RESEARCH BRIEF

Interpretation of Rotavirus-positivity Patterns Across India

S VENKATASUBRAMANIAN, CP GIRISH KUMAR AND SANJAY MEHENDALE
From National Institute of Epidemiology, Indian Council of Medical Research, Chennai, India.

Correspondence to:
Dr S Venkatasubramanian,
National Institute of Epidemiology,
Indian Council of Medical Research,
II Main Road, TNHB, Ayapakkam,
Chennai 600 077, India.
subramanianv89@yahoo.co.in
Received: June 10, 2015;
Initial review: September 26, 2015;
Accepted: May 05, 2016.

Objective: To analyze variation in rotavirus-positivity using simple alternative statistical measures.

Methods: Hospital-based rotavirus surveillance among children admitted with acute gastroenteritis between 2005 and 2009. Prevalence, adjusted proportions and symmetrized index were calculated.

Results: Rotavirus prevalence was 40% (range 37% - 44%). Adjusted proportion analysis revealed higher level of deviation from annual prevalence in seasons (December – February and September – November); age groups (<12 months and 12-23 months) and regions (East and South). Analysis of symmetrized index revealed higher estimates of variation in all years, except in 2006.

Conclusion: Proposed statistical measures are useful as refined measures to study extent of disease spread in surveillance programmes, aiding evaluation of the load and pattern of disease burden in different regions over time.

Keywords: *Diarrhea, Epidemiology, Prevalence, Rotavirus infection*.

In developing countries, rotavirus-associated gastroenteritis is a leading cause of morbidity and mortality in early childhood. The burden of rotavirus gastroenteritis is highest among very young children and decreases rapidly thereafter. In temperate areas, seasonality of rotavirus infection with a single winter peak is well documented. In tropical climates, seasonality is less defined, with occasional increase noticed during cold and dry periods [1, 2].

The Indian 'National Rotavirus Surveillance Network' (NRSN) was initiated in 2005 by the Indian Council of Medical Research (ICMR) as a collaborative task-force study to generate nationally representative data on rotavirus burden and its molecular epidemiology in India. The multi-site, hospital- based surveillance has been carried out in two rounds (Round 1: 2005-2009 and Round 2: 2012-2016). The data from 2005-2009 has been previously analyzed to report association of rotavirus positivity with age, gender and season as well as the distribution of specific genotypes [3,4].

With multi-site data sets, the application of appropriate statistical measures to more meaningfully interpret national level data is important. Such analyses facilitate interpretation of variations by years, by regions, by population characteristics like age and gender, and by seasonality. These may be useful in planning and implementing appropriate prevention and control strategies in different parts of the country. We used two new methodological approaches namely 'Adjusted Proportion' and 'Symmetrized Index' (adjusted variation) to assess temporal, geographical and demographical variations in disease patterns, and have illustrated the application of these measures in the context of rotavirus gastroenteritis in India.

METHODS

In all, 6954 cases of children hospitalized with acute gastroenteritis enrolled in NRSN Round 1 (2005-2009) for which complete data were available have been included in this analysis. Details of study setting, location of sites, and case recruitment in NRSN have been published earlier [3,4]. Data was aggregated as different age groups (<12 months, 12-23 months, 24-35 months, ≥36 months), gender (male, female), seasons (December 6 February, March 6 May, June-August and September 6 November), and Regions (East, West, South, North) for statistical analysis.

In addition to calculating the prevalence of rotavirus positivity (a measure of rotavirus disease burden), to understand the disease pattern in different regions over time and across different groups (levels of aggregations), we calculated the following two statistical measures:

(*i*) Adjusted Proportion, a ratio of two independent binomial proportions can be considered a

standardized ratio (Details provided in *Web Appendix* I). Proportion of rotavirus positive cases by average proportion of rotavirus positive cases was calculated for in various domains / level of aggregations *viz*. seasons and annual time points. Confidence Intervals were calculated [5].

(*ii*) Symmetrized Index (SI), is proposed as an indicator of variability in disease and represents the difference between the highest and lowest prevalences adjusted for extreme observations {[max/(max+min)] – [min/(max+ min)]}. Standard Analysis of Variance (ANOVA) was used to assess the equality of the proposed SI.

RESULTS

Rotavirus prevalence for the period 2005-2009 ranged from 37% to 44% with maximum, minimum and average rotavirus positivity varying across the regions both annually (***Web Fig.* 1**; ***Table* I**) and seasons during the survey period (***Table* I**). The minimum rotavirus prevalence (13.8%) was observed in the third season of 2007 and the maximum prevalence (88.9%) was observed in first season of 2006. The annual average prevalence ranged from 36% (95% CI 29.5- 43.6) to 44% (95% CI 39.6 – 48.9). In the Eastern region during June-Aug of 2007, rotavirus positivity was nil and maximum prevalence of 81% was observed in September – November of 2006 (data not shown). In the Western region, lowest prevalence of 5% was seen during June-August of 2007 and maximum prevalence of 80% was observed in September-November of the same year. In the Southern region, lowest and highest prevalences were 19% (March -May) and 68% (December-February) of 2008, respectively. In the Northern region, lowest positivity was seen in September -November (13%) of 2008 whereas the highest prevalence was observed in December- February (87%) of 2009 (data not shown).

TABLE I ANNUAL AND SEASONAL PREVALENCE OF ROTAVIRUS DISEASE IN INDIA

India Seasons	*Year*	*Max (%)*	*Min (%)*	*Prevalence (95% CI)*	*SI*
Dec - Feb	2006	88.89	49.35	55.70 (55.22-56.18)	29
	2007	55.64	51.85	54.61 (54.61-54.62)	04
	2008	68.66	51.42	59.33 (59.21-59.45)	14
	2009	51.87	41.70	46.21 (46.15-46.27)	11
Mar-May	2006	60.00	50.00	55.81 (55.77-55.85)	09
	2007	53.33	28.57	37.74 (37.49-37.99)	30
	2008	43.01	20.90	34.58 (34.38-34.78)	35
	2009	46.46	14.63	27.55 (27.13-27.96)	52
Jun-Aug	2006	38.93	28.66	32.47 (32.43-32.51)	15
	2007	22.96	13.79	19.16 (19.12-19.19)	25
	2008	30.46	27.57	28.73 (28.73-28.74)	05
	2009	42.75	42.75	42.75 (42.75-42.75)	00
Sep-Nov	2006	55.13	40.68	49.74 (49.65-49.82)	15
	2007	63.69	32.11	49.36 (48.95-49.77)	33
	2008	45.53	34.36	40.00 (39.95-40.05)	14
Annual	2006	60.00	27.14	44.24 (39.6-48.9)	38
	2007	63.69	13.79	36.56 (29.5-43.6)	64
	2008	68.66	24.00	39.49 (33.2-45.8)	48
	2009	51.87	22.50	39.07 (34.9-43.2)	39

SI: symmetrized index.

Using adjusted proportions, differences in rotavirus positivity for various domains (annual, seasons, age group *etc*.) adjusted for reference points (domain prevalence/ mean) with 95% CI are shown in ***Table* II**. Deviation from the annual prevalence (adjusted proportion of >1.0) was observed for the season December-February and September – November in all the regions (***Table* II**). A similar phenomenon was observed for the age groups of <12 months and 12-23 months and in the Eastern and Northern regions.

Symmetrized Index (SI) varied from 4% to 52% during the surveillance period across regions when the rotavirus positivity was accounted for the seasonal periodicity (***Table* I**). For the entire study period, the calculated estimate SI was 67%. When accounting for annual periodicity the SI ranged from 38% to 64%, highlighting the high rotaviral disease burden. ***Web Fig.* 2** depicts the annual-seasonal and regional distribution of SI. Analysis of variance performed on SI revealed that there was no variation among SI accounting for seasons and age of children during the study period (data not shown).

DISCUSSION

In the present surveillance data, although hospitalization with rotavirus gastroenteritis occurred throughout the year, more hospitalizations were seen during months of December to February. Fluctuations in rotavirus disease burden can be detected and modeled using a range of statistical tools [6]. Measurements in terms of rates, percentages, proportions *etc*. can be calculated for different regions or for domains (levels of aggregations) that are mutually exclusive and exhaustive. Patterns become evident when such quantitative measures are compared between various domains, but such summarizations involve loss of information [7]. To compensate for this loss, normalization of data or adjusting for average prevalence or symmetrization of

WHAT THIS STUDY ADDS?

- Adjusted proportion helped in identifying deviations from annual prevalence among regions and seasons.
- Symmetrized index provided rotavirus disease variability estimate of 67%.

extreme observations in prevalence will result in refinement in measurements [8].In the present analysis, comparison of estimates derived by the usual prevalence measurement and the symmetrized index revealed that symmetrization resulted in a higher estimate of 67%. Similar increase or decrease in estimates was observed when SI and prevalence was analyzed for different seasons and regions. By accounting for the difference in extreme observations, the analysis shows that the symmmetrized index will provide more realistic estimates of rotavirus disease variability. This "folded fraction" is particularly useful in settings where the range is large and outlier observations are more.

TABLE II ADJUSTED PROPORTIONS FOR ROTAVIRUS POSITIVITY ACROSS DOMAINS

Levels	*Observed Rotavirus cases*	*Number tested*	*Rotavirus positivity (%)*	*Adjusted Proportion [95% CI]*
Year				
2006	806	1822	44	1.11 [1.04 - 1.17]
2007	623	1704	37	0.92 [0.85 - 0.98]
2008	904	2289	39	0.99 [0.93 - 1.05]
2009	445	1139	39	0.98 [0.9 - 1.06]
	Total	2778	6954	40
Season				
Dec-Feb	1079	1982	54	1.36 [1.3 - 1.43]
Mar-May	607	1909	32	0.8 [0.74 - 0.86]
Jun-Aug	476	1708	28	0.7 [0.64 - 0.76]
Sep-Nov	616	1355	45	1.14 [1.07 - 1.21]
	Total	2778	6954	40
Age (m)				
<12 m	1417	3352	42	1.06 [1.01 - 1.11]
12-23 m	1032	2247	46	1.15 [1.09 - 1.21]
24-35 m	207	696	30	0.74 [0.66 - 0.84]
>=36 m	122	659	19	0.46 [0.39 - 0.55]
	Total	2778	6954	40
Gender				
Male	1750	4375	40	1 [0.96 - 1.05]
Female	1028	2579	40	1 [0.94 - 1.05]
	Total	2778	6954	40
Region				
East	769	1835	42	1.05 [0.99 - 1.12]
West	919	2570	36	0.90 [0.83 - 0.94]
South	753	1678	45	1.12 [0.93 - 1.06]
North	337	871	39	0.97 [1.25 - 1.46]
	Total	2778	6954	40

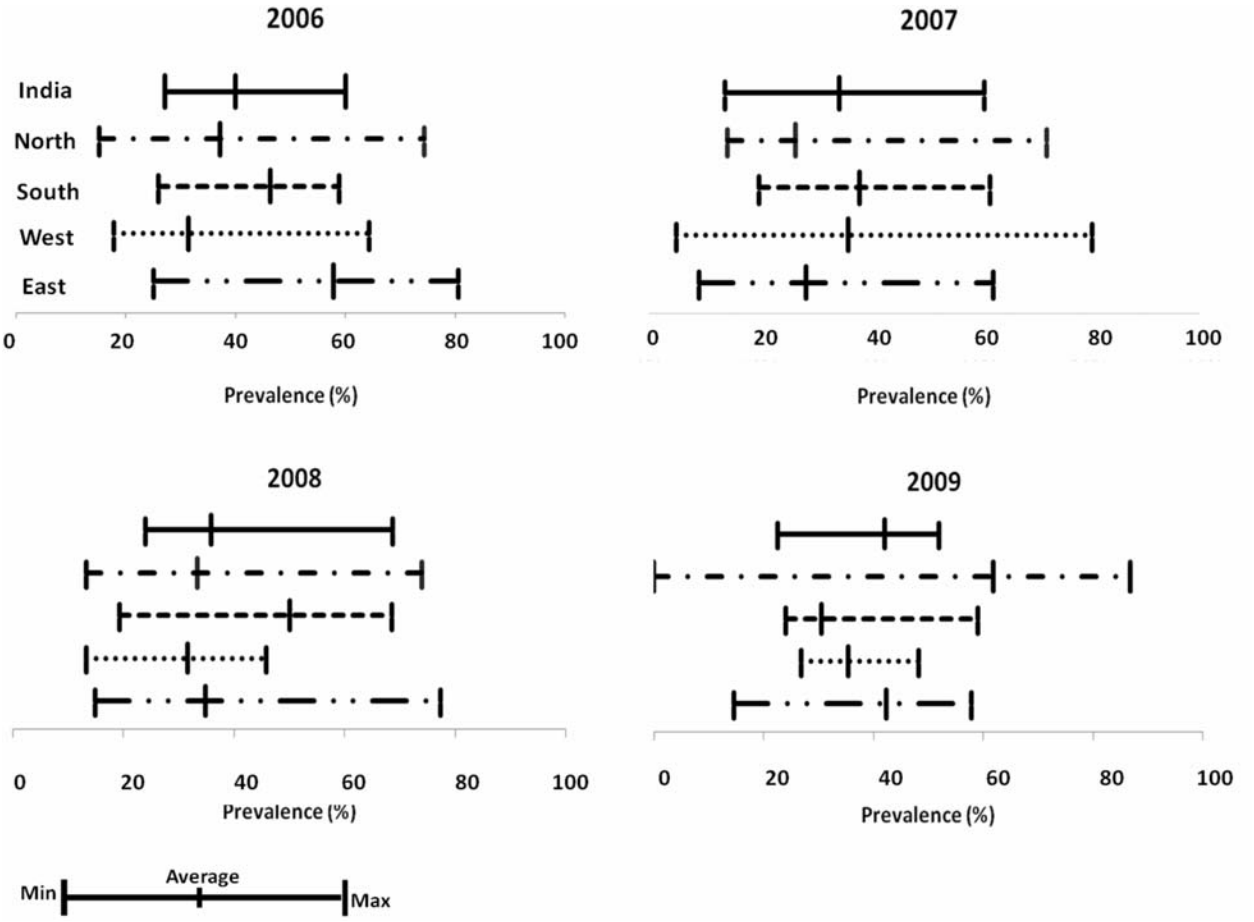
The graphical depiction of prevalence in this paper aids understanding of the distribution of the disease burden. Adjusted proportion analysis explains the variations among differences in annual or seasonal and other such data categorization and calculation of confidence intervals for adjusted proportions adds value to such estimates. This ratio reduces noise from the settings in which little or no rotavirus was observed. It also gives an indication of greater deviation from the average whenever this ratio is greater than one. Usually calculation of confidence intervals is done for simple proportions. The recent development of calculating confidence intervals for ratio of proportions allows application of the indirect test of significance for equality [5]. Using adjusted proportions, we have shown that one can easily identify regions / seasons where the deviation in rotaviral prevalence is more than the average. Therefore this measure can be used to generate pattern of rotaviral disease in the country showing the differences in domains over time.

In conclusion, the analysis using these newer and simple statistical measures has potential for application in surveillance of other major infectious diseases such as meningitis or pneumonias wherein evaluation of the load and pattern of disease burden in different regions over time can be carried out.

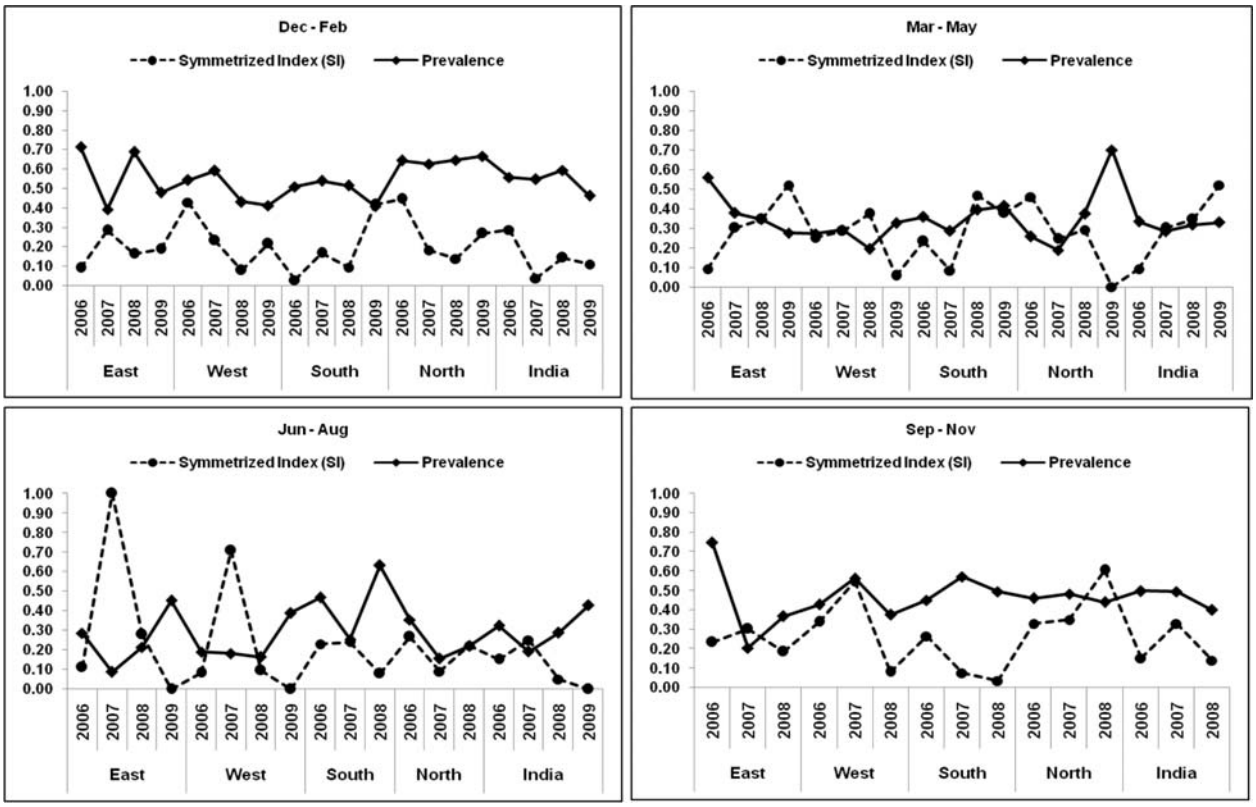
Acknowledgements: M Chiranjeevi, Technical Assistant, NRSN project team at NIE for support in statistical analysis. Dr Gagandeep Kang for reviewing and providing valuable inputs.
Contributors: SV: manuscript conceptualization and development, data analysis; CPGK: concept refinement, data interpretation and manuscript writing; SMM: concept refinement, data interpretation and manuscript writing. All authors approved the final manuscript.
Funding: Indian Council of Medical Research; *Competing Interests*: None

REFERENCES

1. Bresee J, Glass RI, Ivanoff B, Gentsch J. Current status and future priorities for rotavirus vaccine development, evaluation, and implementation in developing countries. Vaccine. 1999;17:2207-22.
2. Jagai JS, Sarkar R, Castronovo D, Kattula D, McEntee J, Ward H, *et al*. Seasonality of rotavirus in South Asia: A meta-analysis approach assessing associations with temperature, precipitation, and vegetation index. PLoS One. 2012;7:e38168.
3. Kang G, Arora R, Chitambar SD, Deshpande J, Gupte MD, Kulkarni M, *et al*. Multicenter, hospital-based surveillance of rotavirus disease and strains among Indian children aged <5 years. J Infect Dis. 2009;200:S147-53.
4. Kang G, Desai R, Arora R, Chitambar S, Naik TN, Krishnan T, *et al*. Diversity of circulating rotavirus strains in children hospitalized with diarrhea in India, 2005-2009. Vaccine. 2013;31:2879-83.
5. Fagerlund MW, Lydersen S, Laake P. Recommended confidence intervals for two independent binomial proportions. Stat Methods Med Res. 2015;24:224-54.
6. Chui KKH, Jagai JS, Griffiths JK, Naumova EN. Hospitalization of the Elderly in the United States for Nonspecific Gastrointestinal Diseases: A Search for Etiological Clues. Am J Public Health. 2011;101:2082-6.
7. Keppel K, Pamuk E, Lynch J, Carter-Pokras O, Kim I, Mays V, *et al*. Methodological issues in measuring health disparities. National Center for Health Statistics. Vital Health Stat. 2005;2:141.
8. To T, Williams JI, Wu K, Theriault ME, Goel V. Comparison of methods to identify outliers observed in health services small area variation studies. Stat Methods Med Res. 2003;12:531-46.



WEB FIG. 1 *Annual and regional prevalence of rotavirus disease in India showing prevalence (minimum and maximum).*



WEB FIG. 2 *Annual - Seasonal and regional distribution of rotavirus disease variability shown as prevalence and symmetrized index.*

WEB APPENDIX 1

Explanation for proving that adjusted proportion is a standardized ratio.

To test the null hypothesis that the expected rotavirus positive rates is same in all regions / levels of aggregations and that the difference in observed rates are no bigger than that expected by chance

H_0:p_1= p_2 =…. = P_j where j = number of regions / levels of aggregations, where $P_j = Y_j / n_j$,(number of rotavirus positive cases in j / total number of diarrheal cases tested in j).

Let $\hat{p} = \sum_{j=1}^{J} Y_j \Big/ \sum_{j=1}^{J} n_j$ be the average prevalence of rotavirus.

Under the null hypotheses the expected no of rotavirus positives in regions / levels of aggregations j, is

$n_j \hat{p}$. i.e. $E_j = n_j \hat{P} = \frac{n_j \cdot \sum Y_j}{\sum n_j}$

Now, $\frac{Y_j}{E_j} = (Y_j) \Big/ \left(\frac{n_j \sum Y_j}{\sum n_j} \right) = Y_j \cdot \frac{\sum n_j}{n_j \sum Y_j} = \frac{Y_j}{n_j} \cdot \frac{\sum n_j}{\sum Y_j}$ (1)

$\frac{P_j}{\hat{P}} = \left(\frac{Y_j}{n_j} \right) \Big/ \left(\frac{\sum Y_{ji}}{\sum n_j} \right) = \frac{Y_j}{n_j} \cdot \frac{\sum n_j}{\sum Y_j}$ (2)

From the above two equations (1) and (2), we have

$\frac{P_j}{\hat{P}} = \frac{Y_j}{E_j}$ 'Standardized ratio'