Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 70(2) 2016 153–158

Small Area Statistics: An Application for Studying Reliability and Validity

Padam Singh[1] and Satyendra Keshari[2]
[1]Invision Communications and Research Pvt. Ltd., New Delhi
[2]TRIOs Development & Support Pvt. Ltd., New Delhi

Received 04 April 2016; Revised 25 June 2016; Accepted 30 June 2016

SUMMARY

The theory of small area statistics is in use for over four decades. The validity has not been examined in the absence of the actual results. The data on TFR which was being calculated at district level using the theory of small area statistics utilizing the information on third and higher order births, have become available through recently conducted Annual Health Survey for EAG States by Office of Registrar General of India. Based on this data, the analysis presented in this paper raises questions on the validity of results using methodology of small area statistics. Even with high values of R^2, the results based on small area statistics combined for EAG states match with actual only in about 44 percent of the cases. Further, undertaking analysis at State level using small statistics, matching improves to 54 percent only.

Keywords: Annual health survey, Covariates, Higher order births, Total fertility rate, Small area statistics.

1. BACKGROUND

The National Population Policy (NPP)-2000 in order to stabilize the population by 2045 set the target of TFR (the average number of children a woman bears in her lifetime) at replacement level of 2.1 by 2010. That time, the TFR for the country as a whole was 3.3 and that for the States of Bihar, Madhya Pradesh, Rajasthan, Uttar Pradesh and Odisha together was 4.2 which was double the desired level 2.1 (Table 1).

In view of this, it was emphasized in the NPP-2000 that the achievement in the backward states of UP, MP, Bihar, Rajasthan and Odisha will determine the time and the year in which the country is likely to achieve population stabilization. It is important to mention that Dr. Ashish Bose coined the term BIMARU to refer to these four grossly under-developed states of (undivided) Bihar, Madhya Pradesh, Rajasthan and Uttar Pradesh in the 1980s. Later, Odisha was also included, to expand it to BIMARUO.

Table 1. Total fertility rate for selected states and country

States	TFR
All India	3.3
Bihar	4.5
MP	4.1
Rajasthan	4.2
UP	4.8
Odisha	3.1

It was felt that in these states and elsewhere there are districts with very high TFR of 3.5 or more which need to be identified and targeted for faster gains. However, the data on TFR at district level were not available. This was a challenge and required use of statistical methodology to arrive at the same. In particular, the theory small area statistics was used to develop the estimates of TFR at district level by Padam Singh who identified 133 districts with TFR more than 3.5 requiring special attention by National Population Commission. This was accepted then by National

Corresponding author: Padam Singh
E-mail address: dr.padamsingh2013@gmail.com

Population Commission, however this work was published later in the Journal of Empirical Research in Social Science under joint authorship of Singh, Padam and Adhikari, Tulsi (2006). For brevity and completeness, the methodology used therein is discussed briefly as under-

***Small Area Statistics*:** The Term "small area" refers to a small geographical area or "small domain", i.e. a particular demographic within an area. The estimates at small area level are needed to support local decision making. If a survey has been carried out at the national and state level, the sample size within any particular "small area" may be too small to generate estimates at the level of small area. There are various methods available for generating estimates for small area. The important among these are the synthetic method, spatial smoothing, and regression.

The reliability and accuracy of three methods used in small–area have been examined and it is reported that the regression method produces the most valid and precise estimates. While using the regression method it is necessary to have information on some covariates that is available for small areas. Therefore, underpinning all of the small area estimation methods is the need for good quality, spatially detailed covariate data relating to the area scale at or below that which the target variable of interest is desired.

Total Fertility Rate: TFR is defined as the average number of children a woman bears in her lifetime. In mathematical terms, this rate is the sum of the age-specific birth rates (5-year age groups between 15 and 49) for female residents of a specified geographic area during a specified time period (usually a calendar year) multiplied by 5. Mathematically.

$$TFR = 5 \sum ASFR_a \text{(for 5-year age groups)}$$

where ASFR*a* = age-specific fertility rate for women in age group *a* (expressed as a rate per woman). The age specific fertility rate (ASFR) is the number of live births per 1000 women in a specific age group for a specified geographic area and for a specific point in time, usually a calendar year.

$$ASFR = \frac{\text{Number of live births to women in specified age group}}{\text{Number of women in same age group}} \times 1000$$

2. APPLICATION OF SMALL AREA STATISTICS

As already mentioned, the information on TFR was not available at district level and thus required choosing an appropriate covariate meeting following requirements:

1. The information on covariate is available at district level.
2. The covariate is highly correlated with TFR and the information on both is available at State level.
3. The covariate should be such that it could be monitored and tracked.

Of various indicators for which information was available at district level, the proportion of third and higher order births (B3+) was seen to be highly correlated with TFR. This indicator reflects what proportions of women have 3 or more children. Higher the value of B3+ consequently higher would be TFR. The complement of B3+ is proportion of women who have up to two children only i.e. the two child norm.

3. IDENTIFICATION OF DISTRICTS WITH HIGH TFR USING SMALL AREA STATISTICS

The regression equation of TFR with third and higher order births (B3+) as derived on NFHS-2 data is as under:

$TFR = 0.321069+0.057467*B3$, $R^2=0.78503$

Evidently above relation has high coefficient of determination. Based on the relationships of TFR with third and higher order births and district level values of third and higher order births, the number of districts having TFR 3.5 or more were estimated as 133 with their location in states as under:

Table 2. Distribution of number of district with TFR 3.5 or more

States	Total No. of Districts	No. of Districts with TFR >=3.5
Bihar	43	33
Madhya Pradesh	45	3
Odisha	30	0
Rajasthan	30	9
Uttar Pradesh	68	53
North East	67	28
Haryana	18	1
J & K	13	6
Total	504	133

***Genesis of Problem*:** Recently data on TFR at district level has become available for all Empowered Action Group States which are the extended states of BIMARUO. It is therefore important to verify as to what extent the theory of Small Area Statistics provided the accurate and reliable result. This paper is an attempt in this direction.

4. MATERIAL AND METHOD

Empowered Action Group States

The Empowered Action Group (EAG) set up to facilitate preparation of area-specific programmes in eight States, namely, Bihar, Jharkhand, MP, Chhattisgarh, Odisha, Rajasthan, UP and Uttarakhand, which have lagged behind in containing population growth to manageable levels.

Table 3. Population and number of district in EAG states

States	No. of Districts	Total Population
EAG States	261	796178
Bihar	37	11824
Chhattishgarh	16	181544
Jharkhand	18	329125
Madhya Pradesh	45	45584
Odisha	30	23391
Rajasthan	32	34061
Uttarakhand	13	24214
UP	70	146435
Other States	332	976862
Total	593	1773040

Though the percentage of the districts under EAG States is 44% accounting for 45% of Country population, their share in Births is much higher (60%).

***Annual Health Survey*:** Recently the information on TFR at district level has become available through Annual Health Survey. The Annual Health Survey (AHS) in India is the largest demographic survey in the world. This survey was conceived at the behest of National Commission on Population, Prime Minister Office and Planning Commission to provide bench mark basic vital and health indicators to map the levels and changes in all the districts of Empowered Action Group (EAG) of eight states viz. Bihar, Jharkhand, Madhya Pradesh, Chhattisgarh, Odisha, Rajasthan, Uttar Pradesh and Uttrakhand. The sample size was estimated for estimating the critical vital parameter Infant Mortality Rate (IMR) at district level with desired confidence level and precision. The numbers of households covered per district were on the average around 14500 HHs. With these sample size AHS covered around 3.8 million households (18.4 million Populations) in 261 districts across all the EAG States. Importantly, the Total Fertility Rate (TFR) at district level has been estimated with much higher precision than IMR.

The fact sheet of AHS was released and is in public domain for the year 2010-11 which has been used for the study.

It is of importance now to see how accurate would be the findings of small area statistics. Accordingly the small area statistics methodology was repeated to estimate TFR values at district level for EAG States and compare with the actual available through AHS.

5. RESULTS AND DISCUSSION

The findings are discussed as under:

***Aggregate Results for EAG States*:** The regression equation of TFR with the B3+ based on data at the state level for EAG States is as under-

$$TFR=0.7500+0.0590*B3,\ R^2=0.7890$$

Based on this relationship and utilizing the value of covariate B3+, the TFR value been estimated. The comparison of results with actual is presented in Table 4.

In-spite of high coefficient of determination ($R^2 = 0.78$), the results using methodology of small area statistics do not provide reliable results. In about one-third of the cases the difference between estimated and actual values of TFR is 0.5 or more. In fact only in about 10% of cases there is completed agreement and another one-third cases within 0.2 points. Thus, it would be seen that there is a wide divergence in large number of cases making the use of small area statistics less irrelevant.

Table 4. Comparison of estimated TFR with actual TFR

Difference in Estimated and Actual TFR	Number of Districts	%age
0	25	9.6
±0.1	46	17.6
±0.2	43	16.5
±0.3	29	11.1
±0.4	34	13.0
±0.5	30	11.5
±0.6	18	6.9
> ±0.6	36	13.8
	261	100.0

Many a times results are required to be presented in terms of categories i.e. low TFR (<2.1), moderate TFR (2.1-2.5), high TFR (2.5-3.0) and very high TFR (>3.0). Accordingly the result has been presented in a cross table: Actual versus Estimated according to different TFR Categories.

Table 5. Cross Tabulation Actual Versus Estimated according to TFR categories

		TFR Category-Actual				Total
		<2.10	2.1-2.5	2.5-3.0	3.0-3.5	
TFR Category-Estimated	<2.10	5	9	2	1	17
	2.1-2.5	15	24	15	0	54
	2.5-3.0	1	11	35	25	72
	3.0-3.5	0	1	20	97	118
Total		21	45	72	123	261

Even in terms of categories in 62% cases (161 out of 261) there is match at broad groups (difference upto 0.5). This analysis puts a question mark in the use of small area statistics.

State wise results for EAG States: The reasons for poor reliability of small area estimates could be due to variations in the relationship between TFR and covariate at the state level. In order to examine this, the methodology has been used by using the state specific relationship- The regression equations of TFR with the covariate B3+ for different state are presented below.

Table 6. Relationship of TFR with B3+

State	R^2	Constant	Beta
Bihar	0.5140	-0.9790	0.1000
Chhattishgarh	0.3070	1.7680	0.0340
Jharkhand	0.7400	0.5420	0.0630
Madhya Pradesh	0.4390	1.4510	0.0470
Odisha	0.4020	1.4490	0.0310
Rajasthan	0.7330	1.0870	0.0560
Uttarakhand	0.3820	0.9410	0.0400
UP	0.3610	1.2510	0.0480
Combined-EAG State at district level	0.6110	1.0410	0.0540
EAG State-State Level	0.7890	0.7500	0.0590

Evidently the R^2 values vary widely from low of 0.30 for Chhattisgarh to high of 0.74 for Jharkhand. The regression coefficient which was around 0.06 for EAG state combined has been observed low (about half of this value) or Chhattishgarh (0.03), Odisha (0.03) and high (about little less than double) for Bihar (0.10).

Here complete agreement in estimated and actual TFR values has been observed in about 12% cases and within ±0.2 in about 43% cases. Taking these together in about 54% cases, there is broad matching within ±0.2. However, this matching varies from low of 33% in UP to 81% in Chhattisgarh. There seems to be no correlation between R^2 and matching of results.

Table 7. State-wise result on comparison of estimate and actual TFR

Difference in Estimated and Actual TFR	Number of Districts									%age
	Bihar	Chhattisgarh	Jharkhand	Madhya Pradesh	Odisha	Rajasthan	Uttar Pradesh	Uttarakhand	Total	
0	5	3	3	1	6	5	5	2	30	11.5
±0.1	7	3	4	7	7	6	5	5	44	16.9
±0.2	8	7	7	12	8	10	13	2	67	25.7
±0.3	5	1	2	7	4	5	10	3	37	14.2
±0.4	5	0	2	5	4	2	9	1	28	10.7
±0.5	4	1	0	7	0	3	9	0	24	9.2
±0.6	2	0	0	2	0	1	9	0	14	5.4
> ±0.6	1	1	0	4	1	0	10	0	17	6.5
	37	16	18	45	30	32	70	13	261	100.0
Proportion of district with difference within (±0.2)	54.1	81.3	77.8	44.4	70.0	65.6	32.9	69.2	54.0	54.1
R^2Value	0.51	0.31	0.74	0.44	0.40	0.73	0.36	0.38	0.79	

Table 8. State-wise cross tabulation actual versus estimated according to TFR categories

		TFR-Actual						TFR-Actual				
Bihar							**Odisha**					
	Category	**<2.1**	**2.1-2.5**	**2.6-3.0**	**>3.0**	**Total**		**<2.1**	**2.1-2.5**	**2.6-3.0**	**>3.0**	**Total**
TFR-Estimated	<2.1	0	0	0	0	0		3	2	0	0	5
	2.1-2.5	0	0	0	0	0		8	11	2	1	22
	2.5-3.0	0	0	1	1	2		0	1	2	0	3
	>3.0	0	0	3	32	35		0	0	0	0	0
	Total	0	0	4	33	37		11	14	4	1	30
Chhattisgarh							**Rajasthan**					
	Category	**<2.1**	**2.1-2.5**	**2.6-3.0**	**>3.0**	**Total**		**<2.1**	**2.1-2.5**	**2.6-3.0**	**>3.0**	**Total**
TFR-Estimated	<2.1	0	0	0	0	0		0	0	0	0	0
	2.1-2.5	0	4	0	0	4		0	2	3	0	5
	2.5-3.0	0	3	7	2	12		0	2	11	0	13
	>3.0	0	0	0	0	0		0	0	2	12	14
	Total	0	7	7	2	16		0	4	16	12	32
Jharkhand							**Uttar Pradesh**					
	Category	**<2.1**	**2.1-2.5**	**2.6-3.0**	**>3.0**	**Total**		**<2.1**	**2.1-2.5**	**2.6-3.0**	**>3.0**	**Total**
TFR-Estimated	<2.1	0	0	0	0	0		0	0	0	0	0
	2.1-2.5	0	2	1	0	3		0	1	0	0	1
	2.5-3.0	0	2	7	0	9		0	3	3	7	13
	>3.0	0	0	2	4	6		0	1	13	42	56
	Total	0	4	10	4	18		0	5	16	49	70
Madhya Pradesh							**Uttarakhand**					
	Category	**<2.1**	**2.1-2.5**	**2.6-3.0**	**>3.0**	**Total**		**<2.1**	**2.1-2.5**	**2.6-3.0**	**>3.0**	**Total**
TFR-Estimated	<2.1	0	0	0	0	0		8	1	0	0	9
	2.1-2.5	0	3	3	1	7		1	2	1	0	4
	2.5-3.0	1	5	5	4	15		0	0	0	0	0
	>3.0	0	0	6	17	23		0	0	0	0	0
	Total	1	8	14	22	45		9	3	1	0	13

Based on these results the matching of results at broad groups improves to about 69% (179 out 261). The results of matching at broad group level for different state is as under-

Table 9. State-wise percentage of districts matching at broad broup

State	Total Number of Districts	Number of Districts with Matching at Broad Group	% of Districts with Matching at Broad Group
Bihar	37	32	86.5
Chhattisgarh	16	11	68.8
Jharkhand	18	13	72.2
Madhya Pradesh	45	25	55.6
Odisha	30	16	53.3
Rajasthan	32	25	78.1
Uttar Pradesh	70	46	65.7
Uttarakhand	13	10	76.9

The matching is low for Odisha and Madhya Pradesh. The high matching for Bihar is because of higher number of district falling in the category TFR > 3.0 and the same is true for UP and some other states.

6. CONCLUSION

The analysis presented in this paper raises questions on the validity of results using methodology of small area statistics even with high values of R^2, the results match only in about one-third of the cases. Further, doing the analysis at lower level of aggregation at state level the matching improves to 54 percent. Thus the estimate of TFR through SAE method is not accurate. This gives further scope to explore the other various SAE methods applicable on such data.

REFERENCES

Bose, Ashish - The man who coined the term *"Bimaru". (www. Livemint. com).*

International Institute for Population Sciences, *National Family Health Survey (NFHS-2), 1998-99: India. Mumbai: IIPS; 2000.*

Singh, Padam and Adhikari, Tulsi (2006). Identification of demographically weak states and districts requiring special attending for achieving targets of TFR 2.1 in India. *J Empirical Res. Soc. Sci.,* 1.

Office of Registrar General of India, *Annual Health Survey* (*2010-11*)*: Government of India,*

Office of Registrar General of India. *Census of India* (*2001*)*: Government of India*

National Population Policy (2000). *Department of Family Welfare, Ministry of Health & Family Welfare, Government of India.*