Components of small area variation in fertility rates among married women in south India

BN Murthy, S Jabbar, T Venkatarao, SK Sureshkumar and MD Gupte
National Institute of Epidemiology (Indian Council of Medical Research), Chennai, India

Accepted 13 March 2003

Objective To adapt and develop a method for finding out what proportion of the variation among small areas in the number of births to married women is excess (systematic) variation over and above the chance (random) variation.

Methods We adopted a two-stage sampling procedure to select 20 sub-centres in south India. We contacted all households and collected information on recent births and socioeconomic variables from all ever married women aged 15–49 years. Systematic and random components of variance were estimated using Poisson regression, adjusting for socioeconomic factors.

Results Of the observed variance in fertility, 65% is explained by the estimated systematic variation. Though many important explanatory variables are considered, the systematic variance is significant mainly among illiterate women, those aged under 18 years at marriage, the marginalized population, and those with labourer husbands.

Conclusions Poisson regression can be adapted to estimate the random and systematic components of variation in fertility among small areas. The systematic component of variance can further be adjusted for socioeconomic factors. Identification of the significant predictors will help health planners develop necessary interventions at the small area level.

Health information systems are needed for small areas such as sub-centres (SC), primary health centres (PHC), or at district level for health planning. This includes monitoring of maternal and child health services and providing warning signals to identify specific problem areas, in order to improve them through necessary interventions.¹ One such problem area has been the high levels of fertility; e.g. crude birth rate (CBR), general marital fertility rate (GMFR), total fertility rate (TFR), children ever born (CEB), children surviving (CS) etc., and constant infant mortality. As there is no routine surveillance system for fertility, information on fertility rate at the local level is often based on specific sample surveys conducted by various research institutes/organizations. It is difficult to monitor the fertility levels in a specified local area.

Most of the theories on fertility change deal mainly with the identification and classification of a set of biological, demographic, social, economic, and cultural variables as potential predictors, and the study of the interrelationships among them and their association with the observed fertility. Recent developments in this regard have assessed quantitatively the effects of these selected predictors on fertility. However, these approaches for estimating the effects on fertility of a set of selected predictors have generally been limited to the macro level, with either the country or a state or a region within the country as the unit of study. However, none of these studies in India and elsewhere has undertaken a detailed ‘small area analysis’ of the effects of selected socioeconomic variables (predictors) and especially the variation in fertility rates between small health areas. Such an opportunity is made available in this study with data available on predictors from a sample survey conducted in south India.

There are many statistical problems associated with analysing variation in rates, ratios, prevalences, and incidences among small areas.² The concept of small area variation has been developed and applied to death rates,³ cancer risks,⁴ surgical operation rates,⁵ hospital admission rates,⁶ health behaviour patterns,⁷ incidence of insulin-dependent diabetes mellitus,⁸ and low birthweight.⁹ Earlier studies showed that the amount of variation explained in the dependent variable by associated predictors depended heavily on the level of aggregation.¹⁰ The lower the level of aggregation, say at village/household/individual level, the lower the variance explained by the associated factors. This might be due to the considerable amount of
random variation between the small areas compared with the variation within. The variation between associated factors also influences the amount of random variation when measuring the variation between small areas. It is necessary to know whether there is more variation between areas i.e., systematic variation than what is expected, and if so, to what extent it exists vis-à-vis the level of random variation.

The study of small area analysis is gaining importance in the health field especially with regard to the variation between indicators of mortality and/or morbidity. The same can be applied to fertility rates and its associated predictors. Among other things, there is a rise in the socioeconomic status of the population and constant use of contraceptives, including sterilization, by the community. In this context, fertility level at the local area level is needed for formulation of intervention strategies or programmes. A number of factors have been identified as predictors of fertility, based on the judgement whether they are manipulable for intervention or not.

In this study, the methodology for small area analysis is developed to estimate the random, as well as the systematic, components of variation in fertility after adjusting for the predictors. The study aims (1) to adapt and develop methods for small area analysis of the non-random component of the variation in fertility, (2) to apply the method to examine what proportion of the variation among small areas in births to married women is systematic (excess) variation over and above the expected random (chance) variation for selected socioeconomic indicators in a district of Tamil Nadu.

**Methods**

The cultural background in India does not permit premarital pregnancies. Such pregnancies do exist but the percentage of unmarried women giving birth is below 0.1%. The data regarding the fertility profile of married women was obtained through a well-planned sample survey conducted in one Health Unit District of Tamil Nadu. The design of the study was as follows. The selected district had 82 PHC covering a total of 551 SC. Twenty SC were randomly selected for the study using a two-stage sampling design. In the first stage, a random sample of 10 PHC was selected using probability proportional to size (PPS) technique. In the second stage, a simple random sample of two SC was chosen from each selected PHC.

In each of the 20 selected SC, well-trained graduate female field investigators interviewed all currently married women aged 15–49 years and collected data on fertility using an standardized pre-tested questionnaire. The data include the outcomes of recent pregnancies, if any, for the women, and different socioeconomic indicators of the household and the individual women during the year 1993. Quality control measures such as spot scrutiny and consistency checks were assiduously employed. The survey was carried out from June 1994 to November 1994. A similar survey was done during 1995 in the same SC to collect similar data on the fertility profile for the year 1994, with approximately similar recall period.

A review of the literature available so far suggests that no framework has emerged. Empirical frameworks developed independently by Davis and Blake and Easterlin (with criticisms) indicated that ‘intermediate variables’ had a direct impact on fertility. This led Bongaarts framework to consider at the aggregate level four proximate determinants viz., marriage, postpartum infertility due to breastfeeding practices, contraception, and fetal wastage including induced abortion. Wilmens observed that Bongaarts framework was not appropriate. Women’s age at marriage was considered in the present analysis because both Jain and James suggested that age at marriage for females had a bearing on fertility. It was also considered by Mason as an indicator of the status of women. The woman’s occupation and that of their husbands was also considered to be a measure of development and thus expected to influence fertility.

Using age-specific rates of socioeconomic characteristics and fertility for the state in 1993, we indirectly standardized all crude rates for each centre for age. Similarly, standardized rates for socioeconomic indicators, as well as for fertility, were computed for 1994. The aim of this study was to examine the extent of systematic variance between the areas and whether it was statistically significant. Since the expected number of births for each predictor was at least five, Cochran’s heterogeneity test was used with $k - 1$ degrees of freedom ($k = \text{number of small areas analysed i.e. 20}$). This test can be related to the binomial distribution. Suppose $o_i$ and $e_i$ are the observed and expected numbers of births respectively, then for the SC, for illiterate women,

$$\sum_{i=1}^{K} \frac{(o_i - e_i)^2}{e_i}$$

with $\chi^2_{(k-1)}$ d.f. The expected number of births for each SC was obtained on the basis of age-specific marital fertility rates applied to the age distribution of the married women in that area. The observed variance was calculated by dividing the $\chi^2$ value by the number of births studied for each predictor. The total number of expected births, by definition, should be equal to the total number of observed births and accordingly the expected number of births for each SC was worked out. Since the population sizes varied considerably between the SC, the expected number of births and the resulting observed variance varied substantially. The observed variance was derived as the variance of the ratio of the observed and expected numbers of births with the expected mean value as 1. The observed variance could be written as the weighted variance with weight being proportional to population size in that area.

With simple algebra the observed variance could be split into two components viz., systematic variance and random variance. Under the null hypothesis all the variation was due to random variance and the expected value of $\chi^2$ was the degrees of freedom i.e., $k - 1$. The random component of the variance $= (k - 1)/n$. The systematic variance component was then

$$\sum_{i=1}^{K} \frac{(o_i - e_i)^2}{e_i} \frac{(k - 1)}{n}$$
The proportion of systematic variance component to the total variance was calculated as the ratio of systematic variance with the observed variance. The above procedure was repeated for each predictor variable considered in the analysis.

To estimate both the systematic and random components of variance for each predictor variable, Poisson regression was carried out with observed births as the dependent variable and the expected number of births as the independent variable. The deviance provided a \( \chi^2 \) value for the test of goodness of fit for the model. In this model, selected predictors, say, illiteracy, age at marriage <18 years, living in kuchcha households (defined as the household having mud floor, mud walls with thatched roof), marginalized population (defined as those people living at the periphery of the village), labourers etc. were introduced as the independent variables. This adjusted value of \( \chi^2 \) for the predictor was reduced indicating the improvement in goodness of fit of the model. This \( \chi^2 \) value was used for the systematic variance given in the earlier formula with k – 2 d.f. The expected random variance component had k – 2 as numerator in the formula.

Similarly, corresponding change was made in the formula for systematic variance. This modified systematic variance was labelled as the predictor-adjusted systematic component of variance in births. The significance of the systematic variance was examined and the corresponding \( P \)-value was computed based on the \( \chi^2 \) value. The predictor-related systematic component of variance in birth rates was calculated as the difference between the adjusted systematic component of variance and the total systematic variance in births.

### Results

To comprehend better, the analysis was done after amalgamating the results for 1993 and 1994. In all, 19 173 households were identified in the 20 SC; of these, 1% could not be contacted as the inhabitants were away at the time of the field visit. The levels of the predictors (socioeconomic factors) varied substantially from 33% to 68% (Table 1). The observed variance was highly significant (\( P < 0.001 \)). The variance was highest (0.148) for women belonging to the ‘marginalized population’ and lowest (0.01) for ‘illiterate women’. The corresponding standard deviation (SD) varied between 10% and 38%. The general marital fertility rate was the lowest among family planning users and highest among women belonging to the marginalized population.

The proportion of estimated systematic variance in the ratio of births for the whole study sample was large and significant (Table 3). The proportion of estimated systematic variance in the ratio of births was significantly high for all the predictors except for family planning users, those living in a kuchcha household, labourers, and those whose husbands were illiterate. The
The highest systematic component of variance in the ratio of births was for ‘age at marriage below 18 years’. The systematic component of variance in the ratio of births varied between 0.023 and 0.046, the corresponding SD varied between 15% and 21%.

For illiterate women the systematic component of variance was estimated as 0.029, (SD = 17%). For women living in kuchcha households and who were labourers, the systematic component of variance was the same at 0.023 (SD = 15%). For illiterate women, women married before age 18 years, women belonging to marginalized population, and whose husbands were labourers, 61% to 65% of the observed variance was accounted for systematic variance. The proportion of systematic component of variance varied between 52% and 59% for other predictors combined, the observed variance was more attributable to systematic variance.

The systematic variance in the ratio of observed to expected births was divided into the predictor-adjusted and predictor-related systematic components of the variance in fertility. The predictor-adjusted systematic variance in the ratio of observed to expected births was significant for the whole study sample as well as for illiterate women, age at marriage below 18 years, and labourer husbands (Table 4). The predictor (marginalized population) related component of variance in the ratio of observed to expected births was significant but low at 0.007.

For the entire sample the systematic variance in the ratio of observed to expected births was marginally lower after adjustment for all the predictors. Similar findings were observed in the systematic variance for each of the predictors except for family planning users, labourer husbands, and husband illiterate. In these predictors the systematic component of variance remains unchanged.

### Table 3

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Total births</th>
<th>Total variance</th>
<th>SCV&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Proportion of SCV&lt;sup&gt;a&lt;/sup&gt;</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All subjects</td>
<td>1762</td>
<td>0.034</td>
<td>0.023</td>
<td>0.68</td>
<td>0.002</td>
</tr>
<tr>
<td>Illiterate</td>
<td>1100</td>
<td>0.046</td>
<td>0.029</td>
<td>0.63</td>
<td>0.02</td>
</tr>
<tr>
<td>Age at marriage &lt;18 years</td>
<td>985</td>
<td>0.035</td>
<td>0.036</td>
<td>0.65</td>
<td>0.008</td>
</tr>
<tr>
<td>Family planning users</td>
<td>541</td>
<td>0.079</td>
<td>0.044</td>
<td>0.56</td>
<td>0.16</td>
</tr>
<tr>
<td>Living in kuchcha households</td>
<td>1211</td>
<td>0.039</td>
<td>0.023</td>
<td>0.59</td>
<td>0.07</td>
</tr>
<tr>
<td>Labourers</td>
<td>883</td>
<td>0.044</td>
<td>0.023</td>
<td>0.52</td>
<td>0.34</td>
</tr>
<tr>
<td>Marginalized population</td>
<td>635</td>
<td>0.076</td>
<td>0.046</td>
<td>0.61</td>
<td>0.05</td>
</tr>
<tr>
<td>Labourer husband</td>
<td>978</td>
<td>0.049</td>
<td>0.030</td>
<td>0.61</td>
<td>0.04</td>
</tr>
<tr>
<td>Illiterate husband</td>
<td>688</td>
<td>0.062</td>
<td>0.034</td>
<td>0.55</td>
<td>0.18</td>
</tr>
</tbody>
</table>

<sup>a</sup> Systematic component of variance.

### Table 4

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Total SCV&lt;sup&gt;a&lt;/sup&gt; in births</th>
<th>Predictor-adjusted SCV&lt;sup&gt;a&lt;/sup&gt; in births</th>
<th>P-value</th>
<th>Predictor-related SCV&lt;sup&gt;a&lt;/sup&gt; in births</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All subjects</td>
<td>0.023</td>
<td>0.021</td>
<td>0.004</td>
<td>0.001</td>
<td>0.06</td>
</tr>
<tr>
<td>Illiterate</td>
<td>0.029</td>
<td>0.028</td>
<td>0.02</td>
<td>0.001</td>
<td>0.38</td>
</tr>
<tr>
<td>Age at marriage &lt;18 years</td>
<td>0.036</td>
<td>0.033</td>
<td>0.01</td>
<td>0.003</td>
<td>0.07</td>
</tr>
<tr>
<td>Family planning users</td>
<td>0.044</td>
<td>0.044</td>
<td>0.12</td>
<td>0.000</td>
<td>0.84</td>
</tr>
<tr>
<td>Living in kuchcha households</td>
<td>0.023</td>
<td>0.022</td>
<td>0.06</td>
<td>0.001</td>
<td>0.46</td>
</tr>
<tr>
<td>Labourers</td>
<td>0.023</td>
<td>0.022</td>
<td>0.29</td>
<td>0.001</td>
<td>0.62</td>
</tr>
<tr>
<td>Marginalized population</td>
<td>0.046</td>
<td>0.039</td>
<td>0.10</td>
<td>0.007</td>
<td>0.04</td>
</tr>
<tr>
<td>Labourer husband</td>
<td>0.030</td>
<td>0.030</td>
<td>0.03</td>
<td>0.000</td>
<td>0.71</td>
</tr>
<tr>
<td>Illiterate husband</td>
<td>0.034</td>
<td>0.034</td>
<td>0.14</td>
<td>0.000</td>
<td>0.89</td>
</tr>
</tbody>
</table>

<sup>a</sup> Systematic component of variance.

### Discussion

The first point of interest from this study is that a specific method is suggested and attempts are made in a novel way to study the variation in fertility rates between 20 SC for selected predictors. Various statistical methods had been suggested for small area variation between geographical areas when analysing relatively rare cases in mortality/morbidity. The concept of the χ² test for testing the heterogeneity of the rates in different geographical areas is quite old (Cochran’s χ²).<sup>20</sup> The concept of weighted observed variance has been used earlier to estimate the components of small area variation. Estimation of systematic variance is important even if it is not significant when the concept of ‘avoidable fertility’ is elucidated in examining the variation among the small areas with high fertility rates.

The second point of interest is that a considerable part of the variation in fertility rates for selected predictors is systematic. The idea of small area analysis meant that the population and the resulting outcomes in the study areas should be relatively small. This indicates that the concept of small area analysis for fertility rates seems appropriate. Poisson regression is used to estimate adjusted systematic component of variance. The systematic component of variance decreased after adjustment, suggesting it is sensitive to changes in the study factor levels. However, a fairly large amount of variation is found to be systematic.

The third point of interest is that the results of the analysis show pathways for further studies at the community level. For example, the illiteracy among eligible women is one of the important variables and so needs attention and intervention. Age at marriage below 18 years, being in the marginalized population, and labourer husbands were the other three important variables for which the systematic component of variance in fertility rates was significant. These predictors deserve special attention for programmed interventions. Though other important variables
from the programme point of view were considered in this model, none of them was found to be significant. It was not possible to include data on other proximate variables such as breastfeeding practices, abstinence, letal wastage, secondary sterility etc., and therefore, it could not be possible to judge whether variation is due to these factors. The present analysis did not explain more than what was due to selected predictors and random effects.

The fertility rates in these SC were found to be higher among the marginalized population and those with labourer husbands. Attention needs to be given to those selected variables (predictors) that can lower the fertility rates at the local area level. The levels of these variables also provide warning signals for likely high fertility. The interrelationships among the predictors influencing fertility might vary between different areas. In the present study, the time period between two fertility surveys was only one year, too short to show trends in fertility pattern and the levels of predictors.

Small area analysis of the consistency between fertility rates and predictors was needed because a high degree of association had been shown between fertility and the predictors. The variations in fertility rates were not fully explained by the variation in the predictors. Other plausible explanations need to be sorted out, for example, the effects of population policy for early identification of predictors affecting fertility. Reproductive and child health programmes (RCH) showed the influence of proximate determinants on fertility. Organized RCH programmes need to be introduced at local area levels at different time points in the state. It would be of interest to study the relation between RCH programmes and the systematic components of variance in fertility rates. The systematic variation in fertility rates is also dependent on age at marriage, children ever born, and children born to eligible women who survived.

It is important to emphasize that measuring adjusted systematic variance in fertility may not solve the problem of separating the effects of predictors from the effects of improved health. The data on predictors and fertility is easily available when compared with the data on health. The predictors and the fertility experienced refer to a district and the potential bias in measuring them is considered small in the SC studied. Inclusion of live births in the study was, however, another important variable for the analysis of possible explanation for systematic component of variation in the predictors as well as in fertility.

The Draft National Health Policy of the Government of India is characterized by rapid progress with new programmes being continually introduced with the aim of reducing fertility. The selection of useful indicators of avoidable fertility under decentralized planning differs from one period to another. Analysis of the systematic component of variation in fertility rates at the local area level should be useful for further studies on population.

Acknowledgements
The authors are grateful to the comments and suggestions made by the referees. The authors are grateful to Dr S Radhakrishna, Ex-Director of Institute for Research in Medical Statistics under whose guidance the study was initiated. The authors are also thankful to Mr MC Satagopan, Senior Research Officer for supervision of field staff and data collection. Mr R Ezhil, Technical Officer, Mr P Kamaraj, Mr K Boopathi, Research assistants for spot scrutiny and consistency checks in the field and Mr S Venkatasubramanian and Mr AK Mathai, Research assistants of the Institute for necessary literature search, and to Mrs R Janaki for secretarial assistance.

KEY MESSAGES

- Small area variation in fertility rates exists.
- The observed variation can be split into systematic and random components of variance.
- The systematic component of variance is adjusted for factors affecting fertility rates. It was significant for illiterate women, those aged under 18 years at marriage, the marginalized population, and those with labourer husbands.
- Reproductive and child health programmes need to be introduced at the local area level to achieve a reduction in fertility.

References

7 Ecob R, Macintyre S. Small area variations in health related behaviours; do these depend on the behaviour itself, its measurement, or on personal characteristics? Health Place 2000;6:261–74.