Exploring the *h*-type measure and its theoretical model in the context of e-commerce

Bingjia Shao[1], Feng Xiao[1] and Star X. Zhao[2,*]

[1]School of Economics and Business Administration, Chongqing University, Chongqing 400030, China
[2]Department of Information Management, Faculty of Economics and Management, East China Normal University, Shanghai 200241, China

The aim of this study is to find whether the *h*-index and Schubert–Glänzel model which was originally used in bibliometrics can work in the broader social and information systems, such as e-commerce. This study introduces the popularity index (h_P) to profile on-line sellers and products and tests the accuracy of the Schubert–Glänzel model of the h_P index using the Levenberg–Marquardt algorithm. The h_P index was tested on Taobao, a leading Chinese e-commerce marketplace. About 300 products on Taobao were randomly selected from 10 categories of the most popular on-line products as samples for this study. It was found that h_P is consistent with the Schubert–Glänzel model the *h*-index. This shows that the existing theoretical bibliometric model which was originally used in the scientific literature, could be helpful in the broader social or information systems.

Keywords: Bibliometric measures, e-commerce, *h*-index, popularity index.

THE *h*-index, originally proposed by Hirsch[1,2], is regarded as one of the most important measures to quantify the scientific output of researchers in scientometrics and informetrics[3–6]. It is considered the state-of-the-art of scientific research evaluation, due to its unique evaluation perspective, easy computing process, robust results, strong recognition and forecasting ability of scientists[7]. Application of the *h*-index has achieved great success in bibliometrics, and scholars are exploring its applications in different aspects, such as information network science[8,9]. In the past decade, e-commerce using information technology has substantially impacted the business world and greatly changed our daily life[10–12]. Every year trillions of transactions are being made on e-commerce markets. Data from these transactions can provide useful information for companies to make effective decisions and also help consumers make better product choices[13]. Exploring the feasibility of *h*-type measure in e-commerce is an interesting area of research.

The *h*-index can be defined as follows: a scientist has an index *h*, if *h* of his or her *N* papers have at least *h* citations each and the other $(N - h)$ papers have $\leq h$ citations each[1]. In recent years, the *h*-index has been thoroughly improved both in bibliometrics and in other fields. For instance, Banks (2006) applied the *h*-index to study topics and compounds, and Bar-Ilan proposed an *h*-type method to evaluate the production in different scientific topics[14]. Glänzel put forward two new bibliometric applications of *h*-related indicators[15]. Jacso discussed the related data sources and databases for computing *h*-index[16]. Schuber and Guillaume extended the *h*-index to evaluate the co-author partnership ability[17,18]. Schubert *et al*. and Zhao *et al*. applied the algorithm of *h*-index as a basic measure for network science, which has led to a new set of indicators characterizing nodes in a network[8,9,19]. In addition, there are some revised *h*-indices; their main purpose has been to complement the *h*-index and extend the same, considering other important variables like time. For example, the *g*-index was proposed as a modification of the *h*-index to give more weightage to highly cited papers[20]. Burrell revised the *h*-index by dividing the number of years of research activities to compare the contribution of researchers[21]. From a theoretical perspective, one of the most interesting and accurate models is the Schubert–Glänzel model for citation analysis[22–24]. It can also be used for information network analysis[17,25].

As both the *h*-type measure and Schubert–Glänzel model show significant and general features in information analysis, this communication discusses a link between e-commerce sales information and the *h*-type measure. It aims to explore the possibility of whether bibliometric measures and models can successfully work in the broader social systems like business.

First, we apply concept of *h*-index to the e-commerce system. Then the theories and methods associated with the popularity index and the Schubert–Glänzel model in e-commerce are discussed. Next we describe an empirical study of the largest Chinese e-commerce company followed by conclusions.

The popularity of on-line shopping has grown immensely with the rapid development of e-commerce[26–28]. A large number of companies and retailers have discovered the business opportunities of e-commerce and are actively participating in on-line platforms to promote their products. Companies and retailers try to sell products that are popular with customers to satisfy their demand, which increases the profits.

Evaluation of popularity of product lies at the heart of the e-commerce system, because it not only directly reflects customer demand, but is also an indicator of market orientation for companies[29,30]. In China, customers use the sales volume to evaluate popularity of a product[27]. However, they do not pay attention to the influence of sellers on product popularity, since the sale of a product is often easily controlled by limited number of sellers[31]. At the same time, a great waste of network resource may be generated by excess sellers. In the case of an oligopoly, several major sellers can lower prices of products and force out other minor sellers. In the case of excessive

*For correspondence. (e-mail: xzhao@infor.ecnu.edu.cn)

sellers, the sales volume of each seller is low. The on-line products must be not popular for sellers when only these products' total volume is high. But only if appropriate numbers of sellers participate in selling, the product will be sold in a fair, free and extensive market, which would attract more sellers. Therefore, the number of sellers is critical to accurately reflect the popularity of a product in a competitive market.

Relevant to the aforementioned characteristics of e-commerce sales, for evaluating the popularity of a product, we define a measure similar to the Hirsch-index, called the popularity index (h_P).

The popularity index of a product shows that this product with an index of h_P has sellers who sell at least P products each, while other sellers have no more than P sales.

From this definition, it can be seen that h_P is largely determined by the total number of sellers and the volume of the product each seller can sell. Only if the number of sellers reaches a certain value, and product sales volume is also large, can the value of popularity index increase. If a small number of sellers is involved in selling a product, but they sell a high volume, it does not mean that this product is popular because some big companies can control the market and manipulate or even fake the sales data. In contrast, if many sellers participate in the selling, but each of them sells only a few products, the popularity index will be very low.

The number of sellers indicates the number of companies and retailers who are optimistic about this product, and also whether the market environment of this product is under the free competitive condition. The sales volume of each seller shows the popularity of the product. The h_P index is a combination of the total number of sellers and the sales number for each product sold by them. The h_P index can help the sellers to scientifically choose new products and optimize the products mix. Moreover, companies can use this index to understand customer demand so they can sell competitive products to attract customers and gain greater market share. For customers, this index can be a useful indicator to help them choose popular products among so many brands.

The Schubert–Glänzel model of h-index can be expressed as follows[22]

$$h = cn^{1/(\alpha+1)} x^{\alpha/(\alpha+1)}, \quad (1)$$

where c is a positive real value, α the Lotka coefficient (classical value = 2), n the sample size, and x is the density (mean citation rate per paper).

As a Hirsch-type index, h_p can be expected to be consistent with the Schubert–Glänzel model of the h-index. In the present work, we suppose that this bibliometric model can successfully be applied to an e-commerce marketplace.

Assuming that the presumptions of the Schubert–Glänzel model also work in the popularity index example, the relationship among h_p, G (the total sellers of a product) and AS (the mean sales volume per month) is given as follows

$$h_p = cG^{1/(\alpha+1)} AS^{\alpha/(\alpha+1)}, \quad (2)$$

where c is a positive real value and α is a coefficient.

Next we provide a real case to test the h_P index and Schubert–Glänzel model in the context of an e-commerce marketplace.

The Chinese e-commerce company, Taobao, was chosen for this empirical study. Sales data were collected from the company website (taobao.com) for May–July 2014. According to the Chinese e-commerce Research Center, the turnover of the Chinese on-line e-commerce market amounted to US$ 1612 trillion in 2013. On-line retail comprises a major portion of the total Chinese e-commerce market. Being the largest C2C and B2C on-line company, Taobao shares 80% of the on-line retail market[32–34]. It is the leading electronic commerce market in China. Its parent company, Alibaba, has become the world's most valuable e-commerce company (US$ 174.9 billion)[35,36]. Taobao dominates the C2C market with 94.5% of C2C sellers and it has more than 26 million professional sellers[26]. Thus, the company can largely represent the growing on-line retail market in China.

Data were selected based on information about 10 categories of the most popular networking products published by 'the Tmall Operation Service White Paper'[37,38]. Table 1 shows the 300 types of products randomly selected from the 10 categories as sample data in the present study. Table 2 shows the characteristics of sample data.

From each of the most popular networking products category, 30 products were randomly chosen. For example, in electronics, Samsung N8000 and Flyco 330 were chosen. In the case of clothing, the Roem Rcow 32313c and Nike 554886 were selected.

For every product, all the sellers were ranked in descending order of sales volume first. Then the sort number is regarded as the value of the empirical popularity index. When the sort number is equal to or larger than the corresponding number of sales volume.

The total number of sellers and the mean sales volumes per month for the product were calculated. Then, according to the Schubert–Glänzel model, the theoretical value of the popularity index h_p was computed.

The h_p is determined by two factors; the number of sellers and the sales volume of each seller. From the analysis of the sample data, it was found that the law of diminishing marginal utility exists between the total sales volume and h_p index. When h_P index is very low, it will increase rapidly when total sales increases. But when h_p reaches a certain value, it requires high sales volume to

Table 1. Examples of the 300 kinds of products

Type of product	Product name	Number of sellers	h_p	Mean sales volume per seller
Electronics	Samsung N8000	104	20	14
Electronics	Flyco FH216	67	20	15
Electronics	Flyco 330	247	55	71
Clothing	Roem Rcow 32313c	38	6	3
Clothing	Nike 554886	359	39	22
Clothing	Staccato 9FH43	201	41	48
Cosmetics	Maxfactor lasting performance	129	16	13
Cosmetics	Maybelline Volum Express	187	25	19
Cosmetics	DHC mild soap	204	39	31
Baby products	Pigeon nursing bottle, 240 ml	333	69	82
Baby products	Pigeon baby liquid laundry, 1000 ml	279	60	51
Baby products	HiPP first infant milk stage I	265	27	12
Appliances	Meizu EP10 earphone	117	30	40
Appliances	Sennheiser CX300II	118	17	19
Appliances	TP-LINK TL-WR845N	261	60	204

Table 2. Characteristics of sample data

Product type	Product characteristics	Degree of product standardization
Women's clothing	Female	Moderate
Beauty products	Female	Low
Underwear and shoes	Female and male	High
Men's clothing	Male	Moderate
Electronics	Male	Low
Mother and baby products	Family	Moderate
Appliances	Family	Low
Home decoration	General	Moderate
Outdoor products	General	High
Regional food	General	Moderate

increase further. In other words, as the total sales volume increases, the positive effect of the same volume on the h_p index gradually decreases and an increase in h_p becomes difficult. Therefore, in order to increase the value h_p index, depending only on increasing the number of sellers but not on the total product sales proves ineffective. At the same time, high sales volume from a limited number of sellers may have a positive effect on h_P at first, but such an effect will be undermined later. h_p index is balanced and comprehensive to evaluate product popularity.

The h_p index value of regional food is much higher than other products, regardless of the total sales volume or the number of sellers. This is related to product characteristics and e-commerce promotion. E-commerce provides an effective platform for sellers. Regional foods can be offered fresh and at the lowest price to customers. In contrast, the h_p index of books was the lowest in the present study because few booksellers provide free shipping to customers. This increases the price of the books. Compared with Amazon and Dangdang, the leading on-line bookseller in China, Taobao has little advantage in selling books. So it is difficult to obtain a high h_P index for books.

In order to test the accuracy of the h_P model, values of the empirical and theoretical popularity index were calculated for the 300 products. Figures 1 and 2 show the correlation between estimated and actual values. The coefficient of determination $r^2 \in [0, 1]$ was used to measure the accuracy ($r^2 \to 1$) of the Schubert–Glänzel model's model with respect to empirical data.

Equation (2) demonstrates that the value of α can strongly influence the relationship between estimated and actual values.

As $\alpha = 2$ is defined as the classic value in informatics, the Schubert–Glänzel model of h_p index is as follows

$$h_p = cG^{1/3}AS^{2/3}, \tag{3}$$

where $c = 0.528$.

From Figure 1, it can be seen that although there is a relationship between the estimated and actual values, the coefficient of determination (Adj. $r^2 = 0.7400$) does not illustrate a strong relationship between the two variables.

In order to increase the accuracy of the Schubert–Glänzel model, the Levenberg–Marquardt algorithm (LMA) was used to optimize the value of the parameters. As a result, when $\alpha = 0.59272$, the coefficient of determination is as high as 0.914 (Adj. $r^2 = 0.914$). This suggests that the Schubert–Glänzel model is more accurate for this value of α when $\alpha = 2$ (Figure 2).

$$h_{p1} = cG^{1/(1.59272)}AS^{0.59272/(1.59272)}, \tag{4}$$

when $\alpha = 0.59272$, most sample values which is estimated by the Schubert–Glänzel model is fitting the Empirical value greatly (Figure 3).

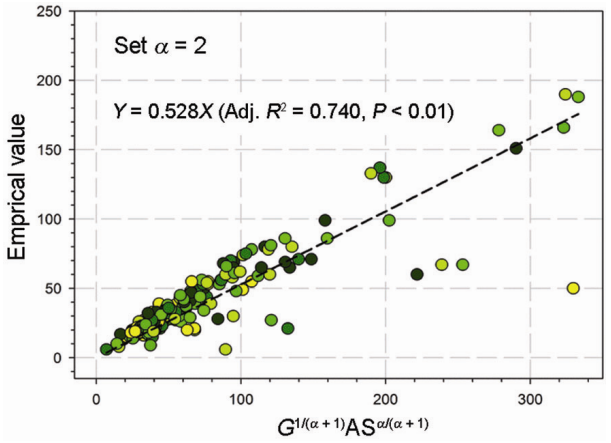

Figure 1. Theoretical (h_p) and the empirical (h_p) values of popularity of 300 products on taobao.com, when $\alpha = 2$.

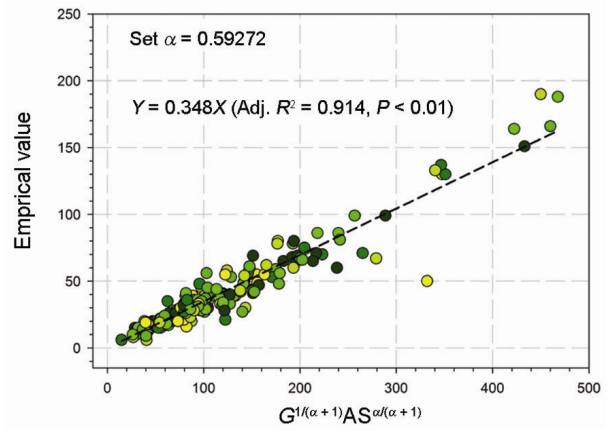

Figure 2. The theoretical and empirical values of popularity of 300 products, when $\alpha = 0.59272$.

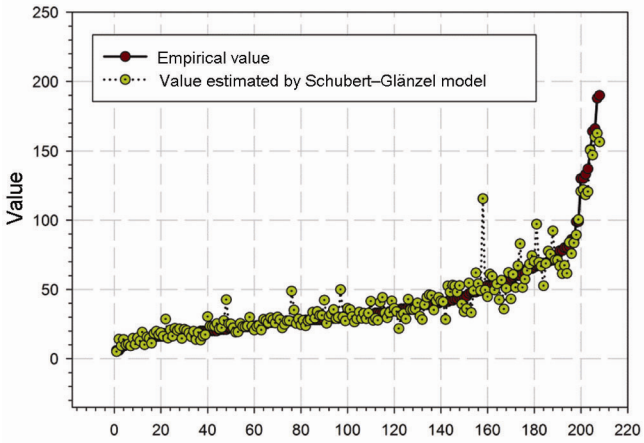

Figure 3. Comparison of theoretical (h_{p1}) and empirical (h_{p1}) values of popularity of products, when $\alpha = 0.59272$.

This study has explored the usage of popularity index (h_p) and validity of the Schubert–Glänzel model in the context of e-commerce. Results show that h_p index can be an effective measure of product popularity of e-commerce system. The Schubert–Glänzel model approximately fits the empirical data collected from the e-commerce company website. The estimated value of the parameters of this model undergo changes, and the classical values in bibliometrics of the parameters may not be optimal values in the e-commerce system. These findings reveal that: (a) specific bibliometric measures, such as the *h*-index, could provide interesting perspectives with regard to business systems, especially in e-commerce and (b) theoretical models from information science and bibliometrics can also be used in e-commerce.

The h_p-index can help companies and individual retailers to choose products and optimize product mixes. Sellers in the e-commerce markets can use h_p-index to modify their product promotion strategies. The number of sellers helps customers to understand the market environment of products. The sales volume helps customers to know about the popularity of the products. It is difficult for a limited number of large sellers to manipulate prices and rankings for products with high h_p value. Furthermore, because viewing other customers' evaluation of a product is becoming an important factor and reference for purchasing products, the h_p index should not only include the total number of sellers of a product and the mean sales volume, but also take the customers' feedback into consideration.

However, there are some limitations in this study. The original *h*-index works in bibliometrics because the scale on the *y*-axis (citations per paper) and *x*-axis (papers) is the same. The h_p-index also inherits shortcomings of the *h*-index, such as ignoring the concrete quantities in *h*-core and lack of distinction in some situations[39]. So the modified *h*-index, for example the *g*-index or *R*-index, may provide more detailed information. We recommend further studies using the Schubert–Glänzel model as starting point to develop new connections and models between information science and e-commerce.

1. Hirsch, J. E., An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 16569–16572.
2. Hirsch, J. E., Does the *h*-index have predictive power? *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 19193–19198.
3. Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E. and Herrera, F., *h*-Index: a review focused in its variants, computation and standardization for different scientific fields. *J. Informetr.*, 2009, **3**, 273–289.
4. Jacso, P., Using Google Scholar for journal impact factors and the *h*-index in nationwide publishing assessments in academia – siren songs and air-raid sirens. *Online Inf. Rev.*, 2012, **36**, 462–478.
5. Bornmann, L. and Marx, W., HistCite analysis of papers constituting the *h*-index research front. *J. Informetr.*, 2012, **6**, 285–288.
6. Tsai, C. F., Hu, Y. H. and Ke, S. W. G., A Borda count approach to combine subjective and objective based MIS journal rankings. *Online Inf. Rev.*, 2014, **38**, 469–483.
7. Yin, C. Y., Do impact factor, *h*-index and eigenfactor of chemical engineering journals correlate well with each other and indicate the journals' influence and prestige? *Curr. Articles*, 2011, **100**, 648–653.

8. Zhao, S. X., Rousseau, R. and Ye, F. Y., *h*-Degree as basic measure in weighted networks. *J. Informetr.*, 2011, **5**, 668–677.
9. Rousseau, R. and Zhao, S. X., A general conceptual framework for characterizing the ego in a network. *J. Informetr.*, 2015, **9**, 145–149.
10. Zhao, S. X. and Ye, F. Y., Exploring the directed *h*-degree in directed weighted networks. *J. Informetr.*, 2012, **6**, 619–630.
11. Moutusy, M. and Mayukh, D., Consumer decision-making across modern and traditional channels: e-commerce, m-commerce, in-store. *Decision Support Syst.*, 2014, **61**, 34–46.
12. Bakos, Y., Lucasm, H. C., Oh, W., Simon, G., Viswanathan, S. and Weber, B. W., The impact of e-commerce on competition in the retail brokerage industry. *Inf. Syst. Res.*, 2005, **16**, 352–371.
13. Yu, Z., Jing, B. and Wei, X. Z., Trust fraud: a crucial challenge for China's e-commerce market. *Electron. Commerce Res. Appl.*, 2013, **12**, 299–308.
14. Gu, B., Konana, P. and Chen, H. W. M., Identifying consumer consideration set at the purchase time from aggregate purchase data in online retailing. *Decision Support Syst.*, 2012, **53**, 625–633.
15. Bar-Ilan, J., Informetrics at the beginning of the 21st century – a review. *J. Informetr.*, 2008, **2**, 1–52.
16. Glänzel, W., On some new bibliometric applications of statistics related to the *h*-index. *Scientometrics*, 2008, **77**, 187–196.
17. Jacso, P., The *h*-index, *h*-core citation rate and the bibliometric profile of the web of science database in three configurations. *Online Inf. Rev.*, 2011, **35**, 821–833.
18. Schubert, A., A Hirsch-type index of co-author partnership ability. *Scientometrics*, 2012, **91**, 303–308.
19. Guillaume, C., Experimenting with the partnership ability φ-index on a million computer scientists. *Scientometrics*, 2013, **96**, 1–9.
20. Schubert, A., Korn, A. and Telcs, A., Hirsch-type indices for characetering networks. *Scientometrics*, 2009, **78**, 375–382.
21. Egghe, L., Theory and practice of the *g*-index. *Scientometrics*, 2006, **69**, 131–152.
22. Burrell, Q. L., Hirsch's *h*-index: a stochastic model. *J. Inf.*, 2007, **l**, 16–25.
23. Schubert, A. and Glänzel, W., A systematic analysis of Hirsch-type indices for journals. *J. Intermetrics*, 2007, **1**, 179–184.
24. Ye, F. Y., An investigation on mathematical models of the *h*-index. *Scientometrics*, 2009, **81**, 493–498.
25. Ye, F. Y., A unification of three models for the *h*-index. *J. Am. Soc. Inf. Sci. Technol.*, 2011, **62**, 205–207.
26. Zhao, S. X., Zhang, P. L., Li, J., Tan, A. M. and Ye, F. Y., Abstracting the core subnet of weighted networks based on link strengths. *J. Assoc. Inf. Sci. Technol.*, 2014, **65**, 984–994.
27. McKinsey Global Institute, China's e-tail revolution: online shopping as a catalyst for growth. McKinsey & Company, USA, 2013; http://www.mckinsey.com/insights/asia-pacific/china_e-tailing.
28. Chen, Y. T. and Chou, T. Y., Exploring the continuance intentions of consumers for B2C online shopping: perspectives of fairness and trust. *Online Inf. Rev.*, 2012, **36**, 104–125.
29. Lee, J., Park, D. H. and Han, I., The different effects of online consumer reviews on consumers' purchase intentions depending on trust in online shopping malls: an advertising perspective. *Internet Res.*, 2011, **21**, 187–206.
30. Dellarocas, C., Gao, G. D. and Narayan, R., Are consumers more likely to contribute online reviews for hit or Niche product? *J. Manage. Inf. Syst.*, 2010, **27**, 127–157.
31. Ahn, H. J., Utilizing popularity characteristics for product recommendation. *Int. J. Electron. Commerce,* 2006, **11**, 59–80.
32. Parente, S. L. and Prescott, E. C., Monopoly rights: a barrier to riches. *Am. Econ. Rev.*, 1999, **89**, 1216–1233.
33. Iresearch, Chinese e-commerce transactions to 8.1 trillion in 2012, raised the proportion of online shopping, 2013; http://ec.iresearch.cn/shopping/20130128/192198.shtml (in Chinese).
34. China Internet Watch, How much it costs to set up online shop on China B2C platforms. 2013; http://www.chinainternetwatch.com/
35. Chinese E-commerce Research Center, Monitoring Data of China's e-commerce Market in the First Half of 2014. 2014.
36. Wang, B. N., Li, J. Q. and Zhu, X. N., China's E-commerce. Ministry of Commerce of the People's Republic of China, 2013 (in Chinese).
37. Chinese E-commerce Research Center, Alibaba group US listing prospectus, 2014; http://www.100ec.cn/detail--6170970.html
38. Alibaba Group Research Center, Tmall Service Provider White Book, Tmall and Taobao, 2012.
39. Ye, F. Y. and Rousseau, R., Probing the *h*-core: an investigation of the tail–core ratio for rank distributions. *Scientometrics*, 2010, **84**, 431–439.

ACKNOWLEDGEMENTS. We acknowledge financial support from the Key Project of National Social Science Fund of China (No. 14AGL023) and National Natural Science Foundation of China (No. 71503083).

Received 25 June 2015; revised accepted 14 March 2016

doi: 10.18520/cs/v110/i12/2276-2280

Nutritional status and infant mortality rate in Saiha district, Mizoram, India

Vishwambhar Prasad Sati* and Lalrinpuia Vangchhia

Department of Geography and Resource Management, Mizoram University, Aizawl 796 004, India

This communication examines nutritional status and its impact on infant mortality rate (IMR) in Saiha district, Mizoram, India. 1650 mothers from 17 villages were surveyed using random sampling method. The district has very high IMR (219.6), significantly higher than Mizoram (35). Meanwhile, per day per capita calorie intake is 1703, which is less than the recommended dietary intake (2400 kcal). All food items which people consume daily were collected and nutritional status was assessed. Our result shows that high IMR in the district is due to food insecurity and malnutrition.

Keywords: Food insecurity, infant mortality rate, nutrition status, pregnant mothers.

NUTRITIONAL status has a significant impact on infant mortality rate (IMR) and health of the people, particularly children. It is closely linked with food adequacy and its proper distribution, and is an outcome of complex and interrelated sets of factors[1]. Its deficiency in the body causes several diseases. On the other hand, inadequacy in

*For correspondence. (e-mail: sati.vp@gmail.com)