Micro-level Estimation of Child Malnutrition Indicators
In Cambodia

Tomoki Fujii*

September 30, 1994

Abstract

Child malnutrition remains a serious problem in developing countries. By combining
demographic and health survey (DHS) and population census data, we disaggregate the
estimates of the prevalence of child malnutrition in Cambodia from currently available
17 strata into 1594 commune. The methodology in this paper is built on the small-
area estimation technique developed by Elbers, Lanjouw and Lanjouw (2002;2003). We
extended it to jointly estimate multiple indicators and to allow for a richer structure
of error terms. We use the commune-level estimates to analyze the current situation
of child malnutrition and explore the correlation with other indicators including con-
sumption poverty. We also apply the methodology to a decomposition analysis of health
inequality and discuss the implications of health inequality for targeting. Finally, we
graphically present the potential gains from geographic targeting carried out at different
levels of geographic aggregation.

*University of California, Berkeley email:fujii@are.berkeley.edu. THIS VERSION IS STILL PRELIMI-
NARY. PLEASE DO NOT CITE. Comments are welcome. I am deeply indebted to Chris Elbers, Jean
Olson Lanjouw and Peter Lanjouw for their advice from the beginning of this research. I also thank Alain
de Janvry, Livia Montana, Mahadevan Ramachandran, Martin Ravallion, H.E. Kim Saysamalen, Elisabeth
Sadoulet, Boreak Sik and H.E. San Sythan. Of course, all the remaining mistakes are mine.
1 Introduction

Malnutrition remains a major public health concern in most developing countries. The World Health Organization (2002) estimates that about 3.7 million deaths among young children worldwide were related to malnutrition in 2000. Similarly, Pelletier et al. (1994) estimate that about one half of childhood deaths in developing countries are caused by undernutrition. Malnutrition has also been associated with mortality and morbidity in later life, delayed mental development, and decreased cognitive and behavioral functioning throughout childhood and adolescence (de Onis et al., 2000) and Galler and Barrett (2001) and poorer performance in school (Shariff et al., 2000; Glewwe et al., 2001).

In Cambodia, almost half of the children under five are malnourished as measured by the height-for-age or weight-for-age indicator (National Institute of Statistics et al., 2001). Given the grave consequences of malnutrition, it is clear that the issue must be tackled seriously. However, as with many other developing countries, the resources available for improving children’s nutritional status are severely limited in Cambodia. Hence, efforts must be made to allocate the available resources in an efficient manner. In particular, geographic targeting, or targeting according to locational information, is often easy to administer and implement, and can be quite effective when malnourished children are concentrated in certain locations. However, the formulation of an effective geographic targeting policy requires knowledge of the location of malnourished children, information which is often not readily available. In Cambodia, the CDHS 2000 data provides information of child nutrition status at the stratum level (a level that is more aggregated than the provincial level.\(^1\)) Since there are only 17 strata for the CDHS 2000 data, such information is too aggregated to be useful for formulating targeting policies.

Such informational constraints are among the central issues concerning the formulation of

\(^1\)In Cambodia, there are four administrative divisions. They are province, district, commune and village in the descending order of aggregation.
targeting policies (Ravallion and Chao, 1989; Kanbur, 1987). This study aims to overcome the problem of excessive aggregated data by producing commune-level estimates of the prevalence of malnutrition in Cambodia. The estimates can be projected onto maps, which in turn allows policy-makers to visually identify areas of severe child malnutrition. These nutrition maps are useful for the analysis of the current situation of malnutrition and for the formulation of geographic targeting policies aimed at assisting the neediest people in a more efficient and transparent manner.

To derive the commune-level estimates, we combine the CDHS 2000 dataset and with individual level Cambodian Population Census data for 1998. The former includes information on child nutrition status but has a limited number of observations, while the latter covers virtually everyone in Cambodia but lacks any specific information on child nutrition status. The approach we take builds on the small area estimation techniques developed by Elbers, Lanjouw and Lanjouw (2000, 2002, 2003; Hereafter ELL). Their small area estimation approach was first applied to Ecuador (Hentschel et al., 2000) and has subsequently been applied to poverty and inequality measures in many countries, including Cambodia (See, Alderman et al. (2002); Demombynes et al. (2002); Fujii (2004)). We extended their methodology to jointly estimate multiple indicators and to allow for a richer structure of error terms. This was a necessary and challenging step to address the issues unique to nutrition indicators.

While the commune-level estimates of the prevalence of malnutrition are in themselves of interest for policy-makers, we take the analysis one step further and illustrate three distinct but related applications of the methodology. First, we explore the relationship between consumption poverty, health and other commune-level variables. Such analysis would not have been possible without the commune-level estimates. Second, we decompose health inequality indicators into between group and within group components by geographic information. This application is useful for elucidating the significance of the geographic information in explaining overall inequality. We propose two decomposable inequality indicators that are useful
for this purpose. One is based on the analysis of variance. The other uses the concentration curve, which is similar to the Lorenz curve but uses a different ranking of individuals on the horizontal axis. We show that the concentration curve is closely linked to several efficiency measures of targeting.

Third, we measure the potential gains from geographic targeting in the presence of nutrition maps. We assume that, in the absence of the nutrition maps, we can target resources only at the stratum level. Given a fixed budget, we can compare reductions in child malnutrition with stratum-level and commune-level targeting. Alternatively, we can ask how much less resource we need to achieve the same goal if targeting is carried out at a smaller geographic levels. How much efficiency we would gain depends on the spatial distribution of malnourished children. As we shall show later, since the spatial distribution of malnourished children is relatively homogenous in Cambodia, the gains from geographic targeting are not extremely large. Still, we cannot ignore the potential of geographic targeting.

This paper is structured as follows: Section 2 reviews the measurement and prediction of nutritional status of children. Section 3 develops the methodology of nutrition mapping. We then discuss the data we used in Section 4. Section 5 presents the results. Section 6 discusses the applications of the methodology. Section 7 concludes.

2 Measurement and prediction of child nutrition status

To measure malnutrition in a non-invasive and inexpensive manner, anthropometry has been widely used among nutritionists and epidemiologists. Among the most commonly used anthropometric measures are weight-for-height, weight-for-age and height-for-age z-scores. Z-scores measure how many standard deviations away a person’s value is from the median of the National Center for Health Statistics (NCHS) growth reference population of the same sex. The deficiency of the weight-for-height, weight-for-age and height-for-age Z-scores is
respectively called “wasting”, “underweight”, and “stunting.” We use the conventional cut-off point of -2 to calculate the prevalence of malnutrition. For example, the prevalence of stunting is defined as the number of children with less than -2 height-for-age Z-score over the total number of children. (See Waterlow et al. (1977); WHO Working Group (1986, 1995) for further discussion on Z-scores).

As WHO Working Group (1986) points out, there are several obvious differences among these measures. First, one can lose weight but not height. Second, linear growth is a slower process than growth in body mass. Third, catch-up in height is possible, but takes a relatively long time even with a favorable environment. Thus, wasting reflects ‘acute’, or short-term, malnutrition whereas stunting reflects ‘chronic’, or long-term, malnutrition with underweight somewhere in between. Given these differences, it should not be surprising if patterns of wasting and stunting are different.

Victoria (1992) found no systematic pattern that holds for an international population between levels of stunting and wasting. Stratum level comparison of the prevalence of wasting, underweight and stunting derived from the CDHS data is consistent with this observation. While the correlation between stunting and underweight is positively and significantly correlated, the correlation between stunting and wasting is negatively and significantly correlated. No significant relationship is found between wasting and underweight.

We report in this study on the prevalence of stunting and underweight, but not wasting. The primary reason for this is that we were unable to construct a regression model for weight-for-height with sufficient explanatory power.\footnote{It is possible to estimate weight-for-height from weight-for-age, height-for-age and age. This is a straightforward extension of this research.} Interestingly, this seems to be the case in other countries. Alderman (2000) created regression models of various anthropometric indicators for Vietnam, South Africa, Pakistan and Morocco, and the variability of weight-for-height was least well-captured among all the anthropometric indicators we used. This may be because of the fact that all of the regressors used in Alderman (2000) reflect the
welfare in a relatively long run. Hence, it should not be surprising that the variation of the weight-for-height, a very short-term measure, is most poorly explained of all. The regression results of Alderman (2000) also suggest that community level effects are of great importance. We also found that the commune-level variables are important for explaining the variation of anthropometric indicators.

As with Pradhan et al. (2003), we used standardized height and weight, which are the z-scores converted back to the corresponding height and weight of the reference age-sex group of 24-month-old girls. The standardized height and weight are an affine transformation of z-scores and preserve all the desirable properties that the original z-scores possess. The additional merit of the standardized height and weight is that they are always positive for practically possible values of z-scores, and that we can compute inequality measures in terms of height or weight. The choice of the reference group of 24-month-old girls is to make this study comparable with (Pradhan et al., 2003).

Since the methodology we develop in this study is built on the association between anthropometric indicators and other socio-economic and geographic indicators, it is instructive to briefly overview the previous studies on the relationship between anthropometric indicators and other indicators. While experiences from other countries are not necessarily applicable to Cambodia, it would make sense to try models that have proved useful in explaining the variation of anthropometric indicators in other countries.3

Li et al. (1999) investigated the issue of malnutrition with various anthropometric indices and examined its correlates in a large sample of poor rural minority children in China. In this study, age, maternal height, water sources, maternal education and very low income were significant correlates. In a study for Vietnam, the similar factors were found to be relevant. Height-for-age Z-scores were significantly correlated with the age of the child, maternal weight and height, parental education, and some indicator variables on access to

3Note also that our model is a predictive one and not intended to describe causal relationship.
water sources (Haughton and Haughton, 1997). In West Africa, residence in a dry zone was found to be associated with wasting, but not with stunting once other variables are introduced as controls (Curtis and Hossain, 1998). Frongillo et al. (1997) used a national-level data and found that higher energy availability, female literacy and gross product were the most important factors associated with lower prevalence of stunting.

Monteiro et al. (1997) investigated the patterns of intra-familiar distributions of undernutrition in Brazil. They analyzed data for four income strata separately and found that undernutrition was significantly associated among household members for the 25 percent poorest families. Khorshed Alam Mozumder et al. (2000) investigated the effects of the length of birth interval on malnutrition in two districts in Bangladesh. They concluded that the results indicate the potential importance of longer birth intervals in reducing child malnutrition. Zeini and Casterline (2002) explored the importance of different levels of geographic clustering in Egypt, including regional level, governorate-level, local level, household level and individual level, using the 2000 Egypt Demographic and Health Survey. They found that spatial clustering does seems to be important in that country. They also found that, even after controlling for socioeconomic factors, significant household-level clustering remained and, interestingly, that individual clustering was also important. They also found little evidence that children suffering from nutritional problems, as evidenced by anthropometric measures, are more likely to suffer from anemia. Finally, while association between various nutritional risks may did exist in general, Zeini and Casterline (2002) found that underweight was correlated with stunting and/or wasting in some, but not all, governorates of Egypt.
3 Methodology

Overview  We describe in this section the methodology used to estimate the prevalence of malnutrition at the level of small geographic areas. The methodology developed here is similar to the small area estimation procedure developed by Elbers et al. (2000, 2002, 2003) for estimating consumption poverty and inequality in that we also combine survey data with census data in order to obtain estimates of malnutrition at a lower level of aggregation than the survey permits. We first provide the general overview of the methodology, and then describe it more formally.

The basic idea of the methodology is simple. We first construct a prediction model of anthropometric indicators using only the variables that are common between the census and the survey, along with geographic indicators available for the entire country at the village or commune level. The geographic indicators include the remotely-sensed data as well as village-level statistics derived from the census data. Common geographic codes in all data sets allow these to be linked to both the survey and the census data.

The parameters of the model are estimated with the survey dataset. An important feature of our study and of the ELL approach is the explicit treatment of the error terms. We estimate regression coefficients and the associated variance-covariance matrix, and also scrutinize the distribution of the disturbance terms in order to carry out simulation. In each round of simulation, we randomly draw regression coefficients and disturbance terms in accordance with their estimated distribution and we impute the anthropometric indicators to each census record. By geographically aggregating the imputed anthropometric indicators, we can estimate the prevalence of malnutrition.

There are four major differences between our methodology and the ELL approach. The first and most obvious difference concerns the type of the survey dataset used for estimation. The ELL approach typically uses consumption expenditure taken from a socio-economic survey, but we used anthropometric measures taken from a Demographic and Health Survey
(DHS). A second difference stems from different units of analysis. Consumption data are usually produced at the household level, whereas anthropometric measures obtain at the individual level. In the ELL approach, disturbance terms are decomposed into a location-specific effect (usually at the cluster-level) and a household-specific effect. However, in our study, it is important to allow as well for an unobserved individual-specific effect in addition to the location-specific effect and household-specific effects.

The third difference is related to the second point. The number of children under five in a household is very limited, no more than two for most of the households. Hence, we cannot use large-sample property to estimate the parameters for individual-specific effect. Thus, unlike the ELL approach, we make finite-sample corrections. The fourth point is the number of left-hand-side variables used in this study. In the ELL approach, only a single consumption or income measure is considered. In our case, we consider multiple indicators. If the unobserved parts of the different indicators are correlated, this should be taken into account when computing the parameter estimates. We shall call such correlation the intra-personal effect, and allow for it in the model. Given that weight-for-age is a medium-term indicator of nutritional status, which is partly affected by height-for-age, it would be natural to assume that intra-personal effect may exist. The results of Zeini and Casterline (2002) also seem to suggest that intra-personal effect may exist in certain locations. Consequently, our approach involves simultaneous estimator of several models, unlike the single-model procedure applied by Elbers, Lanjouw and Lanjouw (2002, 2003).

**Parameter Estimation** Let us now describe the methodology more formally. We denote the set of all clusters by \( \mathcal{C} \), the set of all households in cluster \( c \in \mathcal{C} \) by \( \mathcal{H}_c \) and the set of all individuals in household \( h \in \mathcal{H}_c \) by \( \mathcal{I}_{ch} \). Let \( y_{chi}^{(k)} \) be the \( k \)-th \( (1 \leq k \leq K) \) anthropometric indicator of our interest for the individual \( i \) in cluster \( c \) and household \( h \). In our application, \( K = 2 \) with \( k = 1 \) and \( k = 2 \) being the standardized height and weight respectively. Our
goal is to find an estimate of the aggregate index $W_{Y}^{(k)} = W(\{y_{i}^{(k)}\}_{i \in \mathcal{V}})$ such as the prevalence of malnutrition for a set of individuals $\mathcal{V}$. $y_{\text{ch}_{i}}^{(k)}$ is related to a $d^{(k)}$-vector of observable characteristics, $x_{\text{ch}_{i}}^{(k)}$, through the following anthropometric model.

$$y_{\text{ch}_{i}}^{(k)} = [x_{\text{ch}_{i}}^{(k)}]^{T} \beta^{(k)} + u_{\text{ch}_{i}}^{(k)}$$

where $\beta^{(k)}$ is a $d^{(k)}$-vector of parameters and $u_{\text{ch}_{i}}^{(k)}$ is a disturbance term. $u_{\text{ch}_{i}}^{(k)}$ satisfies $E[u_{\text{ch}_{i}}^{(k)}x_{\text{ch}_{i}}^{(k)}] = 0$ for all $c$, $h$, $i$ and $k$. Let $C \equiv \#(C)$, $H_{c} \equiv \#(H_{c})$, and $I_{ch} \equiv \#(I_{ch})$, where $\#(\cdot)$ is the counting measure. The total number of observations is $N \equiv \sum_{c \in C} \sum_{h \in H_{c}} I_{ch}$ and each cluster has a weight $w_{c}$ which is normalized so that $\sum_{c} w_{c} = 1$. Recall here that the number of children under five in one household is usually small and can be 1 for some households.

As with the ELL study, we allow for location, or cluster-specific, effect and heteroskedastic household-specific effect. In addition, we have multiple indicators and the individual effect correlated across the indicators. Hence, $u_{\text{ch}_{i}}^{(k)} = \eta_{c}^{(k)} + \epsilon_{ch}^{(k)} + \delta_{\text{ch}_{i}}^{(k)}$, where $\eta_{c}^{(k)}$, $\epsilon_{ch}^{(k)}$, and $\delta_{\text{ch}_{i}}^{(k)}$ are respectively the location, household and individual effect. In principle, each of the three components of $u_{\text{ch}_{i}}^{(k)}$ could be heteroskedastic and correlated across the indicators. The particular choice we make is driven by the insights from previous studies and also by the limitation imposed by data. For example, given that $I_{ch} = 1$ for many households in the CDHS data, it is extremely difficult to distinguish household effects from individual effects if both are heteroskedastic. Cluster-level heteroskedasticity is also difficult to estimate because of the limited number of clusters. We allowed for flexibility in the correlational structure of the disturbance terms where such flexibility seems most crucial.

We shall hereafter denote $(u^{(1)}, \ldots, u^{(K)})^{T}$ by $\tilde{u}$, and use similar notation for other variables. We assume that $\bar{\eta}_{c}, \bar{\epsilon}_{ch}$ and $\bar{\delta}_{\text{ch}_{i}}$ are uncorrelated and satisfy $E[\bar{\eta}_{c}] = E[\bar{\epsilon}_{ch}] = E[\bar{\delta}_{\text{ch}_{i}}] = \bar{O}_{K}$, where the last term is the $K$-vector of zeros. For $l, k \in \{1, \ldots, K\}$ with $l \neq k$, we assume $E[\eta_{c}^{(k)} \eta_{c}^{(l)}] = E[\epsilon_{ch}^{(k)} \epsilon_{ch}^{(l)}] = 0$. We denote the variance of each component of the disturbance
term as $(\sigma^2_{\eta}) = E[\eta^2]$, $(\sigma^2_{\epsilon, ch}) = E[\epsilon^2_{ch}]$, and $(\sigma^2_{\delta}) = E[\delta^2_{ch}]$. Note that we need subscripts $ch$ to express the heteroskedasticity of the household effect. Intra-personal correlations are denoted as $(\sigma^2_{\delta}^{(k)}) = (\sigma^2_{\delta}^{(k, \ell)}) \equiv E[\delta_{ch}^{(k)} \cdot \delta_{ch}^{(\ell)}]$. We shall denote simple means by dots such as $\bar{u}_{ch}^{(k)} = \frac{1}{n_{ch}} \sum_{i \in I_{ch}} u_{ch}^{(k)}$ and $\bar{u}_{c}^{(k)} = \frac{1}{N_c} \sum_{h \in H_c} u_{ch}^{(k)}$.

As discussed above, we estimate the distribution parameters of the estimate of the regression coefficients and disturbance term. To do so, we run an OLS for indicator $k$, and get the residual $\hat{u}_{ch}^{(k)}$. Letting $H_c \equiv \{h \in H | n_{ch} > 1\}$, $H \equiv \#\{H_c\}$, $C \equiv \{c \in C | H_c > 0\}$, and $\bar{w}_{c} \equiv \frac{w_c}{\sum_{c \in C} w_c}$, straightforward calculations give

$$E \left[ \sum_{c \in C} \frac{\bar{u}_{c}}{H_c} \sum_{h \in H_c} \sum_{i \in I_{ch}} \frac{(u_{ch}^{(k)} - \bar{u}_{ch}^{(k)})^2}{n_{ch} - 1} \right] = (\sigma^2_{\delta})$$

and

$$E \left[ \sum_{c \in C} \frac{\bar{u}_{c}}{H_c} \sum_{h \in H_c} \sum_{i \in I_{ch}} \frac{(u_{ch}^{(k)} - \bar{u}_{ch}^{(k)}) \cdot (u_{ch}^{(l)} - \bar{u}_{ch}^{(l)})}{n_{ch} - 1} \right] = (\sigma^2_{\delta})^{(k, \ell)}.$$

We obtain consistent estimators $(\sigma^2_{\delta})$ and $(\sigma^2_{\delta})^{(k, \ell)}$ by taking out the expectations operator on the left-hand-side and replacing $u$ by $\hat{u}$ in the respective equation. In the same manner, we obtain a consistent estimator $(\sigma^2_{\delta})$ of the variance of location effect using the following formula (proof in Appendix).

$$E \left[ \sum_{c \in C} \frac{\bar{w}_{c}H_c(u_{c}^{(k)})^2}{\sum_{c \in C} \bar{w}_{c}H_c} \frac{w_c}{\sum_{c \in C} \bar{w}_{c}} \frac{\sum_{h \in H_c} (u_{ch}^{(k)})^2}{n_{ch} - 1} \right] = (\sigma^2_{\delta})^{(k)}$$

There is no guarantee that the estimated variance of the location effect is non-negative, and hence we censor $(\sigma^2_{\delta})$ at zero.

The household and individual effects are difficult to separate from each other because $I_{ch}$ is small for a majority of households. To estimate the distribution parameters of the household effect, it is useful to work with the sum $(s_{ch}^{(k)})^2 = (\sigma_{\epsilon, ch}^{(k)})^2 + (\sigma_{\delta}^{(k)})^2$ of the household and individual effects. Note that the heteroskedasticity of $(s_{ch}^{(k)})^2$ comes only from the household
effect. We can show the following formula (proof in Appendix) for \( c \in C^* \equiv \{ \theta \in C | H_{\theta} > 2 \} :$

\[
E \left[ \frac{H_c \cdot (u_{ch}^{(k)} - u_{c..}^{(k)})^2}{H_c - 2} - \frac{\sum_{h' \in H_c} (u_{ch'}^{(k)} - u_{c..}^{(k)})^2}{(H_c - 1)(H_c - 2)} \right] + \frac{I_{ch} - 1}{I_{ch}} (\sigma_{ch}^{(k)})^2 = (\hat{s}_{ch}^{(k)})^2
\]  

(2)

We let \( \hat{s}_{ch}^2 \) be the left hand side of the equation above with the expectation operator removed and with \( u \) and \( \sigma_{\delta} \) replaced by \( \hat{u} \) and \( \hat{\sigma}_{\delta} \) respectively.\(^4\) For the heteroskedastic model, we propose a following logistic heteroskedastic model similar to the one in Elbers et al. (2003).

\[
\ln \frac{(\hat{s}_{ch}^{(k)})^2 - B^{(k)}}{A^{(k)} + B^{(k)} - (\hat{s}_{ch}^{(k)})^2} = [Z_{ch}^{(k)}]^T \alpha^{(k)} + \tau_{ch}^{(k)}
\]

, where \( A^{(k)} \) and \( B^{(k)} \) are the maximum and minimum of \((\hat{s}_{ch}^{(k)})^2\), and \( Z_{ch}^{(k)} \) is vector of household characteristics, \( \alpha^{(k)} \) heteroskedastic regression coefficient and \( \tau_{ch}^{(k)} \) the residual term. The feature of this formulation is that \((\hat{s}_{ch}^{(k)})^2\) is both upper- and lower-bounded. For the estimation of \( \alpha^{(k)} \), we use \( B^*_s = \min \{ 0, 1.05 \cdot \min_{ch} \{ (\hat{s}_{ch}^{(k)})^2 \} \} \) and \( A^{(k)} = 1.05 \cdot (\max_{ch} \{ (\hat{s}_{ch}^{(k)})^2 \} - B^*_s) \). The use of delta method suggests the following estimate of \((\sigma_{ch}^{(k)})^2\).

\[
(\hat{\sigma}_{ch}^{(k)})^2 = \max \left\{ 0, \left[ \frac{D^{(k)}}{1 + D^{(k)}} + \frac{D^{(k)}(1 - D^{(k)})}{2(1 + D^{(k)})^2} (\hat{\sigma}_{\delta}^{(k)})^2 \right] A^{(k)} + B^*_s - (\hat{\sigma}_{\delta}^{(k)})^2 \right\}
\]

(3)

, where \( D^{(k)} \equiv \exp([Z_{ch}^{(k)}]^T \hat{\alpha}^{(k)}) \), and \((\hat{\sigma}_{\delta}^{(k)})^2\) is the estimated variance of \( \tau_{ch} \). The max\{\cdot,\cdot\} function is introduced to ensure the non-negativity of \((\hat{\sigma}_{ch}^{(k)})^2\). The consequence of this is that \((\sigma_{ch}^{(k)})^2\) may be upward-biased. Hence, the standard errors for the estimates of nutrition measures at the level of small geographic areas are conservative. We can now estimate the Omega matrix, and carry out a (feasible) GLS regression to obtain the regression coefficients.

\(^4\)When \((\hat{\sigma}_{\delta}^{(k)})^2 = 0\), we use \( E[(u_{ch}^{(k)})^2] = (\sigma_{ch}^{(k)})^2 + \frac{(\hat{\sigma}_{\delta}^{(k)})^2}{I_{ch}} \) instead.
\( \hat{\beta} \) and \( \overline{\text{var}}[\hat{\beta}] \) for all the indicators at once.

We then find the empirical distributions of each disturbance component. We approximate the distributions of \( \eta_c, \epsilon_{ch} \), and \( \delta_{ch} \) respectively by the distributions of \( u_c \), \( u_{ch} - u_c \), and \( u_{ch} - u_{ch} \) standardized to have mean zero and a unit standard error.

**Simulation** We explicitly take into account the model and idiosyncratic errors by Monte-Carlo simulation. Let \( R \) be the number of simulations, which must be sufficiently large in order to make the computational errors small enough. In our study, we set \( R = 100 \). In \( r \)-th simulation where \( r \in \{1, \ldots, R \} \), we need the following parameters: \( \hat{\beta}_{(r)} \), \( \hat{\alpha}_{(r)} \), \( \hat{\sigma}_{\tau,(r)} \), \( \hat{\sigma}_{\delta,(r)} \), \( \hat{\sigma}_{\eta,(r)} \), \( \hat{\sigma}_{\epsilon,(r)} \), \( A_{u_{(r)}} \), \( B_{u_{(r)}} \) for all \( k \neq l \). We randomly draw \( \hat{\alpha}_{(r)} \) and \( \hat{\beta}_{(r)} \) from the normal distribution with mean \( \hat{\alpha} \) and \( \hat{\beta} \), and variance-covariance matrix \( \overline{\text{var}}[\hat{\alpha}] \) and \( \overline{\text{var}}[\hat{\beta}] \) respectively. For the rest of the parameters, we created a two-stage bootstrapping sample of \( \hat{u} \) in each round of simulation, and computed the parameters using the bootstrapping sample.\(^5\) It is straightforward to calculate \( \hat{\sigma}_{\epsilon, ch,(r)} \), \( \hat{A}_{u_{(r)}} \), \( \hat{B}_{u_{(r)}} \) from Eq(3).

For each census record, we draw the standardized cluster, household and individual effects. We use a two stage-draw for the cluster and household effects, but the individual effect is drawn from the entire survey sample.\(^6\)

Note that we use the same survey record for all the indicators to retain the correlation across indicators. Let us write the drawn standardized component as \( \tilde{\epsilon}_{c,(r)} \), \( \tilde{\epsilon}_{ch,(r)} \), and \( \tilde{\delta}_{ch} \). Then, the \( k \)-th imputed anthropometric indicator for the individual in census in \( r \)-th simulation is:\(^7\)

\[
\tilde{y}^{(k)}_{ch,i,(r)} = X^{(k)}_{ch,i} \tilde{\beta}_{(r)} + \tilde{\eta}^{(k)}_{c,(r)} \cdot \tilde{\alpha} + \tilde{\epsilon}^{(k)}_{ch,(r)} \cdot \tilde{\sigma}_{\epsilon, ch,(r)} + \tilde{\delta}_{ch} \cdot \tilde{\delta}_{(r)}
\]

\(^5\)Another possible implementation is to draw \( \hat{\alpha} \) and \( \hat{\beta} \) from the bootstrapping sample.

\(^6\)We did not draw the individual effects for two or more children in the same census households from the same survey household because the number of children under five in the household is small.

\(^7\)To eliminate extreme values, we drop census observations for which the point estimate \( X^{(k)}_{ch,i} \beta^{(k)} \) is not in the range of \( y^{(k)} \) for at least one indicator. In each simulation, we censored \( y^{(k)}_{ch,i,(r)} \) at the minimum and maximum observed in the survey.
This allows us to calculate the health indicator of our interest as $\mathbf{W}_{[r],\nu}^{[k]} = \mathbf{W}(\{\tilde{y}_{i,[r]}^{(k)}\}_{i \in \nu})$
in $r$-th simulation. The point estimate and its associated standard error are estimated at the mean and standard deviation of $\mathbf{W}_{[r]}^{[k]}$ taken over the subscript $r$.

4 Data

Our basic building blocks comprise a survey dataset, a census dataset and a dataset of geographic variables. The survey dataset we used is CDHS 2000, which was designed to collect health and demographic information for the Cambodian population, with a particular focus on women of childbearing age and young children. The sample covered 12,236 households in 17 strata across the country. In addition to detailed information about each household, its members, and housing characteristics, one half of these households were systematically selected to participate in the anthropometric data collection. All children under 60 months of age in the sub-sampled households were weighed and measured. After excluding children for which information on height or weight is missing or implausible, 3,596 observations were used for this analysis (For further details, see National Institute of Statistics et al. (2001)). We first derived the height-for-age and weight-for-age z-score measures, which were then converted to the height and weight for 24 month-old females with the same z-score measures.

The second source of data was the Cambodian National Population Census, the first population census to be conducted in Cambodia since 1962. The census covered virtually all persons staying in Cambodia at the reference time of midnight of March 3, 1998.\(^8\) The census data contains information on housing characteristics as well as information on each usual household member and visitors present on the reference night, including the relationship to the head of household, sex, age, marital status, migration, literacy, education and employment. The census also contained questions on fertility of females aged 15 and over.

\(^8\)Due to military operations, about 0.5 percent of the population was not covered.
and infant mortality. The census dataset contains about 1.5 million records of children under five.

A set of geographic indicators was also used in this analysis. Because Cambodia has a rich collection of geographic data, indicators on a range of characteristics could be generated. These indicators included distance calculations, land use and land cover information, climate indicators, vegetation, agricultural production and flooding. A number of datasets from various sources were compiled into a GIS and these indicators were generated for all villages and communes in Cambodia. Very coarse resolution data was summarized at the commune-level, while high resolution data was attributed to individual villages. Distances from villages to roads, other towns, health facilities, and major rivers were calculated from the center of the villages. Indicators based on satellite data with varying temporal resolutions included land use within the commune (agricultural, urban, forested, etc.), a vegetation greenness indicator to proxy agricultural productivity, and the degree to which the area was lit by nighttime lights as a proxy of urbanization. Relatively stable indicators including soil quality, elevation, and various 30-year average climate variables were derived from other composite datasets. We have also generated the village-level means from the census data. It should be noted that the village-level means do not have to be taken from the variables that also exist in the CDHS dataset. This is because the village-level means, as with other geographic variables, can be linked to both the census and the survey datasets. Inclusion of these geographic variables and their cross terms with other individual-level and household-level have improved substantially the ability to fit the data.

5 Results

We constructed an anthropometric model in each of the five zones (“ecozones”) of Urban, Plain, Tonlesap, Coastal and Plateau. We combined provinces that are similar in agro-
climatic and socio-cultural characteristics because some of the strata had too few observations to carry out meaningful analysis. We ran regressions in each ecozone separately. While individual-level and household-level variables can explain only around twenty to thirty percent of the variation in the standardized height and weight, we were able to increase the explanatory power of the model to about forty to sixty percent by including geographic variables and interaction terms. This observation seems to be consistent with Curtis and Hossain (1998). We checked the robustness of the regression coefficients by randomly dropping some households or clusters as was done in Elbers et al. (2002).

The GLS regression results for the Coastal ecozone are presented in Table 4 and Table 5 in the Appendix. The point estimate of the variance of the location effect was zero in all strata, even though in some rounds of the simulation $(\hat{\sigma}_{\eta(r)}^{(k)})^2$ was strictly positive because of bootstrapping. The average proportion of individual effect to the sum of all the disturbance components was found very high with $\frac{(\hat{\sigma}_{\hat{\delta}}^{(1)})^2}{(\hat{\sigma}_{\hat{\delta}}^{(1)})^2}$ ranging from 0.81 to 0.99 and $\frac{(\hat{\sigma}_{\hat{\delta}}^{(2)})^2}{(\hat{\sigma}_{\hat{\delta}}^{(2)})^2}$ ranging from 0.66 to 0.92. This in turn means that the household effect is relatively small, though it is in general not negligible. In all the ecozones, the magnitudes of intrapersonal correlations as measured by $\frac{\hat{\delta}_{\hat{\delta}}^{(1,2)}}{\hat{\delta}_{\hat{\delta}}^{(1)}\hat{\delta}_{\hat{\delta}}^{(2)}}$ were found high, ranging from 0.44 to 0.53. This strongly suggests the importance of the inclusion of intrapersonal correlation.

After the predictions for standardized height and weight for each child in the census were made in each round of simulation, they were aggregated to the commune level in Cambodia to arrive at the prevalence of stunting and underweight. Due to missing data in the census and geographic datasets for a small number of communes, we obtained commune-level estimates for a total of 1,594 communes out of the 1,616 communes in Cambodia. By linking the estimates to the shape file, the estimates were converted into maps as shown in Figure 1. They show the point estimates of the prevalence of stunting and underweight as of the census year 1998 at the commune level, and the darker areas represent a worse situation. For

---

9 We followed the definition of the ecozones previously used by WFP (2001).

10 Other regression results are omitted to save space. They are available from the author upon request.
example, most densely populated parts of Cambodia surrounding Phnom Penh, the provinces of Kandal, Prey Veaeng, Svay Rieng, Kampong Cham, and Kampong Chhnang exhibit large areas of high prevalence of stunting.

While these maps are presented in a user friendly format, they may be misleading as such presentation does not take into account the fact that the commune-level estimates are subject to statistical errors. One could also present the maps in terms of the difference between the commune-level estimate and a reference level such as the national average divided by the standard error of the commune-level estimate. Such presentation emphasizes how significantly different the prevalence of malnutrition in each commune is from the national average. Another way to present the results is to depict the density of malnourished children. Such a representation may be most relevant when a proposed policy intervention is likely to benefit the target location as a whole. Construction of clinics might be an example of such a project. What format is most appropriate would depend on the purpose and intended audience of the map, but it is essential that the decision makers understand the meaning of the estimates and representation.

To evaluate the reliability of the estimates, the ecozone level CDHS prevalence of stunting and underweight were compared with the ecozone-level estimates in this study. Table 1 summarizes these results. The differences are within two standard errors of the CDHS estimates suggesting that the differences between the CDHS estimates and this study can be attributed to the random errors. It should also be noted from Table 1 that the standard errors from DHS Only are higher than the DHS+CENSUS except for the underweight model in Tonlesap. The standard errors from Coastal ecozone in the survey is quite high, and our methodology improved the ecozone level estimates quite substantially. We also arrive at the same conclusion when comparison is made at the level of 17 strata.

We also checked the magnitude of standard errors associated with commune-level estimates of prevalence of malnutrition. The simple mean of the standard are 9.0 percent for
Figure 1: Commune-level Prevalence of Stunting (top) and Underweight (bottom) for Cambodia.
Table 1: Comparison of the stratum-level estimates. Standard error for DHS Only was calculated by 100-time two-stage bootstrapping.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Stratum</th>
<th>DHS Only</th>
<th>DHS+CENSUS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SE</td>
</tr>
<tr>
<td>Stunting</td>
<td>Urban</td>
<td>37.89</td>
<td>3.30</td>
</tr>
<tr>
<td></td>
<td>Plain</td>
<td>47.58</td>
<td>2.49</td>
</tr>
<tr>
<td></td>
<td>Tonsap</td>
<td>43.23</td>
<td>2.09</td>
</tr>
<tr>
<td></td>
<td>Coastal</td>
<td>47.21</td>
<td>5.52</td>
</tr>
<tr>
<td></td>
<td>Plateau</td>
<td>47.10</td>
<td>2.99</td>
</tr>
<tr>
<td>Underweight</td>
<td>Urban</td>
<td>39.58</td>
<td>2.93</td>
</tr>
<tr>
<td></td>
<td>Plain</td>
<td>47.80</td>
<td>2.44</td>
</tr>
<tr>
<td></td>
<td>Tonsap</td>
<td>45.84</td>
<td>1.95</td>
</tr>
<tr>
<td></td>
<td>Coastal</td>
<td>38.95</td>
<td>5.28</td>
</tr>
<tr>
<td></td>
<td>Plateau</td>
<td>46.37</td>
<td>3.87</td>
</tr>
</tbody>
</table>

stunting and 10.1 percent for underweight. The simple average of the coefficient of variation 3.8 percent for stunting and 4.2 percent for underweight. They are reasonably small as they are about the same magnitude as the standard errors in the survey at the ecozone level. We should note, however, that there are communes for which the standard errors are quite high, as the maximum standard error was 22.7 percent for stunting and 18.5 percent for underweight. At the commune-level, the correlation between the estimates of prevalence of stunting and underweight was 0.33. Overall, the accuracy of commune-level estimates for stunting and underweight is about the same.

6 Application

6.1 Correlation between Poverty, Inequality and Malnutrition

We illustrate an application of the commune-level estimate to the analysis of the relationship between poverty, inequality and malnutrition. Consumption and anthropometric indicators both measure certain aspects of welfare, and we would expect that the poverty rate
is positively correlated with the prevalence of malnutrition. While it is indeed the case for underweight, the unweighted Pearson correlation was only 0.11. Moreover, we found that the correlation the prevalence of stunting and poverty rate is not significantly different from zero. We arrive at the same conclusion with the Spearman ranking correlation. Though the correlation is underestimated because we cannot correctly take into account the error term that are correlated between consumption and nutrition measures, the low-level of correlation suggests that there are important factors associated with nutrition outcome that are not correlated with consumption. Such factors may include child-care practices and the prevalence of diseases such as malaria.

The correlation between consumption inequality and the prevalence of malnutrition, however, would be more immune to such underestimation. We found that the prevalence of malnutrition is significantly and negatively correlated with consumption inequality measures such as the Gini coefficient and the Theil’s L index (GE(0)). This holds for both stunting and underweight. We also regressed malnutrition indicators on poverty rate and a inequality measure, and found that the coefficient on the inequality measure is negative and significant. Table 2 shows the results for the Theil’s L index. We ran the same regressions with the provincial-level dummies, and the conclusion was the same. This is a surprising result because inequality has been generally associated with worse health outcome in various geographic locations and various level of aggregation (See, for example, Wilkinson (1996); Kawachi et al. (1997)), even though a majority of the studies use mortality as an indicator of health outcome.

What we found from this analysis seems to have an important implication for the current debate on the significance of inequality on the health outcome. Some researchers in public health, including Wilkinson (1996), have argued that, even after controlling for confounding factors, higher income inequality leads to worse health outcome partly because income inequality leads to higher levels of psychosocial stress. Economists have been skeptical about
this line of argument (Deaton, 2003; Wagstaff and van Doorslaer, 2000).

In Cambodia, we also found unattractive the argument that income inequality harms health because we simply cannot explain the negative correlation between the prevalence of malnutrition and poverty rate. However, if consumption can improve children’s nutrition status only above a certain threshold, the negative correlation can be easily explained. What determines the prevalence of malnutrition in this case is the percentage of those above the threshold. If the mean is well below the threshold and the consumption distribution is unimodal, increasing mean consumption keeping the inequality constant may not change the prevalence of malnutrition much. On the other hand changing inequality keeping the mean consumption may decrease the prevalence of malnutrition substantially, even though this argument depends on the shape of the distribution of consumption.

6.2 Decomposition analysis of health inequality

While the relationship of consumption inequality with the prevalence of malnutrition is of interest, the inequality of child health per se, as measured by nutrition outcomes, is also interesting to look at. Health inequality captures a certain aspect of the inequality of basic capabilities consumption-based poverty measure inadequately captures. Hence the existence of child health inequality in itself can be of concern to policy makers, regardless of consumption inequality, as health is an important aspect of human welfare. This is even so given

Table 2: Commune-level regression of malnutrition indicators on consumption poverty and inequality. Regressions were run separately. N=1594.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Stunting Coef</th>
<th>T-value</th>
<th>Underweight Coef</th>
<th>T-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.50</td>
<td>71.02</td>
<td>0.44</td>
<td>63.85</td>
</tr>
<tr>
<td>GE(0)</td>
<td>-0.13</td>
<td>-6.00</td>
<td>-0.09</td>
<td>-4.65</td>
</tr>
<tr>
<td>Poverty Rate</td>
<td>-0.02</td>
<td>-1.38</td>
<td>0.03</td>
<td>3.08</td>
</tr>
</tbody>
</table>

21
that some studies found child health indicators are important in predicting the future social standing of the child (Nystrom Peck, 1992; Montgomery et al., 1996).

In what follows, we shall focus on the decomposition analysis of inequality indicators. The analysis we carry out is different from most of other studies on health inequality in at least three respects. First, despite the growing literature on health inequality, there have been relatively a few studies carried out at a spatially disaggregated level, especially in developing countries. Our methodology allows one to analyze health inequality at a spatially disaggregated level even in the absence of a large number of observations of health indicators collected at that level.

Second, while the majority of health inequality literature use mortality data, we use the standardized height and weight. Mortality has the advantage of being objective, and often relatively cheap to collect in developed countries. On the other hand, it may not be the best indicator at a very disaggregated level, because death is a rare event. Also, the probability of death is influenced by what the individual experiences over the course of their lives. This makes it difficult to compare populations with different age compositions. Standardized height and weight for children have some advantages here. As with the mortality, they are objective. Even though these indicators are still influenced by the experience of children over the course of their lives, comparison across different populations is easier. Pradhan et al. (2003) discusses additional advantages of the standardized height measure.11

Third, we emphasize direct policy implications of inequality for targeting. In much of health inequality studies, the focus is placed on the different health outcomes for different groups defined by the socioeconomic status (SES). For the formulation and targeting of nutrition policy, such analysis may not be of great use unless we can target resources on the basis of the SES. Even if it is possible, one can still combine geographic targeting with a SES targeting.

---

11Pradhan et al. (2003) prefers height to weight because too much weight is not good for health. Given that less than 1 percent of the children under five is overweight in this study, this is a minor concern.
Let us now decompose the health inequality. We took an approach similar to the one proposed by Pradhan et al. (2003). They apply the Theil’s L index to the standardized height to decompose the world health inequality into the between-country component and the within-country component. They adjust the index to take into account of the natural variation in height. They found that around 31 percent of the total inequality is due to the between country component.

Table 3 provides the decomposition results for different measures. We found that, regardless of the parameter of the general entropy measure, the share of the between-group component was stable. When we adjust GE(0) for the natural inequality, we have a figure that is comparable to the world inequality of 31 percent. Compared with the global health inequality, the share of the between-group inequality is low in Cambodia.
<table>
<thead>
<tr>
<th>Indicator</th>
<th>Decomposable Index</th>
<th>Cambodia</th>
<th>Province</th>
<th>District</th>
<th>Commune</th>
<th>Unit-record</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standardized Height</td>
<td>FGT(0)</td>
<td>0.00 (0.00)</td>
<td>0.95 (0.29)</td>
<td>1.92 (0.31)</td>
<td>4.23 (0.45)</td>
<td>6.28 (0.50)</td>
</tr>
<tr>
<td></td>
<td>FGT(1)</td>
<td>0.00 (0.00)</td>
<td>0.95 (0.31)</td>
<td>2.10 (0.38)</td>
<td>4.98 (0.70)</td>
<td>7.22 (0.75)</td>
</tr>
<tr>
<td></td>
<td>FGT(2)</td>
<td>0.00 (0.00)</td>
<td>0.73 (0.24)</td>
<td>1.70 (0.32)</td>
<td>4.31 (0.75)</td>
<td>6.40 (0.81)</td>
</tr>
<tr>
<td></td>
<td>GE(-1)</td>
<td>0.00 (0.00)</td>
<td>1.12 (0.35)</td>
<td>2.55 (0.42)</td>
<td>6.62 (0.87)</td>
<td>9.37 (0.96)</td>
</tr>
<tr>
<td></td>
<td>GE(-0.5)</td>
<td>0.00 (0.00)</td>
<td>1.12 (0.35)</td>
<td>2.55 (0.42)</td>
<td>6.65 (0.89)</td>
<td>9.40 (0.97)</td>
</tr>
<tr>
<td></td>
<td>GE(0)</td>
<td>0.00 (0.00)</td>
<td>1.12 (0.36)</td>
<td>2.55 (0.42)</td>
<td>6.67 (0.90)</td>
<td>9.43 (0.98)</td>
</tr>
<tr>
<td></td>
<td>GE(0.5)</td>
<td>0.00 (0.00)</td>
<td>1.12 (0.36)</td>
<td>2.55 (0.42)</td>
<td>6.69 (0.91)</td>
<td>9.46 (1.00)</td>
</tr>
<tr>
<td></td>
<td>GE(1)</td>
<td>0.00 (0.00)</td>
<td>1.12 (0.35)</td>
<td>2.54 (0.41)</td>
<td>6.70 (0.92)</td>
<td>9.47 (1.01)</td>
</tr>
<tr>
<td>Standardized Weight</td>
<td>FGT(0)</td>
<td>0.00 (0.00)</td>
<td>0.75 (0.20)</td>
<td>1.85 (0.27)</td>
<td>3.92 (0.47)</td>
<td>6.48 (0.57)</td>
</tr>
<tr>
<td></td>
<td>FGT(1)</td>
<td>0.00 (0.00)</td>
<td>0.78 (0.24)</td>
<td>2.04 (0.34)</td>
<td>4.31 (0.70)</td>
<td>7.01 (0.81)</td>
</tr>
<tr>
<td></td>
<td>FGT(2)</td>
<td>0.00 (0.00)</td>
<td>0.59 (0.21)</td>
<td>1.52 (0.30)</td>
<td>3.20 (0.62)</td>
<td>5.49 (0.71)</td>
</tr>
<tr>
<td></td>
<td>GE(-1)</td>
<td>0.00 (0.00)</td>
<td>1.02 (0.27)</td>
<td>2.51 (0.37)</td>
<td>5.54 (0.72)</td>
<td>8.79 (0.88)</td>
</tr>
<tr>
<td></td>
<td>GE(-0.5)</td>
<td>0.00 (0.00)</td>
<td>1.05 (0.28)</td>
<td>2.58 (0.38)</td>
<td>5.71 (0.74)</td>
<td>9.05 (0.90)</td>
</tr>
<tr>
<td></td>
<td>GE(0)</td>
<td>0.00 (0.00)</td>
<td>1.08 (0.29)</td>
<td>2.64 (0.39)</td>
<td>5.85 (0.76)</td>
<td>9.27 (0.93)</td>
</tr>
<tr>
<td></td>
<td>GE(0.5)</td>
<td>0.00 (0.00)</td>
<td>1.09 (0.29)</td>
<td>2.68 (0.40)</td>
<td>5.96 (0.77)</td>
<td>9.45 (0.95)</td>
</tr>
<tr>
<td></td>
<td>GE(1)</td>
<td>0.00 (0.00)</td>
<td>1.11 (0.30)</td>
<td>2.71 (0.40)</td>
<td>6.04 (0.78)</td>
<td>9.58 (0.97)</td>
</tr>
<tr>
<td>Consumption</td>
<td>FGT(0)</td>
<td>0.00 (0.00)</td>
<td>7.10 (0.50)</td>
<td>12.00 (0.71)</td>
<td>19.06 (0.88)</td>
<td>37.14 (1.16)</td>
</tr>
<tr>
<td></td>
<td>FGT(1)</td>
<td>0.00 (0.00)</td>
<td>8.20 (0.86)</td>
<td>14.29 (1.36)</td>
<td>23.15 (1.76)</td>
<td>46.72 (2.01)</td>
</tr>
<tr>
<td></td>
<td>FGT(2)</td>
<td>0.00 (0.00)</td>
<td>7.20 (1.00)</td>
<td>12.98 (1.81)</td>
<td>21.53 (2.36)</td>
<td>45.87 (2.82)</td>
</tr>
<tr>
<td></td>
<td>GE(-1)</td>
<td>0.00 (0.00)</td>
<td>15.20 (11.07)</td>
<td>24.22 (10.47)</td>
<td>36.56 (9.26)</td>
<td>57.87 (7.22)</td>
</tr>
<tr>
<td></td>
<td>GE(-0.5)</td>
<td>0.00 (0.00)</td>
<td>17.64 (14.30)</td>
<td>27.49 (13.48)</td>
<td>41.01 (11.33)</td>
<td>62.17 (7.55)</td>
</tr>
<tr>
<td></td>
<td>GE(0)</td>
<td>0.00 (0.00)</td>
<td>18.69 (16.32)</td>
<td>28.84 (15.84)</td>
<td>43.19 (13.16)</td>
<td>64.14 (8.68)</td>
</tr>
<tr>
<td></td>
<td>GE(0.5)</td>
<td>0.00 (0.00)</td>
<td>17.04 (12.66)</td>
<td>27.18 (14.47)</td>
<td>42.25 (12.99)</td>
<td>63.66 (9.09)</td>
</tr>
<tr>
<td></td>
<td>GE(1)</td>
<td>0.00 (0.00)</td>
<td>11.71 (2.59)</td>
<td>19.91 (3.54)</td>
<td>34.42 (4.06)</td>
<td>57.24 (6.12)</td>
</tr>
</tbody>
</table>

| Number of Observations | 1     | 24    | 180   | 1594  | 1424907/2130544 |

Table 3: Decomposition results. Unit record for standardized height and weight is individual and it is household for consumption. Standard errors in parenthesis.
As Pradhan et al. (2003) acknowledge, the choice of the reference population of 24-month old girl is arbitrary, and the decomposition results is influenced by this choice. It is not clear, therefore, whether the apparently high levels of between group effects for consumption is real. To address this issue, we propose another way of decomposing the health inequality into between group and within group effects. Before moving on, one should note that the prevalence of malnutrition is analogous to the FGT measure of poverty. Hence thinking of the height or weight corresponding to the z-score of -2 as the poverty line, we can define FGT measures. Hence, letting \( y \) be the FGT measure of interest, we can decompose the variance of \( y \) into between-group and within-group effects as follows:

\[
\frac{1}{N} \sum_{g \in G} \sum_{i \in g} (y_{gi} - \bar{y})^2 = \frac{1}{N} \sum_{g \in G} \sum_{i \in g} (y_{gi} - \bar{y}_g)^2 + \frac{1}{N} \sum_{g \in G} N_g (\bar{y}_g - \bar{y})^2
\]

\[
\bar{y} = \frac{1}{N} \sum_{g \in G} \sum_{i \in g} y_{gi}
\]

\[
\bar{y}_g = \frac{1}{N_g} \sum_{i \in g} y_g
\]

where \( G \) is the set of groups, and \( N \) and \( N_g \) are respectively the population and the number of people in group \( g \). In this formulation, the proportion of the between-group component of the total variance is constant with respect to an affine transformation. Hence it does not suffer from the arbitrary choice of reference group and makes clearer the meaning of comparison across different indicators.

One should note that this measure does not look at the entire distribution. Whether this is desirable property or not would depend on the purpose of the index. When we are interested in the targeting policy, we are more interested in the lower tail of the distribution. For example, very tall child getting taller by “stealing height” from a little less taller child would be much less concerning than a very short child getting shorter by having his height stolen from a less shorter child.

We looked at the between-group component of the variance of FGT measures with pa-
rameter 0, 1 and 2. Table 3 shows the results. As with the generalized entropy measures, the between group health inequality is lower for standardized height and weight than consumption. For both anthropometric and consumption measures, the proportion of between-group inequality within Cambodia is much lower the proportion of between-country inequality in the world.

6.3 Concentration curve and Targeting

Another approach we can take to decompose the inequality is by way of the concentration curve. The concentration curve is similar to the Lorenz curve, but is based on the group individual belongs to. Let \( G = \{ g_1, g_2, \ldots, g_j, \ldots, g_M \} \). With a little abuse of notation, we write \( \bar{y}_{g_j} \equiv \tilde{y}_j \) and \( \bar{N}_{g_j} \equiv \tilde{N}_j \). Let \( a_k \equiv \sum_{j=1}^{k} \frac{N_j}{N} \) and \( b_k \equiv \sum_{j=1}^{k} \frac{\bar{y}_j - \bar{N}_j}{\bar{y} - \bar{N}_j} \). In words, \( a_g \) mean the share of the cumulative population in a group numbered at most \( k \), and, similarly, \( b_g \) mean the ratio of cumulative FGT up to group \( k \) over the total FGT in the population. The concentration curve is drawn by connecting \((0, 0)\) and \((a_j, b_j)\) for \( j = 1, \ldots, M \). Twice the area between the concentration curve and the 45-degree line is called the concentration index. The concentration curve and the concentration index have been often used for analyzing the health inequality across groups defined by SES or income (Wagstaff et al., 1991; van Doorslaer et al., 1997; Wagstaff et al., 2003). In this study, we order the areas in the descending order of the prevalence of malnutrition.

Figure 2 gives an example of a concentration curve. The horizontal axis measures the population share whereas the vertical axis measures the share of FGT measure. If last geographic unit \( g_M \) has malnourished children (i.e. \( \bar{y}_M \leq 0 \)), the concentration curve hits point C. In this case the concentration curve looks like the dashed line \( OC \). The concentration index in this case is twice the size of \( Z \), or the area defined by the dashed line and the line \( \overline{OC} \). In the graph, \( \overline{OD} \) is the concentration curve when targeted defined at the individual level. This corresponds to the Lorenz curve drawn from point C. Hereafter,
we suppose now that the dashed curve overlineOC is the drawn at the commune level. As is clear from the figure, the maximum size of Z depends on the shape of overlineOD. hence we also consider the normalized concentration index defined by \( \frac{Z}{Y+Z} \). This takes between 0 and 1.  

There are several other things to note in Figure 2. The proportion overlineOG of the population corresponds to the poor people since the bold line corresponds to perfect targeting. Hence, \( \overline{OG} \) is the poverty incidence. Note also that OD is a straight line when we use parameter \( \alpha = 0 \) for the FGT measure. Let us now suppose that the goal we face is to reduce FGT measure by \( \overline{OE} \). Then, \( \overline{OI} (= \overline{OE}) \) corresponds to the benchmark case of no targeting. Proportion \( \overline{OH} \) of the population must be assisted when imperfect targeting is carried out. \( \overline{OF} \) corresponds to the perfect targeting. Therefore, \( \overline{HI} \) is the reduction in leakage for commune-level targeting and \( \overline{IT} \) for perfect targeting. We can call this measure “equivalence gains” because it measures

\[12\] If we we do not order communes by the FGT measure, we may have negative values.
how much less population we need to target to achieve the same reduction in the FGT measure. Because how much equivalence gains one can achieve from the commune-level targeting depends on $FI$, we can also consider “relative equivalence gains” defined as $\frac{HI}{FI}$. The relative equivalence $\overline{FG}$ is the proportion of the poor people denied assistance. Hence the rate of Type I error is $\frac{\overline{OE}}{\overline{FG}}$ for the imperfect targeting. $\overline{FH}$ is the leakage to nonpoor, and hence the rate of Type II error for imperfect targeting is $\frac{\overline{FH}}{\overline{OF}}$. Hence, Figure 2 shows clearly that, as the size of the targeting (i.e. $\overline{OE}$) goes up, the rate of Type I error is reduced and the rate of Type II error is increased. Thus it depicts the trade-off between the Type I and Type II errors. If the cost of targeting is approximately in proportion to the share of population targeted, we can consider $\frac{DE}{AI}$ as the gains in the reduction of FGT when budget is fixed.

We drew the concentration curve in Figures 3 and 4. We found that the concentration indeed for malnutrition indicators is not very large in comparison with consumption. This is true in absolute and relative measures. Yet, we find significant gains from targeting at the commune level when compared with the stratum level targeting.

The concentration curve is good when the cost of targeting is approximately in proportion to the share of the population covered by targeting. However, if the cost of improving the nutrition status differs across the individuals, we need to take that into account. This is particularly important when we use higher values of parameter for FGT measures.

(I am going to explain this in a little more detail and present the results).

7 Conclusion

We have developed a methodology that is applicable to mapping of the prevalence of stunting and underweight. We have extended the small area estimation technique to include an individual effect correlated across different indicators, along with the location and het-
Figure 3: Concentration curve for stunting. The horizontal axis measures the share of population and the vertical axis measure the share of stunted children.
Figure 4: Concentration curve for underweight. The horizontal axis measures the share of population and the vertical axis measure the share of underweight children.
erokskedastic household effects. While estimated standard errors in our study are quite high for some communes, the magnitude of the standard errors for the estimated prevalence of stunting and underweight at the commune level is, on average, small enough to be useful. It should be noted that, as the number of communes targeted increases, the idiosyncratic component of the error tends to decrease. Hence, if a proposed policy intervention delivers assistance to a relatively large number of communes, high levels of standard errors may not necessarily be worrisome. While we applied the methodology to Cambodian data, it can be easily applied to other countries where a census dataset and a survey data with anthropometric component are available.

The results brought about by this study can be applied for targeting. Previously, estimates of the prevalence of child malnutrition from the CDHS were only available at the stratum level. Such estimates would be useful to target interventions in areas with an extremely high prevalence of malnutrition throughout the stratum, such as the northeastern Cambodia. However, provincial estimates often mask great disparities in the prevalence of malnutrition within the province. Targeting based on such estimates is likely fail to capture many malnourished children and to misallocate resources to many well-nourished children. Hence, the maps we created can substantially increase the efficiency of child nutrition programs.

Besides targeting, we demonstrated three applications of our methodology. First, we applied to the analysis of correlation between consumption poverty, inequality and malnutrition. Second, we conducted decomposition analyses of child health inequality. We consistently found that the between group component is smaller for health inequality than consumption inequality. We also found that the between group inequality in the country is small in comparison with the between country inequality in the world. Finally, we evaluated the efficiency of geographic targeting. We found that, while the gains from commune-level targeting is much less than perfect targeting, we can substantially gain from targeting in
comparison with stratum-level targeting.

The power of nutrition maps can be multiplied when they are combined with other maps. It might be assumed that there is insufficient access to food where high rates of malnutrition overlap with high poverty rates. An overlay of stunting prevalence with women’s education as a proxy for child care and access to water would be a first step towards identifying areas of high malnutrition along with their causes. The conclusions drawn from such combination of maps can provide program planners with valuable information in designing and coordinating interventions within and across different institutions in a more efficient, effective and transparent manner.

The methodology we discussed may be further extended to include other indicators. When a survey dataset includes both consumption and anthropometric components, we could produce nutrition maps and a consumption poverty map at once. It is conceivable that there exist household effects common between anthropometric and consumption indicators. Such exercise would lead to more reliable estimate of the correlation between commune-level prevalence of malnutrition and commune-level poverty rate.

References


ysis of changes in levels of child malnutrition since 1980.’ *Bulletin of the World Health Organization* 78, 1222–1233


35
Appendix

Proof of Eq (1)  First note the following:

\[
E[(u_{ch_1}^{(k)})^2] = (\sigma_{\eta}^{(k)})^2 + (\sigma_{e,ch}^{(k)})^2 + \frac{1}{I_{ch}}(\sigma_{\delta}^{(k)})^2
\]

\[
E[(u_{ch_2}^{(k)})^2] = (\sigma_{\eta}^{(k)})^2 + \frac{1}{H_c} \sum_{h' \in \mathcal{H}_c} (\sigma_{e,ch}^{(k)})^2 + \frac{1}{H_c^2} \left( \sum_{h' \in \mathcal{H}_c} \frac{1}{I_{ch'}} \right) (\sigma_{\delta}^{(k)})^2
\]

Hence,

\[
\sum_{c \in C} \frac{w_c}{H_c} \sum_{h \in \mathcal{H}_c} E[(u_{ch})^2] = \sigma_{\eta}^2 + \sum_{c \in C} w_c H_c \left( \frac{1}{H_c} \left( \sum_{h' \in \mathcal{H}_c} \sigma_{e,ch'}^2 \right) + \frac{1}{H_c^2} \left( \sum_{h' \in \mathcal{H}_c} \frac{1}{I_{ch'}} \right) \sigma_{\delta}^2 \right)
\]

\[
= \sigma_{\eta}^2 + \sum_{c \in C} w_c H_c \left( E[(u_{ch})^2] - \sigma_{\eta}^2 \right)
\]

\[
= (1 - \sum_{c \in C} w_c H_c) \sigma_{\eta}^2 + \sum_{c \in C} w_c H_c E[(u_{ch})^2]
\]

Solving for \( \sigma_{\eta}^2 \), we have Eq(1).

Proof of Eq (2)  First note the following:

\[
E[(u_{ch} - u_{ch'})^2] = \frac{H_c - 2}{H_c} \left\{ \frac{\sigma_{\delta}^2}{I_{ch}} + \sigma_{e,ch}^2 \right\} + \frac{1}{H_c^2} \left\{ \sum_{h' \in \mathcal{H}_c} \left( \frac{\sigma_{\delta}^2}{I_{ch'}} + \sigma_{e,ch'}^2 \right) \right\}
\]

By summing over the households in each cluster, we have

\[
E[\sum_{h' \in \mathcal{H}_c} (u_{ch} - u_{ch'})^2] = \frac{H_c - 1}{H_c} \left\{ \sum_{h' \in \mathcal{H}_c} \left( \frac{\sigma_{\delta}^2}{I_{ch'}} + \sigma_{e,ch'}^2 \right) \right\}
\]

Therefore, for all the households in \( \mathcal{C}^* \),
\[\sigma_{\epsilon,eh}^2 = \frac{H_c}{H_c - 2} E[(u_{eh} - u_{e..})^2] - \frac{1}{(H_c - 1)(H_c - 2)} E[\sum_{h' \in H_c} (u_{eh'} - u_{e..})^2] - \frac{\sigma_\delta^2}{I_{eh}}\]

Adding \(\sigma_\delta^2\) to both sides of the equality and arranging the terms, we get Eq(2).
<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>77.48</td>
<td>0.73</td>
</tr>
<tr>
<td>(Max years of educ for HH)*(Head has no educ)</td>
<td>1.06</td>
<td>0.16</td>
</tr>
<tr>
<td>(Head’s years of educ)*(Water from dug well)</td>
<td>0.50</td>
<td>0.11</td>
</tr>
<tr>
<td>(# professional women in HH)*(Head has some secondary educ)</td>
<td>-16.41</td>
<td>2.85</td>
</tr>
<tr>
<td>(# women in HH for less than 5 years)*(# members aged 0)</td>
<td>-8.66</td>
<td>1.53</td>
</tr>
<tr>
<td>(Head’s years of educ)*(Total deaths of children in HH last 12mth)</td>
<td>-0.67</td>
<td>0.10</td>
</tr>
<tr>
<td>(Child deaths over live births)*(Age of spouse)</td>
<td>0.35</td>
<td>0.07</td>
</tr>
<tr>
<td>(Rain water)*(Head completed primary educ)</td>
<td>10.83</td>
<td>1.68</td>
</tr>
<tr>
<td>(Sex=Female)*(Head has some secondary educ)</td>
<td>5.95</td>
<td>1.01</td>
</tr>
<tr>
<td>(Roof made of rock)*(Age=4)</td>
<td>-8.91</td>
<td>0.91</td>
</tr>
<tr>
<td>(Rain water)*(Age=3)</td>
<td>-6.44</td>
<td>1.21</td>
</tr>
<tr>
<td>(Water not from pipe/well/bottle)*(Age=0)</td>
<td>11.95</td>
<td>2.32</td>
</tr>
<tr>
<td>(Ratio of female in HH)*(Age=3)</td>
<td>-0.16</td>
<td>0.03</td>
</tr>
<tr>
<td>(# women in HH worked last 12mth)*(Variance of NDVI in Nov)</td>
<td>12.41</td>
<td>3.15</td>
</tr>
<tr>
<td>(Toilet on premise)*(Distance to river)</td>
<td>2.97E-04</td>
<td>6.18E-05</td>
</tr>
<tr>
<td>Ratio of head in village with secondary education</td>
<td>21.98</td>
<td>4.45</td>
</tr>
<tr>
<td>(Age of head)*(Water from dug well)</td>
<td>-0.09</td>
<td>0.02</td>
</tr>
<tr>
<td>(# women worked)*(Child death over births)</td>
<td>-22.71</td>
<td>4.94</td>
</tr>
<tr>
<td>(No women in HH have educ)*(Head has some secondary educ)</td>
<td>-5.67</td>
<td>1.91</td>
</tr>
<tr>
<td>(Head is male)*(age=0)</td>
<td>3.05</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Table 4: GLS regression results for standardized height in Coastal Eco-zone.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>10.21</td>
<td>0.21</td>
</tr>
<tr>
<td>(Max years of educ for HH)*(Head has no educ)</td>
<td>0.20</td>
<td>0.04</td>
</tr>
<tr>
<td>(Head’s years of educ)*(Water from dug well)</td>
<td>0.11</td>
<td>0.03</td>
</tr>
<tr>
<td>(# women in HH for less than 5 years)*(# members aged 0)</td>
<td>-2.07</td>
<td>0.38</td>
</tr>
<tr>
<td>(Ratio of those ever attended school)*(Max year of educ for females in HH)</td>
<td>1.74E-03</td>
<td>4.32E-04</td>
</tr>
<tr>
<td>(Roof made of rock)*(Age=4)</td>
<td>-1.34</td>
<td>0.25</td>
</tr>
<tr>
<td>(# members aged 0-4)*(Age=0)</td>
<td>0.85</td>
<td>0.10</td>
</tr>
<tr>
<td>(Floor material wood)*(Area of rice field in village)</td>
<td>3.99E-04</td>
<td>7.21E-05</td>
</tr>
<tr>
<td>(Max female educ in HH is some primary educ)*(Change b/w 1993 and 1997 in coverage of ?)</td>
<td>-7.13E-08</td>
<td>1.72E-08</td>
</tr>
<tr>
<td>(Flood prone commune)*(Age=4)</td>
<td>1.88</td>
<td>0.34</td>
</tr>
<tr>
<td>(# members aged 0-4)*(Water from dug well)</td>
<td>-0.43</td>
<td>0.10</td>
</tr>
<tr>
<td>(Age of head)*(toilet not on premise)</td>
<td>-0.02</td>
<td>3.78E-03</td>
</tr>
<tr>
<td>(# younger children)*(Roof made of wood/plastic)</td>
<td>1.15</td>
<td>0.23</td>
</tr>
<tr>
<td>(# students)*(Child death over live births)</td>
<td>-1.20</td>
<td>0.35</td>
</tr>
<tr>
<td>(toilet not on premise)*(head is male)</td>
<td>0.63</td>
<td>0.16</td>
</tr>
<tr>
<td>(Roof made of wood/plastic)*(Age=3)</td>
<td>-0.95</td>
<td>0.25</td>
</tr>
<tr>
<td>(Head has some primary educ)*(Age=4)</td>
<td>-0.96</td>
<td>0.23</td>
</tr>
<tr>
<td>(Head is male)*(Age=1)</td>
<td>-0.73</td>
<td>0.22</td>
</tr>
<tr>
<td>(# members 65 or over)*(Age=0)</td>
<td>1.95</td>
<td>0.48</td>
</tr>
<tr>
<td>(Max year educ. for females)*(Age=3)</td>
<td>-0.21</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table 5: GLS regression results for standardized weight in Coastal Eco-zone.