wooden planks / bamboo strips 3. Cement 4. Parquet / polished wood 5. Polished stone / marble 6. Vinyl 7. Ceramic tiles 8. Other (specify)	Some but not all categories are identical; amalgamation of categories likely to be required to match census and survey information CSES has “first material” and “second material” as separate sub-questions using the same CSES codes 1-8

Table B1:

Links between Cambodian Census of Population and Housing 2008 and Cambodian Socio-economic Survey (CSES) 2003/4

Household size, Head of household - Household level

Topic	***2008 Census Question***	***2008 Census Categories***	***2003/4 CSES Question***	***2003/4 CSES Categories***	***Notes***
Household size	Page 3, Form B: Household questionnaire Part 2, derivable as sum of number of entries (ie names) in col 2.	Numeric: 1, 2, 3,…..	Form 3, Section 1A, derivable from "Listing of household members", Col 2, p2	Numeric: 1, 2, 3,…..	Usually need to watch during modelling for household sizes in census that are very much in excess of those in the survey, especially if a 'household size squared' term is included in the survey-based model.
Household head	Page 1 Form B, Household Questionnaire, col 3; also collected in Form A, col 8.	In Form A, name and sex of head of household only is collected. In Form B, "Relationship to head of household" has the instruction "(Write in words)" below it on the questionnaire.	Form 3, Section 1A, via "Relationship to head of household", col 6.	Code '01' denotes head of household	The name sex of head of household is available from census Form A, and name and other personal characteristics including sex from Form B. Note that census question is asked twice. Personal characteristics of head of household including name are also available from the CSES questionnaire.

Table B2:

Links between Cambodian Census of Population and Housing 2008 and Cambodian Socio-economic Survey (CSES) 2003/4

Sex, Age – Individual level

Topic	***2008 Census Question***	***2008 Census Categories***	***2003/4 CSES Question***	***2003/4 CSES Categories***	***Notes***
Sex	Page 3, Form B: Household questionnaire Part 2, Col 4.	1. Male 2. Female	Form 3, Section 1A, derivable from "Sex", Col 3, p2	1. Male 2. Female	Hence proportion of males or proportion of females (or proportions of males and females within age ranges using sex, and age below)
Age	Page 3, Form B: Household questionnaire Part 2, Col 5.	Codes 0-97 correspond to age in completed years, 98 to 98 or more	Form 3, Section 1A, derivable from "Age in completed year", Col 5, p2	Codes 0-95 correspond to age in completed years, 96 to age 96 or more, and 98 to don't know	Take care with '98' code for CSES; can cross check age against date of birth question for CSES only

Table B3:

Links between Cambodian Census of Population and Housing 2008 and Cambodian Socio-economic Survey (CSES) 2003/4

Marital Status, Languages Spoken – Individual level

Topic	***2008 Census Question***	***2008 Census Categories***	***2003/4 CSES Question***	***2003/4 CSES Categories***	***Notes***
Marital Status	Page 3, Form B: Household questionnaire Part 2, Col 6	1. Never married 2. Married (ie currently married) 3. Widowed 4. Divorced 5. Separated	Form 3, Section 1A, derivable from "Age in completed year", Col 9, p2	1. Never married 2. Married (ie currently married) 3. Live together 4. Widowed 5. Divorced 6. Separated	Categories are identical, except for additional code "living together" in CSES 2003/4. Note: no code for males who have lost partners or wives
Languages spoken	Page 3, Form B: Household questionnaire Part 2, Col 7, which allows a maximum of three	29 options including all six of those specified in CSES 2003/4	Form 3, Section 1A, derivable from "Age in completed year", Col 12a, 12b, 12c, ie maximum of three	0. None 1. French 2. English 3. Chinese 4. Vietnamese 5. Thai 6. Lao 7. Other (specify)	The CSES 2003/4 question is, "Can you speak languages other than Khmer?" – all a foreign; the census 2008 question is, "Mother Tongue" Note: Ethnicity is collected in CSES 2003/4 but not in census

Table B4:

Links between Cambodian Census of Population and Housing 2008 and Cambodian Socio-economic Survey (CSES) 2003/4

Literacy – Individual level

Topic	***2008 Census Question***	***2008 Census Categories***	***2003/4 CSES Question***	***2003/4 CSES Categories***	***Notes***
Literacy	Page 4, Form B: Household questionnaire Part 2, Col 13 (a) and (b)	1. Yes 2. No.	Form 3, Section 2, Col 2 and 3, p7	1. Yes 2. No	Census asks "read and write with understanding" in two parts: (a) Khmer (b) any other language. CSES 2003/4 has separate questions for reading (col 2) and writing (col 3) "a simple message" in any language.

Table B5:

Links between Cambodian Census of Population and Housing 2008 and Cambodian Socio-economic Survey (CSES) 2003/4

Education – Individual level

Topic	***2008 Census Question***	***2008 Census Categories***	***2003/4 CSES Question***	***2003/4 CSES Categories***	***Notes***
Education	Page 4, Form B: Household questionnaire Part 2, Col 14 (a) and (b)	(a) Attending school 1. Never 2. Now 3. Past (b) Highest grade - Never attended 00 No class completed 01 Class 1 completed …. 12 Class 12 completed 13 Lower 2^{ndary} diploma 14 2^{ndary} school baccalaureate holder 15 Technical / vocational pre-2^{ndary} diploma 16. Technical / vocational post-2^{ndary} diploma 17. Undergraduate 18. Graduate 19. Postgrad and above 20. Other (specify)	Form 3, Section 2, Col 5 asks 'highest level'	Highest grade 90 None 98. Don't know 00. Preschool / kindergarten 01 Class 1 completed …. 12 Class 12 completed 13 2^{ndary} school certificate 14. Technical / vocational pre-2^{ndary} diploma 15. Technical / vocational post-2^{ndary} diploma 16. Undergraduate 17. Graduate 18. Postgrad and above 19. Other (specify)	The equivalent to census 14(a) - the attending school question – can in principle be derived from the CSES 2003/4 The codes on highest grad are very similar except for codes -, 00, 90, 98, and census codes 13 and 14 being combined so that census codes and CSES 2003/4codes over 13 differ by one.

Table B6:

Links between Cambodian Census of Population and Housing 2008 and Cambodian Socio-economic Survey (CSES) 2003/4

Employment – Individual level

Topic	***2008 Census Question***	***2008 Census Categories***	***2003/4 CSES Question***	***2003/4 CSES Categories***	***Notes***
Employment	Page 4, Form B: Household questionnaire Part 2, Col 16-23. Questions asked are: Main activity Employment period Occupation Employment status Industry, trade or service Sector 2ndary economic activity Place of work or school	See census questionnaire page 4, Form B	Form 3, Section 13, A-C, pages 38-40 Questions asked are For the last 7 days - Economic activity Number of occupations Primary occupation Secondary occupation Type of economic activity Hours worked For the last month: Days worked Employment status Type of employer Earnings if paid For the last 12 months Main activities 1ary and 2ary occupations Kind of economic activity	See CSES questionnaire Form 3, Section 13, A-C, for details	The questions in the employment category in census and CSES 200/4 are similar but not identical. For example, the periods for which the questions apply are different, as do the times of year to which they apply since CSES 2003/4 and census were not conducted at the same time of year. Some additional information may be able to be extracted, but to simplify primary emphasis should be placed on whether employed or not, plus agricultural links. Further advice from NIS also advised.

Table B7:

Links between Cambodian Census of Population and Housing 2008 and Cambodian Socio-economic Survey (CSES) 2003/4

Fertility - Individual level

Topic	***2008 Census Question***	***2008 Census Categories***	***2003/4 CSES Question***	***2003/4 CSES Categories***	***Notes***
Number of children born alive	Page 5, Form B: Household questionnaire Part 3, Col 4 (a) male (b) female	Count data	Form 3, Section 10, Col 5a and 5b, p33 (a) male (b) female	Count data	
Number of children living	Page 5, Form B: Household questionnaire Part 3, Col 5 (a) male (b) female	Count data	Form 3, Section 10, Col 6a and 6b, p33 (a) male (b) female 7a and 7b, p33 (a) male (b) female	Count data	The CSES 2003/4 Col 6a and 6b are "..living in this household" and 7a and 7b are "..not living in this household"; the census 2008 question is simply, "…living".
How many children have died?	Page 5, Form B: Household questionnaire Part 3, Col 6 (a) male (b) female	Count data	Form 3, Section 10, Col 6a and 6b, p33 (a) male (b) female	Count data	Sensitive question: A similar pair of questions in survey and census elicited very different responses due to context, with severe under-reporting for the census

Table B8:

Links between Cambodian Census of Population and Housing 2008 and Cambodian Socio-economic Survey (CSES) 2003/4

Housing conditions: Source of light – Household level

Topic	***2008 Census Question***	***2008 Census Categories***	***2003/4 CSES Question***	***2003/4 CSES Categories***	***Notes***
Main source of light	Page 6, Form B: Household questionnaire Part 4, Col 2	1. City power 2. Generator 3. Both 1 and 2 4. Kerosene 5. Candle 6. Battery 7. Other (specify)	Form 3, Section 3, Q8, p10	1. Publicly owned electricity 2. Privately generated electricity / generator 3. Battery 4. Kerosene lamp 5. None 6. Other (specify)	Similar categories. Check whether percentages match after amalgamating categories for both survey and census before fitting models for any final small area estimation

Table B9:

Links between Cambodian Census of Population and Housing 2008 and Cambodian Socio-economic Survey (CSES) 2003/4

Housing conditions: Main cooking fuel – Household level

Topic	***2008 Census Question***	***2008 Census Categories***	***2003/4 CSES Question***	***2003/4 CSES Categories***	***Notes***
Main cooking fuel	Page 6, Form B: Household questionnaire Part 4, Col 3	1. Firewood 2. Charcoal 3. Kerosene 4. LPG 5. Electricity 6. None 7. Other (specify)	Form 3, Section 3, Q23, p11	1. Firewood 2. Charcoal 3. Firewood and charcoal 4. LPG 5. Kerosene 6. Publicly provided electricity 7. Gas and electricity 8. Privately generated electricity 9. None / don't cook 10. Other (specify)	Similar categories. Check whether percentages match after amalgamating categories for both survey and census before fitting models for any final small area estimation

Table B10:

Links between Cambodian Census of Population and Housing 2008 and Cambodian Socio-economic Survey (CSES) 2003/4

Housing conditions: Toilet facility within premises – Household level

Topic	***2008 Census Question***	***2008 Census Categories***	***2003/4 CSES Question***	***2003/4 CSES Categories***	***Notes***
Toilet facility within premises	Page 6, Form B: Household questionnaire Part 4, Col 4	1. Not available 2. Connected to sewerage 3. Septic tank 4. Pit latrine 5. Other type (specify)	Form 3, Section 3, Q20, p11	1. Connected to sewerage 2. Septic tank 3. Pit latrine 4. Other without septic tank 5. Public toilet 6. Shared toilet 7. Open land 8. None 9. Other (specify)	Similar categories except that census only asks what toilet facilities are on the premises. Off-premise categories for survey can be recoded as equivalent to census code 1. 'not available' (on premises). Check whether percentages match after amalgamating categories for both survey and census before fitting models for any final small area estimation

Table B11:

Links between Cambodian Census of Population and Housing 2008 and Cambodian Socio-economic Survey (CSES) 2003/4

Housing conditions: Main source of drinking water – Household level

Topic	***2008 Census Question***	***2008 Census Categories***	***2003/4 CSES Question***	***2003/4 CSES Categories***	***Notes***
Main source of drinking water	Page 6, Form B: Household questionnaire Part 4, Col 5 "main source of drinking water supply"	1. Piped water 2. Tube / pipe well 3. Protected dug well 4. Unprotected dug well 5. Rain 6. Spring, river, stream, lake/pond 7. Bought 8. Other (specify)	Form 3, Section 3, Q9 (wet season), Q13 (dry season) p10	1. Piped in dwelling or on premises 2. Public tap 3. Tubed / piped well or borehole 4. Protected dug well 5. Unprotected dug well 6. Pond, river or stream 7. Rainwater 8. Tanker, truck, vendor or otherwise bought 9. Other (Specify)	Similar categories, but watch for seasonal changes in water sources given timing of survey and census. Check whether percentages match after amalgamating any categories for both survey and census before fitting models for any final small area estimation Note there are other questions on water in both survey and census. These, while different, may still be able to help with matching of categories.

Table B12:

Links between Cambodian Census of Population and Housing 2008 and Cambodian Socio-economic Survey (CSES) 2003/4

Housing conditions: Number of rooms occupied by household - Household level

Topic	***2008 Census Question***	***2008 Census Categories***	***2003/4 CSES Question***	***2003/4 CSES Categories***	***Notes***
Number of rooms occupied by household	Page 6, Form B: Household questionnaire Part 4, Col 7	1. One room 2. Two rooms 3. Three rooms 4. Four rooms 5. Five rooms 6. Six rooms 7. Seven rooms 8. Eight rooms or more	Form 3, Section 3, Q3, p10	Coded as number of rooms	Census excludes kitchen, bathroom, toilet and storeroom. Survey excludes kitchen, toilet and bathrooms. Note: Floor area is included in CSES as Form 3, Section 3, Q2, p10, but is not available in the census. Check whether percentages match after amalgamating categories for both survey and census before fitting models for any final small area estimation

Variables included on the NCDD National Committee for Decentralisation and Deconcentration, ie Seila database

FAMILY	Total number of families
FEM_TOT	Total number of females
MAL_TOT	Total number of males
FEM_0_5	# Girls 0-5 years old (under 6's)
MAL_0_5	# Boys 0-5 years old (under 6's)
FEM_6_14	# Girls 6 to 14 years old
MAL_6_14	# Boys 6 to 14 years old
F6_14_go	# Girls 6 to 14 who go to school
M6_14_go	# Boys 6 to 14 who go to school
FEM15_17	# Women 15 to17 years old
MAL15_17	# Men 15 to17 years old
FEM18_64	# Women 18 to 64 years old
MAL18_64	# Men 18 to 64 years old
F_OVER65	# Women over 65 years of age
M_OVER65	# Men over 65 years of age
F_ILT_15	# Illiterate women over 15 years old
M_ILT_15	# Illiterate men over 15 years old
THATCH_R	# Houses with thatched roof
TILE_R	# Houses with tiled roof
FIBRO_R	# Houses with fibro roof
ZINC_R	# Houses with zinc roof
CONCR_R	# Houses with concrete roof
TOILET	# Latrines
H2O_HOUSE	# Families with piped water, private pump well or private ring well, usable year round, at their house, less then 150m.
H2O_150M	# Families with a communal tap, pump well or ring well, usable year round, within 150m of their house.
H2O_OTHER	Most common source of water for other families: pond, river, rain water, other.
COW_FAMI	# families with cattle and buffalo in village
PIG_FAMI	# families with pigs in the village
PAD_PRICE	Average farm gate price of paddy in Riel for this month, December
MOTO_NUM	# Motorcycles
CAR_NUM	# Tractors/koyons/cars
OX_CART	# Horse carts and ox carts
BICY_NUM	# Bicycles
ROW_BOAT	# Row boats
MOTOBOAT	# Boats with motor
TV_MUM	# TVs
TBA_USE	# families who used a traditional birth attendant in the past year
MWIFE_USE	# families who used a trained midwife in the past year
TBA_NUM	# traditional birth attendants in the village
MWIFE_NUM	# government trained midwives in the village
NUM_IRRI	# family who have some irrigated rice land
NUM_FERT	# family using chemical fertilizer in the past year
NUM_PEST	# family using pesticide in the past year
INSECURITY	# murder, robbery, theft cases in the past year
NUM_LANDC	# land conflict case in the past year
F_HH_UD5	# female headed household/families, where the head is a mother with one or more children with under 5 yrs old
HH_VIOLEN	# families having problems with violence in home
MARK_LOC	Name of the nearest market villager frequently go to buy goods
HRS_MARK	Time taken to get from village to this nearest market by motor or motorboat
KM_ROAD	Distance in Km to nearest year-round road (4 wheel motor vehicles)
HRS_ROAD	Time taken to get from village to nearest year-road by motor or motorboat

GARBAGE	# Houses which have access to garbage collection by a garbage collector?
PIPWATER	# Houses which have access to piped water
ELECTRIT	# Houses with electricity
TOIL_FAM	# Families with latrines
MOTO_FAM	# families with motorcycles
CAR_FAM	# families with tractors/koyons/cars
BICY_FAM	# families with bicycles
TV_FAM	# families with TVs
TRAFFICKI	# trafficking cases reported in the past year
CROOM_P	# primary school classrooms in the commune
CROOM_S	# secondary school classrooms in the commune
TEACH_P	# primary school teachers in the commune
TEACH_S	# secondary school teachers in the commune
WET-RAINFED	Area wet season rain fed rice land in Ha
WET _IRRI	Area wet season supplemental irrigated rice land in Ha
WET _PADDY	Rice production in wet season, MT
DRY_IRRI_FULL	Area of full-irrigated dry season rice land in Ha
DRY_RECESS	Area of recession dry season rice land in Ha
DRY_PADDY	Rice production in dry season, MT
CROOM_K	# kindergarten classrooms
TEACH_K	# kindergarten teachers in the commune
F_HHH	# Female household headed
M_ILT15_17	# illiterate men from 15-17ys
M_ILT18_64	# illiterate men from 18-64ys
M_ILT_Ov65	# Illiterate men over 65 years old
M15_14SCH	# boy 15 to 17ys who go to school
M6_17SCH	# boy 6 to 17ys who go to school
F_ILT15_17	# illiterate women from 15-17ys
F_ILT18_64	# illiterate women from 18-64ys
F_ILT_Ov65	# Illiterate women over 65 years old
F15_14SCH	# girl 15 to 17ys who go to school
F6_17SCH	# girl 6 to 17ys who go to school
H_THATCH	# family living in thatched roof
H_TILE	# family living in tiled roof
H_FIBRO	# family living in fibro-cement roof
H_ZINC	# family living in zinced roof
H_CONCR	# family living in concrete roof
Lat_that	# Latrine in total thatch house
Lat_tile	# Latrine in total tiled house
Lat_fibro	# Latrine in total fibro-cement house
Lat_zinc	# Latrine in total zinc house
Lat_conc	# Latrine in total concrete house
TV_that	# TVs in total thatch house
VT_tile	# TVs in total tiled house
TV_fibro	# TVs in total fibro-cement house
TV_zinc	# TVs in total zinc house
VT_conc	# TVs in total concrete house
Baby_born	# Women deliver baby in village
Pri_class	# Primary classes in commune
Sec_class	# Secondary classes in commune
Kid_class	# Kindergarten classes in commune
Ric_area	Rice land area in commune
M0_5SCH	# boy 0 to 5ys who go to school
F0_5SCH	# girl 0 to 5ys who go to school

References

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.

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.

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

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

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, The World Bank.

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.

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.

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

Healy A.J., Jitsuchon S. and Vajaragupta, Y. (2003) "Spatially disaggregated estimation of poverty and inequality in Thailand", preprint.

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

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 United Nations World Food Programme.

Jones, G., Haslett, S. and Parajuli, D. (2006) *Local Estimation of Poverty, Undernourishment and Malnutrition in Nepal*, The Central Bureau of Statistics of His Majesty's Government of Nepal / United Nations World Food Programme.

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.

Ministry of Health, Cambodia / World Food Programme / Measure DHS+ - ORC Macro (2003) *Micro-level Estimation of the Prevalence of Stunting and Underweight Among Children in Cambodia*, March 2003, iii+19 pages.

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.

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. (2003) *Small Area Estimation*, Wiley, New York.

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.

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.

Zhao, Q. (2006) *User manual for PovMap,* The World Bank.

http://siteresources.worldbank.org/INTPGI/Resources/342674-1092157888460/Zhao_ManualPovMap.pdf