

RESEARCH ARTICLE

Open Access



# Diversity in substance use behaviour among street children of Delhi under Bayesian paradigm

Vivek Verma<sup>1</sup>, Ashwani Kumar Mishra<sup>2\*</sup>, Anju Dhawan<sup>2</sup> and Dilip C. Nath<sup>3</sup>

## Abstract

**Background:** Shannon's index is one of the measures of biodiversity, which is intended to quantify both richness and evenness of the species/individuals in the ecosystem or community. However, application of Shannon's index in the field of substance use among the street children has not been done till date.

**Methods:** This paper is concerned with methods of estimating Shannon's diversity index (SDI), which can be used to capture the variation in the population due to certain characteristics. Under the consideration that the probability of abundance, based on certain characteristics in the population, is a random phenomenon, we derive a Bayesian estimate in connection with Shannon's information measure and their properties (mean and variance), by using a probability matching prior, through simulation and compared it with those of the classical estimates of Shannon. The theoretical framework has been applied to the primary survey data of substance use among the street children in Delhi, collected during 2015. The measure of diversity was estimated across different age profiles and districts.

**Results:** The results unrevealing the diversity estimate for street children corresponding to each region of Delhi, under both the classical and Bayesian paradigms. Although the estimates were close to one another, a striking difference was noted in the age profile of children.

**Conclusions:** The Bayesian methodology provided evidence for a greater likelihood of finding substance-using street children, belonging to the lower age group (7-10, maximum Bayesian entropy-3.73), followed by the middle (11-14) and upper age group (15-18). Moreover, the estimated variance under the Bayesian paradigm was lesser than that of the classical estimate. There is ample scope for further refinement in these estimates, by considering more covariates that may have a possible role in initiating substance use among street children in developing countries like India.

**Keywords:** Response driven sampling, Street children, Shannon diversity index, Prior distribution, Posterior distribution

## Background

In ecology, a diversity index is a parameter of interest intended to measure biodiversity within an ecosystem [1]. The widely used diversity indices are Shannon's index and Simpson's index. They can be used to assess the diversity of any population in which each member belongs to a

unique species. Shannon's diversity index (SDI) has been derived from the information theory, originally in the context of information in telephone systems. The SDI measures enables in ascertaining the membership of species chosen randomly to a particular class from the collection of all possible classes. This average uncertainty increases as the number of species increases. It attains maximum as the relative abundances of species are equal. A diversity index is a mathematical measure of species diversity in a community. Diversity indices provide more information

\*Correspondence: [akmaiims@gmail.com](mailto:akmaiims@gmail.com)

<sup>2</sup>National Drug Dependence Treatment Centre (NDDTC), All India Institute of Medical Sciences (AIIMS), New Delhi, India

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

about community composition than simply species richness (i.e., the number of species present); they also take the relative abundances of different species into account. The term diversity although seems to be terminology of an ecosystem, its analogous statistical terminology is dispersion. Hence, application of a diversity index can be reviewed as a mathematical measure of species or a particular community under investigation, which facilitates precise delimitation of community composition, which can be characterized by the statistical measures of Mean and Variance.

In earlier studies, estimation of SDI using various techniques viz., the maximum likelihood estimator [2], jack-knife method [3], combined Horvitz-Thompson estimator [4] and sample coverage [5], were discussed by various authors. SDI [6] has two properties: (i) Shannon’s index is zero if and only if there is one specie in the community, and (ii) Shannon’s index is maximum only when all species are present within a studied ecosystem with the same probabilities. Typically the values lies between 1.5 and 3.5 in most ecological studies [7, 8], and the index is rarely greater than 4. SDI increases as both the richness and the evenness of the community increase. The measure greater than 1 implies more diversity and thereby greater likelihood of finding more species or individuals with characteristics under investigation, for example in this study substance using street children. Higher the value of indices more is the richness in terms of the species/individuals.

In the present article, we derived a probability matching prior for SDI under Bayesian paradigm, and demonstrated the relevance of posterior estimate through numerical simulation and real-life application, and compared with classical estimate of SDI. The primary data used for the study was collected during August, 2015 to November, 2015 through a representative survey by implementing a Respondent Driven Sampling (RDS) among street children in all nine districts of Delhi. The suggestive methodology is used to demonstrate the derivable benefits in terms of better characterization of age profile and variation across districts of substance using street children in Delhi.

**Methods**

The methodology of the present article has been elaborated by mathematical formulation of SDI, under the classical and Bayesian paradigm. Under both suggested analytical approaches, the derivations of mean and variance have been used to demonstrate its performance for describing the diversity in population with certain characteristic. The comparative performance of these estimators has been evaluated and presented, by applying the methodology on real life based data of substance using children in Delhi.

**Mathematical formulation of SDI under classical paradigm**

Suppose that there are  $S$  number of different species in a community and  $S$ , is assumed to be a known quantity and is relatively small. Let  $p_i$  denotes the probability of richness of the  $i^{th}$  species in the community, for all  $i = 1, 2, \dots, S$  and is such that  $\sum_{i=1}^S p_i = 1$ . One of the widely used mathematical measures of species diversity in a ecological community is the SDI, which is the linear function of probabilities,  $p_1, p_2, \dots, p_s$ , defined as

$$H(\mathbf{p}) = - \sum_{i=1}^s p_i \log(p_i). \tag{1}$$

Suppose that a random sample of  $n$  has been collected and corresponding to each selected unit let  $X_{ij}$  is an indicator function which takes value ‘1’ if the  $j^{th}$  selected unit belonging to  $i^{th}$  species and ‘0’ otherwise, for all  $j = 1, 2, \dots, n$  and  $i = 1, 2, \dots, S$ . Let us define the variables  $Y_i = \sum_{j=1}^n X_{ij}$ , denoting the number of sampled units

belongs  $i^{th}$  species in the community. Further suppose that,  $X_1, X_2, \dots, X_S$  are independently distributed as  $Y_i \sim Binomial(n, p_i)$ . Under the assumption that the parameter  $p_i$ , the probability of abundance of the  $i^{th}$  species in the community, is an unknown but a fixed quantity, then by following the maximum likelihood principle  $p_i$  is estimated as  $\hat{p}_i = \frac{x_i}{n}$ . Based on the estimated  $p_i$ , say  $\hat{p}_i$ , the mean and variance of SDI,  $H(\mathbf{p})$ , have obtained as

$$\begin{aligned} \hat{H}(\mathbf{p}) &= E(H(\mathbf{p})) = -E \left[ \sum_{i=1}^s p_i \log(p_i) \right] \\ &= -E \left[ \sum_{i=1}^s \frac{x_i}{n} \log \left( \frac{x_i}{n} \right) \right] \\ &= \frac{-1}{n} E \left[ \sum_{i=1}^s x_i (\log x_i - \log n) \right] \\ &= \frac{-1}{n} \sum_{i=1}^s E(x_i \log x_i - x_i \log n) \\ &= \frac{-1}{n} \sum_{i=1}^s (E(x_i \log x_i) - np_i \log n). \end{aligned} \tag{2}$$

$$\begin{aligned} V(\hat{H}(\mathbf{p})) &= V \left[ \sum_{i=1}^s p_i \log(p_i) \right] = V \left[ \sum_{i=1}^s \frac{x_i}{n} \log \left( \frac{x_i}{n} \right) \right] \\ &= \frac{1}{n^2} V \left[ \sum_{i=1}^s x_i (\log x_i - \log n) \right] \\ &= \frac{1}{n^2} \sum_{i=1}^s V(x_i \log x_i - x_i \log n) \\ &= \frac{1}{n^2} \sum_{i=1}^s (V(x_i \log x_i) + np_i(1-p_i) (\log n)^2). \end{aligned} \tag{3}$$

Let  $g(X_i) = x_i \log(x_i)$  and  $p_i$ 's are independent, then the  $E(g(X_i))$  and  $V(g(X_i))$  have derived using the Delta method [9] as

$$E(g(X_i)) \approx g(E(X_i)) + \frac{1}{2}g''(E(X_i))V(X_i) = np_i \log(np_i) + \frac{(1-p_i)}{2}, \tag{4}$$

$$V(g(X_i)) \approx [g'(E(X_i))]^2 V(X_i) = [1 + \log(np_i)]^2 np_i(1-p_i). \tag{5}$$

Using the above equations, the  $E(H(\mathbf{p}))$  and  $V(H(\mathbf{p}))$  is given by

$$E(H(\mathbf{p})) = \frac{-1}{n} \sum_{i=1}^s \left( np_i \log(np_i) + \frac{(1-p_i)}{2} - np_i \log n \right) = - \sum_{i=1}^s p_i \left( \log(np_i) + \frac{(1-p_i)}{2n} - \log n \right) \tag{6}$$

$$V(H(\mathbf{p})) = \frac{1}{n^2} \sum_{i=1}^s \left( [1 + \log(np_i)]^2 np_i(1-p_i) + np_i(1-p_i) (\log n)^2 \right) = \frac{1}{n} \sum_{i=1}^s p_i(1-p_i) \left( [1 + \log(np_i)]^2 + (\log n)^2 \right) \tag{7}$$

**Mathematical formulation of SDI under Bayesian paradigm**

In the previous section, the usual estimation procedure of SDI,  $H(\mathbf{p})$ , has been discussed, which is based on the assumption that the probability of abundance of the  $i^{th}$  species in the community,  $p_i$ , is an unknown but a fixed quantity. But in practical situations  $p_i$ 's might be a random quantity, and hence under this situation the randomness of SDI,  $H(\mathbf{p})$ , can be quantified under Bayesian paradigm by specifying a suitable prior distribution for  $H(\mathbf{p})$ . To suggest a prior distribution for a function of  $p_i$ 's,  $H(\mathbf{p})$ , is difficult to obtain directly. Here, we derived a probability matching prior, distribution for  $H(\mathbf{p})$  for a discrete binary population, which is free from any hyper-parameter. In principle, a probability matching prior holds the promise of providing a possible agreement between frequentist and Bayesian inferential procedures, which can be used for routine use in Bayesian inference [10, 11].

**Theorem 1** Suppose  $q(\cdot)$  defines the prior distribution for  $H(\mathbf{p}) = - \sum_{i=1}^s p_i \log(p_i)$ , a linear function of probabilities of abundance of different species in the community, then  $p_1, \dots, p_i, \dots, p_s$ , takes the following distributional form

$$q(\mathbf{p}) = q(p_1, \dots, p_s) \propto \left\{ \prod_{i=1}^s (1 + \log(p_i))^2 p_i(1-p_i) \right\}^{\frac{1}{2}} \prod_{i=1}^s [p_i(1-p_i)(1 + \log(p_i))]^{-1}. \tag{8}$$

*Proof* The proof of Theorem 1 is given in the Appendix. □

The posterior distribution of  $\mathbf{p}$  for the given information will then be

$$h(\mathbf{p}|data) \propto L(\mathbf{p}|data) q(\mathbf{p})$$

$$h(\mathbf{p}|\mathbf{x}) \propto \left\{ \prod_{i=1}^s (1 + \log(p_i))^2 p_i(1-p_i) \right\}^{\frac{1}{2}} \prod_{i=1}^s [p_i^{x_i}(1-p_i)^{n-x_i}(1 + \log(p_i))]^{-1}. \tag{9}$$

Here, the posterior distribution,  $h(\mathbf{p}|\mathbf{x})$ , does not have any explicit form and  $p_1, \dots, p_s$  are independent, for this reason one has to get samples from  $h(\mathbf{p}|\mathbf{x})$  to get the posterior distribution of  $H(\mathbf{p})$ . This is done by simulating  $N$  values from the posterior distribution as  $\{p_1^{(t)}, p_2^{(t)}, \dots, p_s^{(t)}; t = 1, 2, \dots, N\}$ , then by computing  $H^{(1)}(\mathbf{p}), H^{(2)}(\mathbf{p}), \dots, H^{(N)}(\mathbf{p})$ , where  $H^{(t)}(\mathbf{p}) = - \sum_{i=1}^s p_i^{(t)} \log(p_i^{(t)})$ . The procedure of Markov Chain Monte Carlo (MCMC) simulation technique is adopted to estimate the posterior mean and variance of  $H(\mathbf{p})$  and can be approximated as

$$\hat{H}(\mathbf{p})^B = E(H(\mathbf{p})|\mathbf{x}) \simeq \frac{1}{N} \sum_{t=1}^N H^{(t)}(\mathbf{p}) \tag{10}$$

and

$$V(\hat{H}(\mathbf{p})^B) = V(H(\mathbf{p})|\mathbf{x}) \simeq \frac{1}{N} \sum_{t=1}^N [H^{(t)}(\mathbf{p})]^2 - \left[ \frac{1}{N} \sum_{t=1}^N H^{(t)}(\mathbf{p}) \right]^2. \tag{11}$$

**Numerical study**

In the present section, we illustrate the proposed procedures numerically through a simulation study. For demonstration purpose we here considered that the probability of abundance, say  $p$  lies between  $\{(0.1,0.2), (0.3,0.5), (0.6,0.8)\}$ , we vary the number of species in the community  $S = 2, 3, 4$ , and number of collected samples

$n = 10, 50, 100$ . For different combinations of  $S, p, n$ , both classical and posterior estimates ( $H(\mathbf{p})$  and  $H(\mathbf{p})^B$ ) and their respective variances are computed. Corresponding to each of these combinations we computed both the classical and Bayesian estimates of SDI and reported through Table 1. Here, we computed the posterior estimates through MCMC simulation technique with the help of Metropolis-Hasting's algorithm (given in Appendix). The posterior estimate of SDI is obtained using simulation of  $N = 500,000$  samples, with burn-in period 10000, which is minimum number of sample required for the Markov Chain to reach stationarity. The obtained result under Table 1 shows that with increasing sample size, say  $n$ , under both the classical and Bayesian estimates of SDI got closer to one another, irrespective of number of species ( $n$ ) and probability of abundance ( $p$ ). Even with closer of estimated values of SDI, the posterior variance was detected to be significantly lower than that of the variance obtained under classical paradigm. In addition to that for all combinations of  $S; p; n$ , two estimators are more or less equally but the posterior variance of SDI under Bayesian set-up is again found to be significantly lower than that of classical variance of SDI.

### Application

In the field of ecology, the Shannon index is known alternatively as - Shannon's diversity index, the Shannon-Weiner index, and the Shannon entropy [12]. This index was coined by Claude Shannon to the problem of textual presentation of the strings for quantification of the entropy [13]. Conceptually, it was thought when there is occurrence of different alphabets and approximately equitable distribution of the relative abundances in the textual presentation, there is less likelihood of correctly predicting appearance of next string. Consequently, there is a phenomenon of randomness. Likewise, in the field of biology too we do have variation in the phenomenon of interest. Biological communities vary in the number of species that are present and this number of species corresponding to that community is defined as richness of species. In addition to richness of a species in a community, information about the relative abundant species is also important. It is a well-known that the value of a diversity index increases when the number of types of species increases and their evenness increases. For example, communities with many species that are evenly distributed are the most diverse and communities with few species that are dominated by one species are the least diverse. By considering this distribution we expect to achieve the exact level of quantification. This index has not been used in the field of drug use epidemiology to date. Hence, we thought of applying this index concerning substance use behaviour among the street children by considering the age and the geographical variation. Our logic was that since diversity

index is reflective of the number of different species and how evenly they are distributed, what incremental knowledge is achievable with respect to the substance using street children, where the correct prediction of finding any age-specific adolescent across different geographical areas is an onerous task. In this study, an attempt has been made to explore the diversity index of district viz., Central Delhi, East Delhi, New Delhi, North East Delhi, North West Delhi, South Delhi, South East Delhi, South West Delhi and West Delhi, substance use among street children in Delhi. The diversity across age-wise in these areas under the classical and Bayesian paradigms has also been evaluated. The sample size, inclusion criteria, implementation of RDS, seed plan, process of recruitment, management of coupons and incentives, data collection and ethical considerations can be found from our previous publication [14]. In this investigation due to the non-availability of the estimated size of street children across all the nine districts of Delhi, the district level representative figure of substance using street children could not be provided. The variation across different districts of Delhi is essential for developing local level intervention strategies. Hence, by applying theoretical concept of SDI, we attempted to provide more precisely the variation in the age profile of the street children.

### Results

Shannon measure is based on logarithmic scale and usually expressed in 2, 10 and e bases. For the computational purpose, we considered logarithm with base 2 in Eq. (1) for estimation of SDI of substance use under both classical and Bayesian paradigms. Here, the diversity index less than 1 denotes the low diversity and thereby low likelihood of finding substance using substance street children. Here, we have compared both classical and proposed Bayesian estimates of SDI of street children among three age groups viz., 7-10, 11-14 and 15-18 years, in nine areas of Delhi in Table 2. Table 2 depicts both age-wise and district wise street children. The range of SDI estimates, viz.  $\hat{H}(\mathbf{p}), \hat{H}(\mathbf{p})^B$ , under classical (0.86-1.57) and Bayesian (0.9-1.57) set-ups were close. This implies that across all regions of Delhi the likelihood of locating substance using street children cannot be undermined. On the other hand, in Central, Eastern, Southern, Western, North East and North Western regions of Delhi, both the estimated diversity indices values were found closer to 1.5, and South East and South West regions of Delhi were closer to 1.3. In case of age-wise diversity among children, the classical estimates for all three age groups viz., 7-10, 11-14 and 15-18 years, was detected to be close to 2.9. Whereas the Bayesian estimates exhibited more diversity in the index among age-group 7-10 years (3.74) then 11-14 years (3.42) children and comparatively less among 15-18 years children (3.20). A systematic pattern was also

**Table 1** Classical and Bayesian estimates of Shannon diversity index (SDI)

S	p	n	$\hat{H}(p)$				$\hat{H}^B(p)$				$V(\hat{H}(p))$				$V(\hat{H}^B(p))$			
			S				S				S				S			
			1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
2	(0.1,0.2)	10	1.37	1.53			1.86	1.86			1.16	1.19			0.040	0.030		
		50	1.82	1.90			1.77	1.94			1.06	0.84			0.040	0.010		
	(0.3,0.5)	10	1.76	1.88			1.78	1.94			1.17	1.05			0.046	0.014		
		50	1.96	1.90			2.01	2.02			0.96	1.01			0.007	0.007		
	(0.6,0.8)	10	1.96	1.90			2.01	2.02			0.96	1.01			0.007	0.007		
		50	1.96	1.97			2.04	2.02			0.69	0.57			0.003	0.002		
(0.1,0.2)	100	1.93	1.95			2.05	1.99			0.81	0.63			0.008	0.003			
3	(0.1,0.2)	10	1.73	2.09	0.76		2.54	2.67	2.45		1.14	1.18	0.74		0.147	0.011	0.160	
		50	2.52	2.54	2.57		2.74	2.68	2.69		0.82	0.81	0.75		0.010	0.023	0.022	
	(0.3,0.5)	10	2.46	2.54	2.55		2.75	2.82	2.85		1.02	1.02	0.97		0.007	0.001	0.050	
		50	2.56	2.57	2.58		2.75	2.65	2.74		0.51	0.52	0.47		0.004	0.013	0.007	
	(0.6,0.8)	10	2.55	2.54	2.57		2.60	2.69	2.65		0.85	0.90	0.83		0.021	0.051	0.024	
		50	2.55	2.56	2.58		2.60	2.65	2.76		0.59	0.59	0.53		0.018	0.013	0.009	
(0.1,0.2)	100	2.55	2.56	2.58		2.60	2.65	2.76		0.59	0.59	0.53		0.018	0.013	0.009		
4	(0.1,0.2)	10	2.63	2.88	2.88	2.16	3.33	3.66	3.45	3.37	1.12	1.14	1.10	1.20	0.039	0.030	0.048	0.039
		50	2.91	2.87	2.87	2.93	2.91	3.00	3.31	3.15	0.63	0.79	0.84	0.75	0.006	0.109	0.099	0.093
	(0.3,0.5)	10	2.88	2.76	2.80	2.92	2.95	3.21	3.37	3.13	0.85	0.99	1.03	0.95	0.023	0.058	0.073	0.082
		50	2.98	2.97	2.97	2.98	2.91	3.02	3.27	2.90	0.39	0.48	0.51	0.45	0.006	0.062	0.101	0.006
	(0.6,0.8)	10	2.96	2.96	2.95	2.96	3.00	2.90	3.02	2.94	0.73	0.75	0.78	0.76	0.045	0.006	0.062	0.074
		50	2.96	2.96	2.97	2.97	2.83	2.91	3.02	2.91	0.55	0.52	0.53	0.47	0.000	0.006	0.062	0.006
(0.1,0.2)	100	2.96	2.96	2.97	2.97	2.83	2.91	3.02	2.91	0.55	0.52	0.53	0.47	0.000	0.006	0.062	0.006	

observed among the Bayesian derived SDI among age-groups, which was not observed in case of classical SDI estimates.

**Discussion**

In the field of drug use epidemiology, the methodological challenges are inevitable [15]. Numerous studies have been performed in several countries like Bangladesh [16]; Pakistan [17]; Nepal [18]; Tehran [19]; Kenya [20]; Brazil [21] and India [22], but majority of them utilized purposive sampling for recruitment of street children. The lack of representativeness, has limited utility as these programs, for prevention and treatment intervention for substance using street children. There is a dearth of literature on this vulnerable population (street children) with very few studies from the developing countries. One of our previous study [14] was the first study from India that estimated the size of substance using street children in a representative manner by utilizing the Respondent Driven Sampling (RDS) survey. Hence, we attempted to move further for better characterization of the diversity. In that study we did mentioned that district level size estimation of the substance using street children could not be obtained because of the lack of count of district wise street children. Hence, under our present work we attempted the posterior characterization of the SDI by applying the

figures obtained in our previous study for getting fresh insights using simulation strategies.

It is emphasized that the diversity indices estimated for street children, corresponding to each regions of Delhi, under both classical and Bayesian paradigms were close to one another, but the Bayesian estimates are much more precise than that of classical estimates. In case of age-wise diversity of street children, Bayesian estimates variation in richness towards substance using street children and were higher than that of classical estimates. Hence, the results indicate that Bayesian paradigm was more precise (lesser variance) than classical estimates of diversity. The phenomena of substance use among the adolescent's age group are not uncommon. Based on our previous large epidemiological study on the assessment of pattern, profile, and correlates of substance use among children (both living at home or on the streets) in India, some deleterious figures of health consequences were observed on a sample of 4,024 children [23, 24] between the age of 5-18 years [Mean(SD): 15.6(2.1)]. It was found that approximately one-sixth of the children living at home and more than one-fourth of the street children were involved in the sexual behavior under the effects of the substance use. Additionally, one-sixth (16.9%) of the children at home and one fifth (20%) of the street children indulged in sexual behavior in exchange for either substances or

**Table 2** Classical and Bayesian estimates of Shannon diversity index (SDI) area-wise and age-class wise substance use among children in Delhi

Area	Age classes (in years)			Classical		Bayesian	
	7-10	11-14	15-18	$\hat{H}(p)$	$V(\hat{H}(p))$	$\hat{H}(p)^B$	$V(\hat{H}(p)^B)$
Central Delhi	2	2	3	1.51	1.2023	1.51	0.0106
East Delhi	10	12	8	1.56	0.9388	1.57	0.0005
New Delhi	0	14	35	0.86	0.5153	0.90	0.0111
North East Delhi	16	35	16	1.47	0.6123	1.49	0.0046
North West Delhi	7	19	13	1.47	0.8007	1.56	0.0005
South Delhi	11	14	8	1.54	0.8953	1.56	0.0007
South East Delhi	2	13	8	1.29	0.8988	1.28	0.0042
South West Delhi	7	36	35	1.34	0.5482	1.29	0.0047
West Delhi	14	20	16	1.56	0.7579	1.57	0.0001
$\hat{H}(p)$	2.74	2.92	2.84				
$V(\hat{H}(p))$	0.6847	0.4337	0.4659				
$\hat{H}(p)^B$	3.73	3.42	3.20				
$V(\hat{H}(p)^B)$	0.0644	0.0174	0.0271				

money. The physical and psychological problems related to substance use were reported by one-half of the children and that substantial proportion also reported legal problems on account of their substance use. The complications [experienced tolerance (55%-63%) or withdrawals (56%-67%)] was found to be higher among street children and out-of-school children. Hence, the detrimental consequence of substance use in the adolescent age group cannot be ignored as it may have more serious ramifications on other dimensions of life such as social, physical, psychological, legal, etc. The estimates of the SDI under the classical approach was closure to 2.9 for all the three age groups however, Bayesian estimates exhibited more diversity in the index in the lower age-group [7-10 years (3.74); 11-14 years (3.42) children] and comparatively less among 15-18 years children (3.20). This finding demonstrates that the policy-makers should take cognizance of substance use even in the lower age-group while devising a mechanism for age appropriate treatment strategies program so that various facets of physical, psychological, social, legal, and, etc. can be addressed effectively, in the middle and late childhood.

The term species evenness refers to a situation where the number of species in various possible classes are approximately closer to each other, in the environment of their habitat. Mathematically it is defined as a diversity index, a measure of biodiversity which quantifies how equal the community is numerically. Hence, findings under the present study provide a deeper insight to the fact that there is greater likelihood of finding street children in Delhi, who use substances and belong to lower as well as the middle age group. The major strength of

the present study is the theoretical extension of SDI, to the real world application. The novel mathematical framework, clearly demonstrates added benefit as it characterizes more specifically, the age profile of substance using street children in Delhi. The information gained under the Bayesian paradigm, demonstrates that in light of wider age range for substance using street children it is important to consider this factor for planning local level intervention targeted at ameliorating the conditions of street children in Delhi.

The limitation of the present study is that it took only two selected variable, viz., age and geographical location into consideration. Since we aimed at establishing the extension of theoretical aspects to real world data, we thought of using single variable. Moreover, age as a variable has important public health significance. From the perspective of the descriptive epidemiology the factor age holds its own importance. According to the 'Gateway hypothesis' [25], there is relationship between an early drug use and its progression in the later age. Hence, we considered the variable age. Age is also an important variable for devising age appropriate interventions. According to the World Health Organization (WHO), the age group 10-19 [26] constitutes adolescents. Our premise of using this age group was based on the consensus by the stakeholders involved in the implementation of the research study. The stakeholders were the personnel working in the Non-Governmental Organization (NGOs) with the street children and they opined that during the implementation phase the seed plan should allow intake of street children in these age groups. Further details in this regard can be found from our previous publication [14]. By this

age grouping we were able to have three levels with equal age spacing. We considered the broad age group given by the WHO with flexibility for the lower age range to be a part of the seed plan in implementing the RDS survey and ensuring equal age spacing. The geographical area is yet another important covariate affecting the phenomena of drug use. There are different settings for the street children such as railway tracks, garages, places of religious importance, restaurants, flyovers, traffic signals, and, etc. Any local level programs need to understand the geographical spread of the phenomena so that one can devise focused prevention strategies for decreasing the harm associated due to substance use. The districts of a particular province/state are one such natural local areas for understanding and undertaking focused attention. Moreover, it was for the first time that the posterior characterization of the SDI was attempted in the drug use epidemiology and thus we thought that these two variables would be good way to initiate the model characterization under Bayesian paradigm. The future studies should include more variables and further extend this concept for more better and comprehensive characterization of the profile of substance using street children or even other populations.

**Conclusion**

In the present article, we have investigated both classical and Bayesian in connection with the estimation of overall SDI in a population. The proposed Bayesian estimate of SDI is compared with their classical counterpart. In Bayesian statistics, the characterization of a parameter depends on the selection of the appropriate prior distribution for the analysis of data and the derived posterior is then used to draw the final conclusion. Here, we derived a particular prior based on the definition of SDI and the simulated result shows that Bayesian derived estimates of SDI are uniformly better than the classical estimate of SDI.

In the vulnerable population like that of the substance using street children, age is an important factor. Any intervention should be tailored around the age composition of street children. The study clearly demonstrates the utility of Bayesian paradigm of index based measure as the methodology for deeper understanding of age profile. The results demonstrates that such methodological innovation, unravel better characterization of age profile as it provides the evidence of propensity for finding substance using street children, who belongs to lower age group (7-10 years, highest measured entropy of 3.73) followed by middle age-group (11-14 years) and upper age group(15-18). In addition to that a systematic pattern was also captured through the Bayesian derived SDI among age-groups with lesser variability than classical SDI estimates, that can be utilized by policy makers. There is ample scope for further refinement in these estimate, by incorporating

more covariates that have a possible role in initiating substance use among street children in developing countries like India.

**Appendix**

**Proof of Theorem 2:** Let us assume that  $X_1, X_2, \dots, X_S$  are independent binomial random variables with  $X_i$  follows  $Binomial(n, p_i)$ , for all  $i = 1, 2, \dots, S$ . The Fisher's information of the probability of abundance of the  $i^{th}$  species in the community,  $p_i$ , using standard notation has been obtained as

$$I(p_i) = \sum_{y_i=0}^n \left( \frac{\delta}{\delta p_i} \log f(y_i|p_i) \right)^2 f(y_i|p_i) \tag{12}$$

and the inverse of the Fisher's information matrix of species classified probabilities vector, say  $\mathbf{p} = (p_1, \dots, p_S)$ , has given by

$$I^{-1}(\mathbf{p}) = I^{-1}(p_1, p_2 \dots, p_S) = \begin{bmatrix} \frac{p_1(1-p_1)}{n} & 0 & \dots & 0 \\ 0 & \frac{p_2(1-p_2)}{n} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{p_S(1-p_S)}{n} \end{bmatrix}$$

For the given linear function,  $H(\mathbf{p})$ , of Shannon index in Eq. (1), the gradient of  $\mathbf{p}$  has obtained as

$$D_H^T(\mathbf{p}) = \left[ \frac{\partial H(\mathbf{p})}{\partial p_1} \quad \frac{\partial H(\mathbf{p})}{\partial p_2} \quad \dots \quad \frac{\partial H(\mathbf{p})}{\partial p_a} \quad \dots \quad \frac{\partial H(\mathbf{p})}{\partial p_c} \right] = [\gamma_1 \ \gamma_2 \ \dots \ \gamma_a \ \dots \ \gamma_s]$$

where  $\gamma_i = -[\log(p_i) + 1]$ .

Let  $\lambda_i = (I^{-1}(\mathbf{p}))_{ii} = (i^{th} \text{ diagonal element of } I^{-1}(\mathbf{p})) = \frac{p_i(1-p_i)}{n}$ , then

$$D_H^T(\mathbf{p})I^{-1}(\mathbf{p}) = [\gamma_1 \ \gamma_2 \ \dots \ \gamma_s] \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_s \end{bmatrix} = [\gamma_1\lambda_1 \ \gamma_2\lambda_2 \ \dots \ \gamma_s\lambda_s] \tag{13}$$

$$D_H^T(\mathbf{p})I^{-1}(\mathbf{p})D_H(\mathbf{p}) = \sum_{i=1}^s \gamma_s^2 \lambda_s \tag{14}$$

Let us consider,

$$\phi^T(\mathbf{p}) = \frac{D_H^T(\mathbf{p})I^{-1}(\mathbf{p})}{\sqrt{D_H^T(\mathbf{p})I^{-1}(\mathbf{p})D_H(\mathbf{p})}} \tag{15}$$

$$= \left[ \frac{\gamma_1\lambda_1}{\sqrt{\sum_{i=1}^s \gamma_s^2 \lambda_s}} \quad \frac{\gamma_2\lambda_2}{\sqrt{\sum_{i=1}^s \gamma_s^2 \lambda_s}} \quad \dots \quad \frac{\gamma_s\lambda_s}{\sqrt{\sum_{i=1}^s \gamma_s^2 \lambda_s}} \right] = [\phi_1(\mathbf{p}) \ \phi_2(\mathbf{p}) \ \dots \ \phi_S(\mathbf{p})], \tag{16}$$

where,  $\phi_i(\mathbf{p}) = \frac{\gamma_s \lambda_s}{\sqrt{\sum_{i=1}^s \gamma_s^2 \lambda_s}} = -[\log(p_i) + 1] \frac{p_i(1-p_i)}{n} \left( \sqrt{\sum_{i=1}^s [\log(p_i) + 1]^2 \frac{p_i(1-p_i)}{n}} \right)^{-1}$ .

In context of deriving a prior distribution of a parameter, [27] has suggested the criteria that must be satisfied to establish the posterior distribution for a parametric function and is given by

$$\sum_{i=1}^S \frac{\partial}{\partial p_i} \phi_i(\mathbf{p}) q(\mathbf{p}) = 0$$

Let

$$q(\mathbf{p}) = \left( \sum_{i=1}^S [\log(p_i) + 1]^2 \frac{p_i(1-p_i)}{n} \right)^{\frac{1}{2}} \prod_{i=1}^S [p_i(1-p_i)(1 + \log(p_i))]^{-1}$$

then

$$\phi_1(\mathbf{p}) q(\mathbf{p}) = \frac{-1}{n} \prod_{i \neq 1}^S [p_i(1-p_i)(1 + \log(p_i))]^{-1}$$

$$\Rightarrow \frac{\partial}{\partial p_1} \phi_1(\mathbf{p}) q(\mathbf{p}) = 0 \tag{17}$$

and

$$\phi_j(\mathbf{p}) q(\mathbf{p}) = \frac{-1}{n} \prod_{i \neq j=1}^S [p_i(1-p_i)(1 + \log(p_i))]^{-1}$$

$$\Rightarrow \frac{\partial}{\partial p_j} \phi_j(\mathbf{p}) q(\mathbf{p}) = 0. \tag{18}$$

From the above Eqs. (17) and (18) we have

$$\sum_{i=1}^S \frac{\partial}{\partial p_i} \phi_j(\mathbf{p}) q(\mathbf{p}) = 0,$$

which satisfied the condition required to be a prior distribution,  $q(\mathbf{p})$ , of a parameter. Therefore,

$$q(\mathbf{p}) \propto \left\{ \sum_{i=1}^S (1 + \log(p_i))^2 p_i(1-p_i) \right\}^{\frac{1}{2}} \prod_{i=1}^S [p_i(1-p_i)(1 + \log(p_i))]^{-1}; 0 < p_i < 1$$

and hence we get the required proof.

### Metropolis-Hastings algorithm

The basis for opting Metropolis-Hastings (MH) algorithm is to simulate samples from a probability distribution by using the full joint density function and (independent) proposals distributions for each variable of interest. The

steps followed under this algorithm consist of three components namely, propose a proposal distribution to generate sample, computation of the acceptance probability, acceptance criteria of a sample generated through the proposal distribution that are given below:

**Initialize** Initialize the sample value, say  $\pi$ , for each random variable from  $g(\cdot) \sim \text{Uniform}(0, 1)$ .

$$\pi^{(0)} \uparrow g(\pi)$$

**for** iteration  $i = 1, 2, \dots$  do

{  
**Propose:** Generate a proposal (or a candidate) sample  $\pi^{cand}$  from the proposal distribution  $q(\pi^{(i)} | \pi^{(i-1)})$

$$\pi^{cand} \uparrow h(\pi^{(i)} | \pi^{(i-1)})$$

**Acceptance Probability:** Compute the acceptance probability via the acceptance function  $\alpha(\pi^{cand} | \pi^{(i-1)})$  based on the proposal distribution and the full joint density  $g(\cdot)$

$$\alpha(\pi^{cand} | \pi^{(i-1)}) = \text{Min} \left\{ 1, \frac{h(\pi^{(i)} | \pi^{cand}) g(\pi^{cand})}{h(\pi^{cand} | \pi^{(i-1)}) g(\pi^{(i-1)})} \right\}$$

Generate  $u \sim \text{Uniform}(u; 0, 1)$

**if**  $u < \alpha$  **then**

**Accept the proposal:**  $\pi^{(i)} \leftarrow \pi^{cand}$

**else**

**Reject the proposal:**  $\pi^{(i)} \leftarrow \pi^{(i-1)}$

**end if**

**Accept the candidate sample with probability  $\alpha$ , the acceptance probability, or reject it with probability  $1 - \alpha$**

**end for**

}

### Abbreviations

SDI: Shannon's diversity index; RDS: Response driven sampling; MCMC: Markov Chain Monte Carlo

### Acknowledgements

The authors deeply appreciate the active participation of the officials of NGOs (Butterflies, Childhood Enhancement through Training and Action, Don Bosco Ashalayam, Prayas, Saalam Baalak Trust and Society for Promotion of Youth and Masses) and is duly acknowledged. At last but not least, we place on record the kind cooperation of street children without whose participation this study would not have been completed. The authors express their sincere thanks and gratitude to the editorial board and both of the anonymous reviewers for their valuable suggestions, which enabled in substantial improvement of the manuscript.

### Authors' contributions

VV, Primarily involved in the development of the statistical model and wrote the first draft of the manuscript. AKM, Data acquisition, definition of the intellectual content, organizing of the results and editing of the manuscript. AD, Principal Investigator of the research project, substantial editing and interpretation of the results. DCN, Involved in the development of the model and editing of the manuscript. All authors have read and approved the manuscript.

### Funding

The present research work was financially supported by the Department of Women and Child Development, Government of National Capital Territory of



Delhi. ID: IEC/NP-253/03.07.2014, RP-15/2014. The funding agency was not involved towards the scientific or intellectual content of the project under which this study was conducted.

#### Availability of data and materials

The data and the material are under the Government Jurisdiction and if sought necessary to be made available, need the permission for the same through the competent authority. In this reference the corresponding author can be contacted.

#### Ethics approval and consent to participate

The Institutional Ethical Committee of the All India Institute of Medical Sciences (AIIMS), New Delhi, granted the approval and consent for the participation vide the letter number IEC/NP-253/03.07.2014, RP15/2014. Children whose parents' consented verbally were enrolled in the study, which was approved by the ethics committee.

#### Consent for publication

The principal investigator of the study is also the co-author and hence we do have the consent for publications. All the research study carried out at the All India Institute of Medical Sciences (AIIMS), New Delhi, can be published by the investigators and that there are no restrictions.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Neurology, All India Institute of Medical Sciences (AIIMS), New Delhi, India. <sup>2</sup>National Drug Dependence Treatment Centre (NDDTC), All India Institute of Medical Sciences (AIIMS), New Delhi, India. <sup>3</sup>Assam University, Silchar, Assam, India.

Received: 12 August 2020 Accepted: 23 November 2020

Published online: 01 December 2020

#### References

- Pielou EC. Ecological Diversity. *Limnol Oceanogr.* 1977;22(1):174–174.
- Basharin GP. On a statistical estimate for the entropy of a sequence of independent random variables. *Theory Probab Appl.* 1959;4(3):333–6.
- Zahl S. Jackknifing an index of diversity. *Ecology.* 1977;58(4):907–13.
- Chao A, Shen T-J. Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environ Ecol Stat.* 2003;10(4):429–43.
- Chao A, Lee S-M. Estimating the number of classes via sample coverage. *J Am Stat Assoc.* 1992;87(417):210–7.
- Magurran AE. *Ecoogical Diversity and its Measurement.* Princeton: Princeton University Press; 1988.
- Türkmen G, Kazanci N. Applications of various biodiversity indices to benthic macroinvertebrate assemblages in streams of a national park in Turkey. *Rev Hydrobiol.* 2010;3(2):111–25.
- Lamb EG, Bayne E, Holloway G, Schieck J, Boutin S, Herbers J, Haughland DL. Indices for monitoring biodiversity change: Are some more effective than others? *Ecol Indic.* 2009;9(3):432–44.
- Oehlert GW. A note on the delta method. *Am Stat.* 1992;46(1):27–29.
- Reid N, Mukerjee R, Fraser DAS. Some aspects of matching priors. In *Mathematical Statistics and Applications: Festschrift for Constance van Eeden.* Inst Math Stat Lect Notes-Monogr Ser. 2003;42:31–43.
- Scricciolo C. Probability matching priors: A review. *J Ital Stat Soc.* 1999;8(1):83.
- Spellerberg IF, Fedor PJ. A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the Shannon–wiener index. *Glob Ecol Biogeogr.* 2003;12(3):177–9.
- Shannon CE. A mathematical theory of communication. *Bell Syst Tech J.* 1948;27(3):379–423.
- Dhawan A, Mishra AK, Ambekar A, Chatterjee B, Agrawal A, Bhargava R. Estimating the size of substance using street children in delhi using respondent-driven sampling (rds). *Asian J Psychiatry.* 2020;48:101890.
- Vameghi M, Roshanfekar P, Ali D, Noroozi M, Madani S, McFarland W, Mirzazadeh A. Population size estimates of street children in Iran: synthesis of multiple methods. *J Urban Health.* 2019;96(4):549–57.
- Hossain MI, Ahmed Z. Social bonding and drug addiction: a study on the street children in Dhaka city. *ASA Univ Rev.* 2015;9(2):193–214.
- Sherman SS, Plitt S, ul Hassan S, Cheng Y, Zafar ST. Drug use, street survival, and risk behaviors among street children in Lahore, Pakistan. *J Urban Health.* 2005;82(4):113–24.
- Kakchapati S, Shrestha B, Li DY, Rajbhandari R, Poudel T. Drug use, injecting behaviors, and survival sex among street children and youths in Kathmandu Valley, Nepal. *Int J STD AIDS.* 2018;29(6):588–97.
- Ahmadkhaniha HR, Mohammadian M, Naserbakht M, Ghazaiepour F, et al. Street children in tehran and risk factors for substance abuse. *Med J Islam Repub Iran (MJIRI).* 2010;24(2):83–87.
- Bah YM. Drug abuse among street children. *J Clin Res HIV AIDS Prev.* 2018;3(3):12.
- Moura Y. G. d., Sanchez ZM, Opaleye ES, Neiva-Silva L, Koller SH, Noto AR. Drug use among street children and adolescents: what helps? *Cad Saude Publica.* 2012;28:1371–80.
- Sharma N, Joshi S. Preventing-substance abuse among street children in India: a literature review. *Health Sci J.* 2013;7(2):137.
- Dhawan A, Pattanayak R, Chopra A. Assessment of pattern, profile and correlates of substance use among children in India. National Commission for Protection of Child Rights (NCPCR). New Delhi. 2013.
- Dhawan A, Pattanayak RD, Chopra A, Tikoo V, Kumar R, et al. Pattern and profile of children using substances in India: Insights and recommendations. *Natl Med J India.* 2017;30(4):224.
- Nkansah-Amankra S, Minelli M. "gateway hypothesis" and early drug use: Additional findings from tracking a population-based sample of adolescents to adulthood. *Prev Med Rep.* 2016;4:134–41.
- Sacks D, Society CP, Committee AH. Age limits and adolescents. *Paediatr Child Health.* 2003;8(9):577.
- Datta GS, Ghosh JK. On priors providing frequentist validity for Bayesian inference. *Biometrika.* 1995;82(1):37–45.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

