

## How to estimate small area health related characteristics of populations?

Health related characteristics of a population in a society are significant to health promotion programs and to the provision of better health services. The efforts of feasible health planning generally target local areas such as local health regions or small area health units, and the population specific health program planning often requires precise estimates of health behaviours of the population at a fine spatial scales e.g., statistical local areas (SLAs) in Australia or counties in US or Wards in UK. But health related data are not always available at such scales due the policy and privacy constraints.

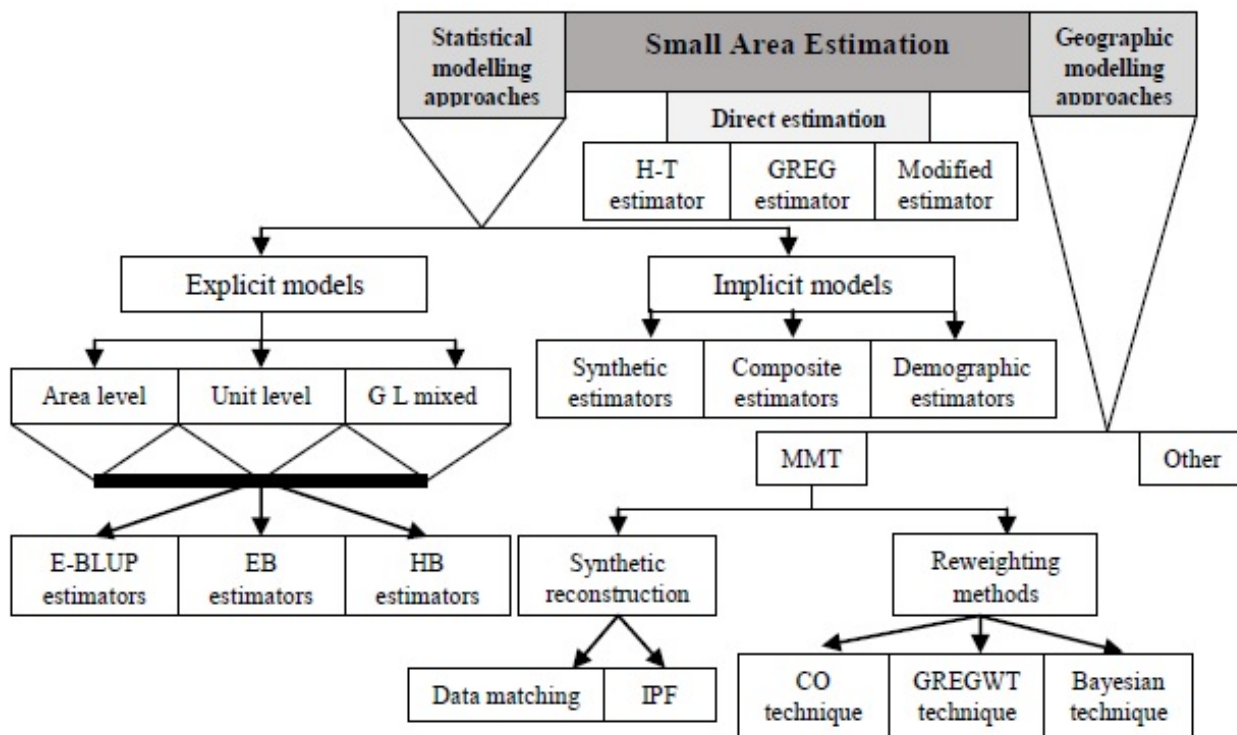


Fig. 1. A summary of different techniques and estimators for SAE.

Policy makers sometimes rely on national level or state level datasets to understand the health needs of their communities. However, the basic problem with these surveys is that they are not designed for efficient estimation at small areas. Any conventional estimators from these data mainly have the two major limitations: (i) prevalence estimates can only be computed for a subset of all areas which contain respondents to the survey; and (ii) for those small sampled areas the achieved sample size will usually be very small and the estimator will thus have low precision, and ultimately statistically unreliable.

The lack of a national dataset having detailed characteristics of individuals at fine spatial scales negatively impacts the ability of local and national agencies to manage serious health issues in the community and associated risks factors. For examples, youth and adults' smoking behaviours, characteristics of overweight and obesity, risk factors and distribution of cardiovascular disease, etc at small areas are not readily available, but crucially important for policy making or evaluation purposes. In this study, an appraisal of the methodologies for the estimation of health related characteristics of populations at small area levels is provided.

The overall methodologies for SAE are depicted in Figure 1. Traditionally there are two types of SAE – direct and indirect model-based estimations. The direct methods include three estimators: H-T estimator, GREG estimator and modified direct estimator. Indirect methodologies of SAE are divided into two approaches – statistical and geographic. The statistical modelling is based on statistical models (i.e. implicit and explicit models) and techniques; while, the geographic approaches use microsimulation modelling.

Most of the developed nations are utilising SAE as an essential means to support the knowledgeable and effective decision making and policy analysis for various issues at local levels. There are three distinct sets of modelling methodologies utilised by practitioners: (i) *indirect standardisation and individual level modelling*; (ii) *multilevel statistical modelling*; and (iii) *microsimulation modelling technology (MMT)*. A highlight of these three sets of methodologies is depicted in Table 1.

<i>Properties</i>	<b>Indirect standardisation and individual level modelling</b>	<b>Multilevel statistical modelling</b>	<b>Microsimulation modelling technology</b>
<i>Methods / Comments</i>	Models are based on individual level covariates from the survey or from the Census data.	Models are based on multilevel covariates from the survey or from the Census data.	Models are based on synthetically created micropopulation datasets that integrate a set of variables from the survey or Census.
<i>Advantages</i>	<ul style="list-style-type: none"> <li>• Very easy and inexpensive to apply and the estimate is unbiased for large samples.</li> <li>• Flexible to calculate estimates at the national level and use them to adjust the method by calculating estimates for different spatial scales.</li> <li>• The estimates produced by this method for each small area within a larger area can be ratio adjusted so that a weighted average of the adjusted small area estimates equals the direct estimate for the larger area.</li> </ul>	<ul style="list-style-type: none"> <li>• Fairly easy to apply and more explanatory model that can provide CIs of the estimates.</li> <li>• Flexible to allow both the effects of an individual's circumstances and at any level of hierarchy of the social system, demographic cluster and physical environment.</li> <li>• Able to explore the extent of any differences between the small areas and small area level characteristics. A range of computing software is also available for multilevel modelling.</li> </ul>	<ul style="list-style-type: none"> <li>• Rather sophisticated and state of the art method that can generate measures of statistical reliability of the estimates.</li> <li>• Very robust approach in terms of the choice of further aggregation or disaggregation of the small area estimates on the basis of different spatial scales or demographic domains.</li> <li>• Doable to utilise the small area level synthetic microdata file for further analysis and updating. It is also possible to measure small area effects of any policy changes. However, the other traditional statistical approaches do not have such robustness.</li> </ul>
<i>Limitations</i>	<ul style="list-style-type: none"> <li>• This approach considers the notion that the national level prevalence rates for each subgroup apply uniformly across all small areas, in fact this is not viable in many cases.</li> <li>• The choice of covariates in the model is restricted by the data requirement to have equivalent covariate information for all of the small areas.</li> <li>• Borrows strength from the overall data but cannot increase effective sample. The estimates are often unreliable due to misclassification of the models and/or use of inconsistent auxiliary data.</li> </ul>	<ul style="list-style-type: none"> <li>• This method imposes quite stringent data requirements as they require an exact match between the covariates used in the model and the counts available from the Census.</li> <li>• Important individual level predictors may be eliminated from the model simply because their distribution at the small area level is unknown.</li> <li>• Estimating the standard errors for the estimates which use both the individual level and area level covariates is considerably more complex than the individual level modelling technique.</li> </ul>	<ul style="list-style-type: none"> <li>• This approach is highly computing intensive with an issue being the size of the final file (e.g., in Australia 1300 areas/columns × 30 weights = 39000 columns; then 39000 columns × 12000 households = 468 million cells in the final file).</li> <li>• Solely depends on a good microdata generation technique. Several techniques are in use, but a much better and commonly available reweighting method is yet to be developed.</li> <li>• When there are too few observations in a sample stratum, the resulting standard error estimates should be statistically unreliable. Also validations of the small area estimates are still challenging for some models.</li> </ul>
<i>Applications</i>	Mostly used when the overall sample size is large and come from reliable agency. Typically, the small area level sample data are not large enough in many areas.	Widely used approach for repeated measures data, multivariate data, data with missing observations, data with clustering information etc.	Quickly becoming a popular methodology across the researchers in the developed nations, and frequently used for small area estimation and social policy analysis.

Tab. 1. A comparison of the three approaches to small area health related characteristics estimation.

The findings suggest that although each of these three modelling approaches has its own strengths in relation to generating small area health related characteristics estimation, the microsimulation modelling technology has more robustness over other methods in the sense that further aggregation or disaggregation is possible on the basis of the choice of spatial scales or domains. As well, it is possible to use the microdata file from MMT for further analysis and updating, and to assess small area effects of policy changes. MMT approaches also allow what if scenarios in terms of policy changes. The other approaches do not have such utilities. Finally, from the overall assessment, it is apparent that MMT is comparatively a precise way to estimate small area health related characteristics and evaluate policy changes. Our future research will employ this new approach to produce the estimates of small area health related characteristics – in particular, the estimates of smoking behaviours of adults and/or estimates of the prevalence of obesity of adults at small area levels in Australia.

**Azizur Rahman**

*Discipline of Mathematics and Statistics, School of Computing and Mathematics,  
Charles Sturt University, Wagga Wagga, NSW 2678 Australia*

## **Publication**

[Estimating small area health-related characteristics of populations: a methodological review.](#)

Rahman A

*Geospat Health. 2017 May 8*