

CHAPTER 6

The Use of Species Diversity Indices and Accumulation Curves to Analyse Ethnobotanical Data

Application of diversity indices to appraise plant availability in the traditional medicinal markets of Johannesburg, South Africa

VIVIENNE L. WILLIAMS*, EDWARD T.F. WITKOWSKI and Kevin Balkwill

*School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, PO Wits, 2050, South Africa; *Author for correspondence (e-mail: vivwill@planetac.co.za; fax: + 27-11-403-1429)*

Received 4 June 2003; accepted in revised form 19 April 2004

Key words: Abundance, Diversity, Evenness, Quantitative ethnobotany, Species accumulation curve, Species richness, Traditional medicinal trade

Abstract. The lack of scientific rigour in analysing ethnobotanical surveys has prompted researchers to investigate ways of quantitatively describing their data, including the use of ecological diversity indices. There are numerous indices and measures available to describe sample diversity. Twenty-two measures of species richness, diversity and evenness were reviewed using six sets of ethnomedicinal data derived from 50 formal *muti* shop traders (of different ethnicities) and 100 informal street traders of traditional medicine in Johannesburg, South Africa, and a seventh data set from traders on the western boundary of the Kruger National Park, South Africa. The diversity measures were coupled with species accumulation curves to construct cumulative diversity curves used to determine the minimum viable sample size on which a diversity index should be based, and to better understand the differences in the relative diversities of the samples. Distinct differences in the relative abundance and diversity of plants sold by street traders and shop traders were evident. Species diversity and evenness was found to be higher in shops, thus resulting in a lower dominance in the sale of certain plant species compared to the street traders. A survey of an informal market should include no less than 35 research participants compared to no less than 20 for the *muti* shops. The use of selected indices of species richness (Margalef's), diversity (Shannon, Simpson's, Fisher's alpha, Hill's numbers) and evenness are recommended as a means of describing patterns exhibited within ethnobotanical data.

Introduction

An emergent trend in ethnobotanical studies has been the use of quantitative methods and models to describe patterns of plant use and availability in surveys or assessments of natural resources (e.g., Prance et al. 1987; Phillips and Gentry 1993a, b; Johns et al. 1994; Begossi 1996; Höft et al. 1999; Hanazaki et al. 2000; Luoga et al. 2000; Cunningham 2001; Wong et al. 2001), thereby allowing for a more rigorous statistical approach to the discipline. While a quantitative approach to the discipline is not always possible or even necessary, the benefits include: a greater depth to the understanding of the subject under investigation; a conscious attempt at reporting and refining methods employed to collect and evaluate the data; attention to sampling effort; economy of

description and examination of patterns in the data; hypothesis testing; and the ability to question and describe more precisely the results of surveys.

Begossi (1996) first demonstrated how diversity indices were useful quantitative tools that could assist researchers analysing ethnobotanical data by allowing comparisons of diversity among different communities in different or similar environments. In ecology there are numerous indices available for exploring species diversity between different communities. Begossi (1996) demonstrated how the Shannon–Wiener index of diversity and evenness, and the rarefaction curve, might be used to compare differences in diversity, uniformity of species use and sampling effort in eight samples from South America, Thailand and Tonga. This paper broadens the spectrum of diversity indices to include 22 measures of richness, evenness and diversity. The goals are to: evaluate the performance of the indices in relation to samples of different sizes, and trader profiles and to examine the kind of information they provide; make recommendations on measures appropriate for quantifying ethnobotanical data; assess whether the survey sites were adequately sampled, and determine the minimum viable sample size on which a diversity measure should be based for the type of survey data collected; and lastly, to compare the species diversity of sites within the formal and informal sector, and thereby appraise plant availability within the ethnobotanical trade in the region.

Species diversity

Diversity indices are used to characterise species abundance relationships in communities (Ludwig and Reynolds 1988). The literature reveals a multifarious array of ecological indices, usually expressing species diversity as a single number (Magurran 1988). How can the appropriate measure be determined? The answer largely depends on the question the index is being used to answer, the component of diversity being measured, and whether the index is simple to use and understand.

Diversity measures take into account two factors: species richness (i.e., the number of species, S) and evenness/equitability (i.e., how uniformly abundant species are in a sample) (Magurran 1988). An ‘index’ of species diversity (also called an index of heterogeneity) incorporates both richness and evenness into a single value. Species diversity measures are broadly divided into four main categories: indices of species richness; indices of evenness; indices of species diversity/heterogeneity; and species abundance/distribution models (Ludwig and Reynolds 1988; Magurran 1988). These concepts may be translated into ethnobotanical terms to answer the following questions: (1) what is the species richness of plants used/sold within a sampled market or group of resource users?; (2) how does the species diversity of plants sold differ between different groups of traders?; (3) are the same plants sold by most traders?; and (4) is sampling effort adequate, i.e., have sufficient numbers of research/survey participants been interviewed?

Recent research by Colwell and Coddington (1995) and Gotelli and Colwell (2001) recommend the measurement and comparison of species richness by the use of taxon sampling or accumulation curves. The curves may be computed from *EstimateS* (Colwell 2001) software that computes randomised species accumulation curves and also several indices of diversity and the parameters to calculate others. Plotting the performance of the indices on a diversity curve demonstrates the performance of the indices and differences in relative abundance as sample size increases.

Study area

Johannesburg is located within a region of Gauteng Province called the 'Witwatersrand', the name given to an extensively urbanised axis of approximately 100×40 km and part of a geological super-group consisting of gold-bearing conglomerates (Lowrey and Wright 1987). The Witwatersrand emerged from a small mining town in the 1880s and labour for the mines were provided by mainly rural people in the migrant labour system. The ensuing rural–urban oscillation of Black labour from around the country enhanced the introduction of activities related to Black 'rural' culture (Dauskardt 1990, 1991). Traditional herbalism was incorporated into the developing urban mine culture to meet the needs of both the Black migrant labourers and the rapidly expanding, permanent urban population for traditional medicine (Dauskardt 1991). Various historical processes shaped the preponderance of different ethnic and language groups within sectors of the emerging South African capitalist economy and the traditional medicine trade. The Witwatersrand is South Africa's second largest market for medicinal plants after the markets in KwaZulu-Natal, and the ethnic diversity of the region's traders, healers, gatherers and consumers is influential in determining the traded floristic diversity and sources of the plants harvested for the multinational trade (Williams et al. 2000).

The trade is differentiated into two sectors, namely formal business and informal street markets (Williams et al. 1997). Traders, including traditional healers, selling traditional medicines from premises called '*muti*' shops, represent the formal sector. In 1994, there were estimated to be 244 shops in the region selling traditional medicines (Williams et al. 1997), the majority of which were owned by Black traditional healers (52%) and Indian merchants (25%).

Commercial gatherers and traders selling plants from the pavements and street markets, on the other hand, represent the informal sector. Located in Johannesburg, the 'Faraday Street' market is the Witwatersrand's only informal wholesale and retail street market for traditional medicine. Ninety-seven percent of the approximately 166 traders are migrants to the Witwatersrand, of whom 90% regard the province of KwaZulu-Natal as "home" (Williams 2003). Customers to the market are primarily traditional healers from Gauteng townships, owners of *muti* shops, and sometimes patients seeking treatment

from the traders that are traditional healers. Most of the traders earn less than R100 a week (US\$1 \approx R6.51, June 2004) (Williams 2003).

Methods

Market surveys

Between February 1994 and January 2001, two semi-quantitative surveys of plants traded for traditional medicine were conducted within the Witwatersrand. The first survey in 1994, based on a stratified random sample of 50 research participants from 50 *muti* shops, appraised the nature of the plant trade in the formal sector. The second survey in 2001, a stratified random survey of 100 street traders in the Faraday Street informal market, was conducted on behalf of the provincial Directorate for Nature Conservation for Gauteng. The surveys were based on questionnaires and structured interviews, and an inventory of all common names of plants sold by each trader was compiled. The characteristics of the trade and the species sold within the shops and at the market have already been quantitatively and qualitatively described (Williams et al. 2000, 2001; Williams 2003).

Synthesis of plant inventories

Identification of the species traded was mainly achieved by matching vernacular names to botanical names from previously published studies that, for the most part, are reliable because of the credible body of literature existing for ethnobotanical names in South Africa. In some cases, species were visually identifiable or were identified later from purchased specimens. Species identification through published records is problematic, and errors in identification are likely to have occurred. However, this was considered the most expedient mechanism for identifying the large numbers of inventoried species sold by each of the traders surveyed. In order to make a distinction between vernacular names synonymous with >1 species, analysis by '*ethnospecies*' instead of by botanical species was used. '*Ethnospecies*' (Hanazaki et al. 2000) takes into account the folk or common name given to one or several species quoted during interviews. The Zulu name '*iNgwavuma*', for example, is the ethnospecies name designating *Elaeodendron transvaalense* (Burt Davy) R.H. Archer, whilst the ethnospecies '*iMphepho*' applies to at least six species of *Helichrysum*. *iMphepho* was cited 17 times during the Faraday market survey, however, only one of the six potential *Helichrysum* species would have been sold at each stall and the most prevalent species is not known. Therefore, wherever appropriate, the data were quantified based on the number and frequency of occurrence of ethnospecies to avoid repetitions and hence any bias/inaccuracies in reporting the results. In this paper, all citations of *species* are citations for plant *ethnospecies*.

Six data sets derived from the Witwatersrand medicinal plant trade were subject to analysis with the 22 diversity measures. This was to evaluate the effect of different sample sizes, trader ethnicities and formal/informal market sectors in the appraisal of patterns in the utilisation and trade of traditional plant medicines. The sample of 50 *muti* shops (abbreviated as *MS: All 1994* in the graphs) was subdivided into three smaller subsamples based on the ethnicity of the shop owner, namely Black ($n = 28$ shops; *MS: Black 1994*), Indian ($n = 20$ shops; *MS: Indian 1994*) and White ($n = 2$ shops; *MS: White 1994*). The Faraday market data (*ST: Faraday*) were not subdivided for the initial appraisal of the indices. An earlier survey conducted by the author in 1992 of seven *muti* shops (*MS: 1992*) was included to compare the effect of sample size (Williams 1992). Whenever appropriate, the performance of the indices was compared with a seventh data set – a sample of 17 informal vendors derived from an inventory compiled for medicinal plants traded on the western boundary of the Kruger National Park, South Africa (Botha 2001; Botha et al. 2001). The dataset is abbreviated as '*ST: WBKP*' in the graphs. Later, by way of an independent example comparing intra-sample diversity for selected indices, the Faraday data matrix was subdivided into 'healer' and 'non-healer' traders.

Calculating indices

The calculation of an index to evaluate ecological diversity relies on information regarding the *number* and *frequency of occurrence* of species in a sampled community. In order to calculate diversity indices for ethnobotanical purposes it is necessary to have data on the number of *individual informants* (e.g., resource users/traders) who *cited* the plant species (Begossi 1996). Since the inventory of plants sold by each trader in this study recorded the presence of an ethnospecies, each ethnospecies was recorded/cited once per trader and, therefore, incidence/occurrence equals abundance of the ethnospecies per trader sample/inventory.

References for the formulae and software used to calculate the indices are listed in Table 1. The statistical distributions used to fit species abundance observations may be used for fitting species occurrences (Hayek and Buzas 1997). N occurrences may be substituted for N individuals (Hayek and Buzas 1997). In the calculations, n = number of samples (e.g., number of *muti* shops or street traders surveyed) and N = number of citations or occurrences of ethnospecies.

The species accumulation curves and cumulative diversity curves were constructed from variables and diversity statistics computed by *EstimateS* (Colwell 2001). In cases where *EstimateS* did not directly compute the diversity measure (e.g., Margalef and Menhinick's indices, Hill's diversity numbers N_1 and N_2 , evenness measure E_1 – E_5), then the formulae cited in Magurran (1988) and the appropriate variables computed by *EstimateS* were used to calculate the indices and construct the graphs. It is important to note that *EstimateS* computes the

Table 1. References for methods used to compute the indices and measures applied to the data.

Index/Measure	Reference
<i>Species richness indices</i>	
# Species (S or N_0)	Discerned from observation of the data set
Margalef (D_{Mg})	Ludwig and Reynolds (1989) – program ‘SPDIVERS.BAS’ or Magurran (1988) and Colwell (2001)*
Menhinick (D_{Ma})	Ludwig and Reynolds (1989) – program ‘SPDIVERS.BAS’ or Magurran (1988) and Colwell (2001)*
<i>Diversity indices</i>	
Shannon–Wiener (H')	Ludwig and Reynolds (1989) – program ‘SPDIVERS.BAS’ or Colwell (2001)*
Shannon–Wiener (H_{max})	Magurran (1988)
Comparing Shannon indices	Zar (1984); Magurran (1988); Murali et al. (1996) or use standard deviation for H' from Colwell (2001) *
Brillouin measure (HB)	Zar (1984); Magurran (1988); Krebs (1989)
Simpson (λ)	Ludwig and Reynolds (1989) – program ‘SPDIVERS.BAS’ or Colwell (2001)*
Simpson ($-\ln \lambda$)	Pielou (1975); Colwell (2001)*
Berger–Parker (d)	Magurran (1988)
McIntosh’s dominance (D)	Magurran (1988)
Fisher’s alpha (α)	Magurran (1988); Krebs (1989); Colwell (2001)*
Hill’s Diversity Number N1	Ludwig and Reynolds (1989) – program ‘SPDIVERS.BAS’ or Magurran (1988) and Colwell (2001)*
Hill’s Diversity Number N2	Ludwig and Reynolds (1989) – program ‘SPDIVERS.BAS’ or Magurran (1988) and Colwell (2001)*
Hill’s Diversity Number N_∞	Magurran (1988)
<i>Evenness measures</i>	
Shannon (J' or $E1$) (or, Pielou’s J)	Ludwig and Reynolds (1989) – program ‘SPDIVERS.BAS’ or Magurran (1988) and Colwell (2001)*
$E2$ (or, Buzas’ & Gibson’s E)	Ludwig and Reynolds (1989) – program ‘SPDIVERS.BAS’ or Magurran (1988) and Colwell (2001)*
$E3$, $E4$ and $E5$	Ludwig and Reynolds (1989) – program ‘SPDIVERS.BAS’ or Magurran (1988) and Colwell (2001)*
Brillouin (J)	Krebs (1989)
McIntosh’s (E)	Magurran (1988)

*Indicates that *EstimateS* (Colwell 2001) either computes the index and/or the parameters that can be inserted into the equations obtained from the other references listed.

reciprocal of Simpson’s λ . In the formulae for diversity indices, any logarithmic base may be used (Zar 1984). As a way of standardising the results, the natural log (\ln) was used throughout.

Results and discussion

Species richness

Numerical species richness (S), or the *number of species* in a sample of a specified size, is an instantly comprehensible expression of species diversity

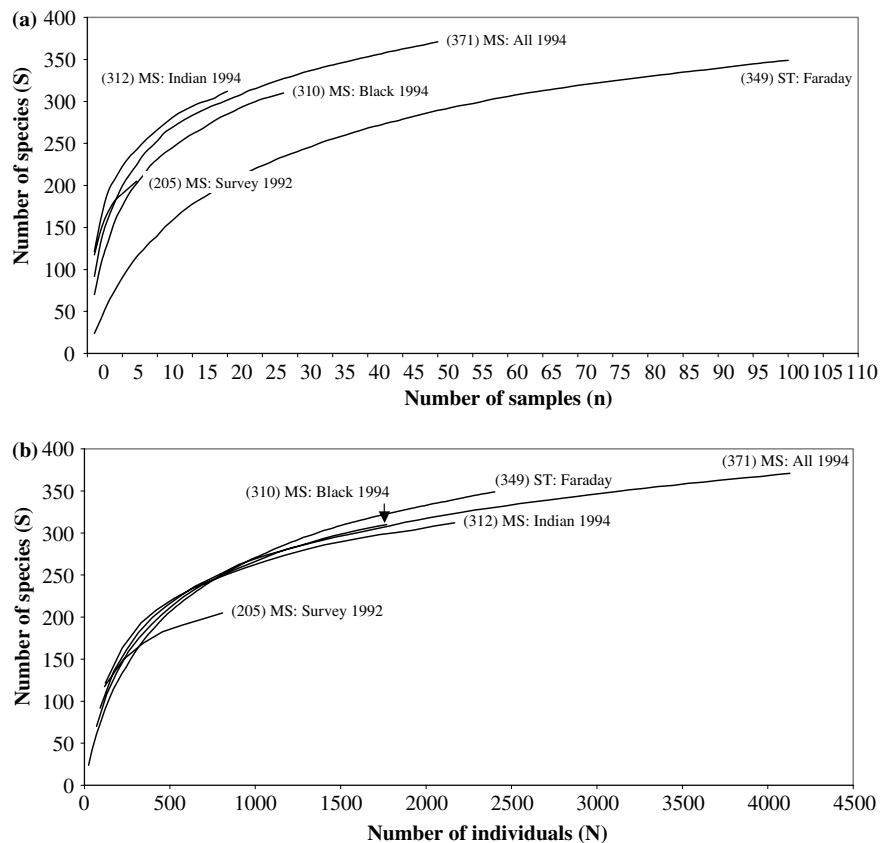


Figure 1. Species accumulation curves (or collector's curves) for plant ethnospecies traded as traditional medicine in muti shops and in a street market in the Witwatersrand. The curves represent successively pooled and randomly ordered samples (a) and individuals (b). The curves were computed using *EstimateS* (Colwell 2001). The total number of ethnospecies per sample (S) is labelled in brackets at the end of each curve. (MS: *Muti shops*; ST: *Street traders*).

(Magurran 1988). S is related to the total number of individuals (N) summed over all S species recorded. As sampling effort increases, more individuals are encountered and more species are likely to be recorded (Hayek and Buzas 1997). The relative abundance of species, however, is important and a number of simple indices have been derived using some combination of S and N . These indices include Margalef's and Menhinick's index of species richness. While these indices are easy to calculate, they are (like S) sensitive to sample size.

The relationship between S and N may be viewed by plotting a *species accumulation curve* (Figure 1a, b), also termed a *species effort curve* or *collector's curve* – so called because the cumulative number of species is plotted against some measure of the effort it took to obtain that sample of species (Hayek and Buzas 1997). Compared to interpreting the single numerical value of species

richness for the randomly pooled samples (Table 2), plotting the curves facilitates improved interpretation of species richness results for different samples of varying size. Comparing raw taxon counts (and index values) for two or more assemblages/samples will quite generally produce misleading results (Gotelli and Colwell 2001). Differences in measured species richness between communities may be the result of differences in underlying species richness, differences in the shape of the relative abundance distribution, or because of differences in the number of individuals counted (Gotelli and Colwell 2001).

Whereas fewer *muti* shop traders were sampled compared to street traders (n) (Figure 1a), the number of individual plants (N) recorded in the shops was greater (Figure 1b), and hence the numerical richness per trader is greater for the *muti* shops. There is also a similarity in shape and clustering of curves for the shop data ('*MS*'), even the 1992 survey of seven shops, making them distinct from the curve of the street trader ('*ST*') sample (Figure 1a). The initial steep gradients of the curves for the shop data show that more ethnospecies per shop are sold (mean = 83) and consequently the accumulation of ethnospecies is more rapid, even for smaller sample sizes. Street traders, by contrast, sell fewer ethnospecies per trader (mean = 24) and consequently the accumulation of ethnospecies is slower. When samples are highly variable in terms of plant diversity amongst traders, then more samples are needed to fully represent the trade in medicinal plants, while if the samples show little variation then fewer traders need to be sampled.

The Margalef and Menhinick indices have been cited as being inadequate by several authors (e.g., Brower and Zar 1977; Magurran 1988; Hayek and Buzas 1997) because the indices lack the ability to differentiate the species richness of samples having similar S and N . Looking at Table 2, the performance of these indices as a single numerical value for pooled samples is difficult to adequately judge. However, plotting the performance of an index as samples are successively pooled and individuals are accumulated is a useful procedure for aiding the interpretation of plant availability within the different trader groups (Figures 2 and 3).

The graphs of Margalef's index show how species richness increases until eventually the curve levels off with increasing sample size and the number of individuals inventoried (Figure 2a, b). The point at which the curve flattens indicates a minimum viable sample size on which a diversity or richness index should be based (Magurran 1988). The curve of S (Figure 1) can also be constructed for this purpose i.e., to determine the minimum requisite sampling effort. The sample of species sold by White shop traders ($n = 2$) is numerically inadequate and only two points of richness could be plotted on the graphs. However, the sampling strategy for the 50 traders selected for the shop survey (including the 2 White traders) was stratified random, and therefore trader ethnicities were proportionately representative within the sample to minimize participation biases (Williams et al. 1997). Sampling more White traders would only have been necessary if it had been an important aim to compare the shops of different trader classes, but it would have biased the overall results of the

Table 2. Species richness indices calculated for six data sets sampled from Witwatersrand traders of traditional medicine. n = the number of samples (traders and/or shops surveyed); N = number of individual ethnospecies observed; S = number of ethnospecies counted.

Index/measure	1992 <i>Muti</i> Shop Survey ($n = 7$) $N = 809$	1994 <i>Muti</i> Shop Survey ($N = 50$)			2001 Street trader survey ($n = 100$) $N = 2402$	
		'White-owned' ($n = 2$) $N = 193$	'Indian-owned' ($n = 20$) $N = 2168$	'Black-owned' ($n = 28$) $N = 1769$	Total ('All') shops ($n = 50$) $N = 4129$	
# Species (S or N_0 or $e^{H_{\max}}$)	205	144	313	310	371	349
Margalef (D_{Mg})	30.5	27.2	40.5	41.3	44.4	44.7
Menhinick (D_{Mn})	7.2	10.4	6.7	7.4	5.8	7.1

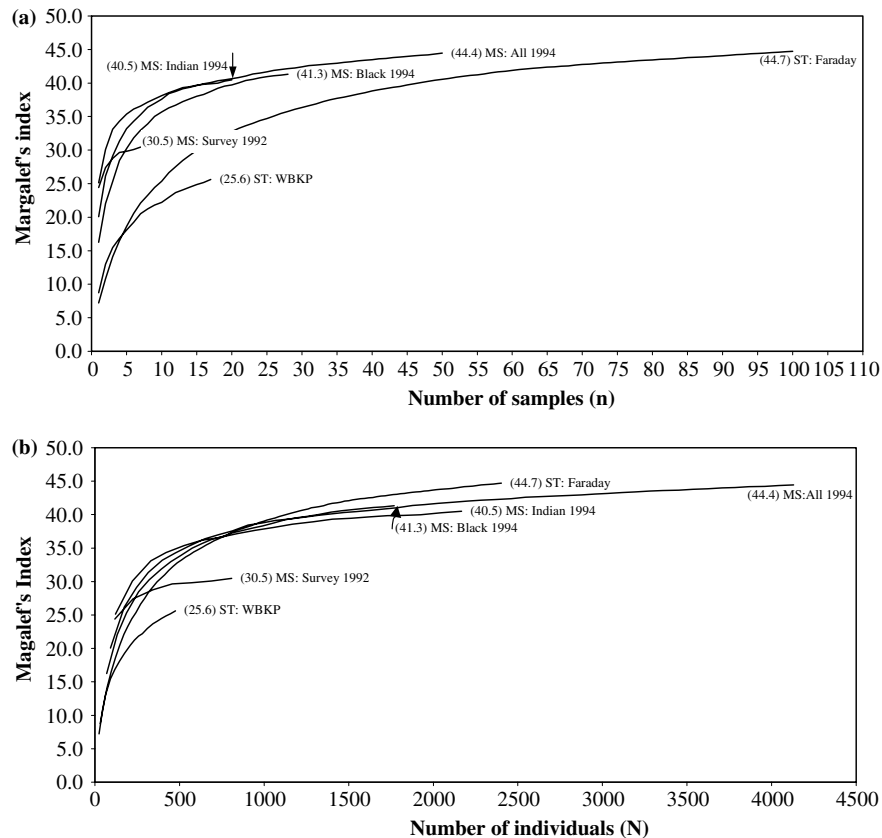


Figure 2. The cumulative species richness curves of Margalef's index for ethnospecies traded on the Witwatersrand for both samples (a) and individuals (b). 'WBKP' is a sample of 17 informal vendors trading medicinal plants on the western boundary of the Kruger National Park, South Africa (Botha 2001). The overall value of the index for the randomly pooled samples is labelled in brackets at the end of each curve. The formula for the index is $D_{Me} = (S-1)/\ln N$. (MS: *Muti* shops; ST: Street traders).

study. The minimum viable sample size (i.e., the number of research participants) necessary for assessing species richness is larger for informal street traders than for shop traders. The evidence for this is reflected in the species accumulation and diversity graphs for street traders, which show that a larger sampling effort is necessary before the curves begin to reach an asymptote (Figures 1 and 2).

Evidence for the distinctive trading patterns in species richness in formal and informal trading sectors are supported by the sample of 17 informal vendors from the western boundary of the Kruger National Park (*ST: WBKP*, Figure 2a). Despite the small sample size (which was conducted with 73% of the vendors at the site), the curve adopts a similar shape to the Faraday Street trader sample (*ST: Faraday*) and exhibits low ethnospecies richness per trader.

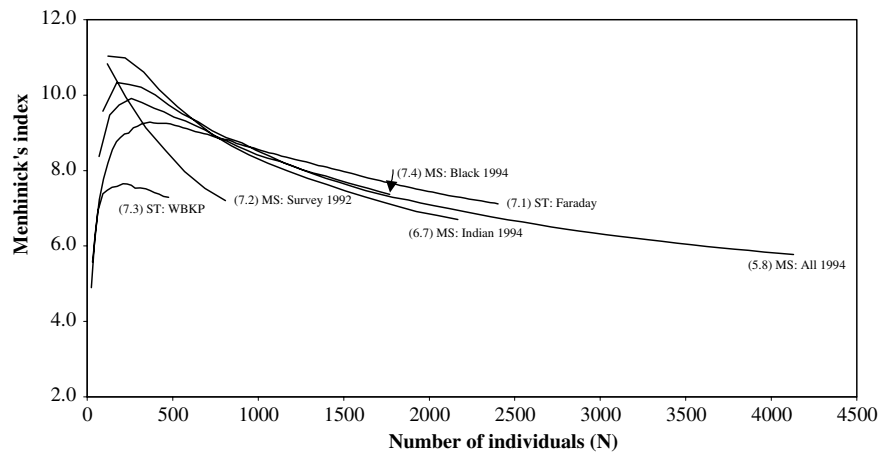


Figure 3. The cumulative species richness curves of Menhinick's index for ethnospecies traded on the Witwatersrand. The overall value of the index for the randomly pooled samples is labelled at the end of each curve. The formula for the index is $D_{Mn} = S/\sqrt{N}$. (MS: Muti shops; ST: Street traders; WBKP: Western Boundary Kruger National Park).

Gotelli and Colwell (2001) recommend that when *comparing species richness between samples*, the number of taxa should be plotted as a function of the accumulated number of *individuals* (not *samples*) because datasets may differ systematically in the mean number of individuals per sample. Figure 2b, therefore, shows that the relative abundance and richness of ethnospecies sold at the Faraday street market is higher than that of the shops. Margalef's index for Faraday is 44.7 compared to 44.4 for 'All shops' (MS: All). The relative abundance and richness of ethnospecies sold by Black traders is slightly higher than that of Indian traders.

The graph of Menhinick's index (Figure 3) corroborates earlier evidence derived from Margalef's index that street traders keep a lower species richness per trader, and therefore increased sampling effort is required for the curve to reach a peak before declining as n and N increase. The ST: WBKP sample shows a similar pattern to the Faraday Street trader data, which is different to the pattern shown by the *muti* shops. The numerical richness values for the index similarly indicate the Faraday sample to be relatively more species rich than the 'All shops', and the Black trader sample to be more species rich than the Indian trader sample.

Species diversity or heterogeneity

Indices of diversity or heterogeneity incorporate both richness and evenness into a single value and are based on *the proportional abundance* of species in a sample (Ludwig and Reynolds 1988; Magurran 1988). These measures are

attractive to researchers because, unlike the species abundance models, no assumptions are made about the underlying distributions of the data (Hayek and Buzas 1997). There are four categories of indices. First are measures derived from *information theory* (e.g., Shannon–Wiener and Brillouin), based on the rationale that diversity or information in a natural system may be measured in a similar way to information contained in a code or message (Magurran 1988). The second category of indices are the *dominance* indices (e.g., Simpson, McIntosh and Berger–Parker), so called because they are weighted towards the abundances of the commonest species (Magurran 1988). A third category of diversity index is ‘*Hill’s diversity numbers*’. The numbers, developed by Hill (1973), show how diversity indices are mathematically related and may be arranged in a sequence depending on whether they measure species richness (weighted towards uncommon species) or dominance (weighted towards abundant species) (Magurran 1988). Interpreting the single statistic for each index can be problematic (Table 3). A fourth category of diversity index is derived from the logarithmic series abundance model, namely Fisher’s alpha.

Information theory indices

The Shannon index (H') measures the average degree of “uncertainty” in predicting to what species individuals chosen at random will belong (Ludwig and Reynolds 1988). Uncertainty may be visualised as being synonymous with diversity (Krebs 1989), therefore, the higher the degree of uncertainty, the higher the diversity and the greater the degree of uncertainty in correctly predicting the identity of the next species chosen at random. The average uncertainty (H') increases as S increases, as seen in Figure 4 when compared with Figure 1. Figure 4a shows that there is a distinction in the curves between formal and informal traders (Figure 4a), with the overall degree of uncertainty and diversity being higher in the formal sector (Figure 4b). The higher predictability of a species’ identity in the street markets is because of the lower mean species richness per trader, as discussed in the previous section. Informal traders sell fewer species and one may more comfortably predict what popular species most traders are likely to sell. In terms of determining the minimum sample size necessary to assess the Shannon index (as indicated by the point at which the curve levels off), the curve for shop traders begins to reach an asymptote at around 15–25 samples, compared to 25–30 for the street traders.

Maximum uncertainty (H_{\max}) will occur when each species in a sample is equally represented (Table 3) (Hayek and Buzas 1997). The more species there are, the more evenly the individuals are spread across the species, the higher will be the value of H' because there will be greater uncertainty as to which species will most likely be observed next time they are chosen at random. It appears that a characteristic of ethnobotanical samples (especially those of large sample sizes) is that H' is high. In examples described in Magurran (1988) for “natural” communities (e.g., diversity of birds in woodlands; species diversity in plantations etc.) H' ranged between 1.38 and 3.54. By contrast, Begossi (1996) estimated H' to be between 2.99 and 5.95 (average = 4.6) for

Table 3. Species diversity indices calculated for six data sets sampled from Witwatersrand traders of traditional plant medicine.

Index/Measure	1992 Muti Shop Survey ($n = 7$) $N = 809$	1994 Muti Shop Survey ($N = 50$)			2001 Street trader survey ($n = 100$) $N = 2402$
		'White-owned' ($n = 2$) $N = 193$	'Indian-owned' ($n = 20$) $N = 2168$	'Black-owned' ($n = 28$) $N = 1769$	Total ('All') shops ($n = 50$) $N = 4129$
Shannon-Wiener (H')	5.16	4.91	5.43	5.38	5.46
Shannon-Wiener (H'_{\max})	5.32	4.97	5.75	5.74	5.92
Brillouin measure (HB)	2.08	1.78	2.25	2.22	2.30
Simpson (λ)	0.0050	0.0026	0.0047	0.0051	0.0049
Simpson ($-\ln \lambda$)	5.29	5.95	5.36	5.28	5.32
Berger-Parker (d)	0.0087	0.0104	0.0088	0.0113	0.0094
McIntosh's dominance (D)	0.954	0.982	0.949	0.947	0.942
Fisher's alpha (α)	88.49	257.28	99.95	108.85	98.75
Hill's Diversity Number $N1(e^{H'})^*$	174.9	135.7	227.5	216.8	233.9
Hill's Diversity Number $N2(1/\lambda)^*$	198.9	378.1	213.2	195.8	200.1
Hill's Diversity Number N_{∞}^*	115.6	96.5	114.1	88.4	105.9

*The result is a count of the number of ethnospecies.

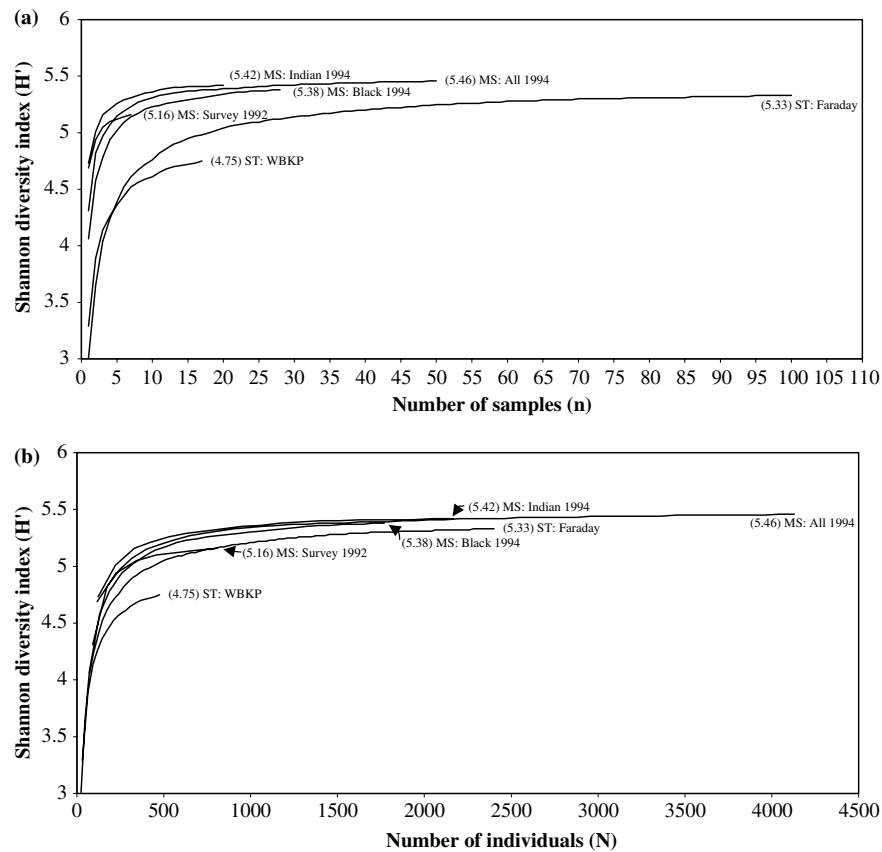


Figure 4. The cumulative diversity curves of Shannon's diversity index (H') for ethnospecies traded on the Witwatersrand for samples (a) and individuals (b). The overall value of the index for the randomly pooled samples is labelled in brackets at the end of each curve. (MS: *Muti* shops; ST: Street traders; WBKP: Western Boundary Kruger National Park).

eight ethnobotanical samples from mainly South America. In the South African study, H' ranged between 4.91 and 5.46 (Table 3).

When the Shannon index is obtained for two or more samples it is possible to test the null hypothesis that the diversities of the samples are equal (Zar 1984). The differences in the index between the samples is mostly significant at $p < 0.000001$, except for three comparisons which are approaching an equal diversity. The diversity of plants sold by Black traders when compared with both the diversities of the Faraday traders and Indian-owned shops is $p = 0.0021$. The least significantly different samples are those of Indian traders and All shops, where $p = 0.0017$ – this would suggest that the characteristics and diversity of All Shops ($n = 50$) is largely due to the influence of the sample of Indian traders ($n = 20$) within it.

The Brillouin index (HB) is similar to Shannon, and the use of this index instead of Shannon is recommended when randomness of a sample cannot be guaranteed (Magurran 1988). The values for HB are lower than H' (Table 3) and show similar numerical rankings for species diversity between samples. The obstacle to using this index is the calculation of very large factorials; additionally, it was not possible to derive the accumulated diversity curve. While many authors recommend the use of HB over H' (e.g., Magurran 1988), the simplicity of calculating Shannon is to its advantage and has led to its widespread acceptability as an index – however, Shannon is sensitive to sample size, thus indices such as Simpson's and Fisher's are sometimes preferred by researchers. Magurran (1988) says that “ideally (Fisher's) alpha should replace the Shannon index as the preferred measure”.

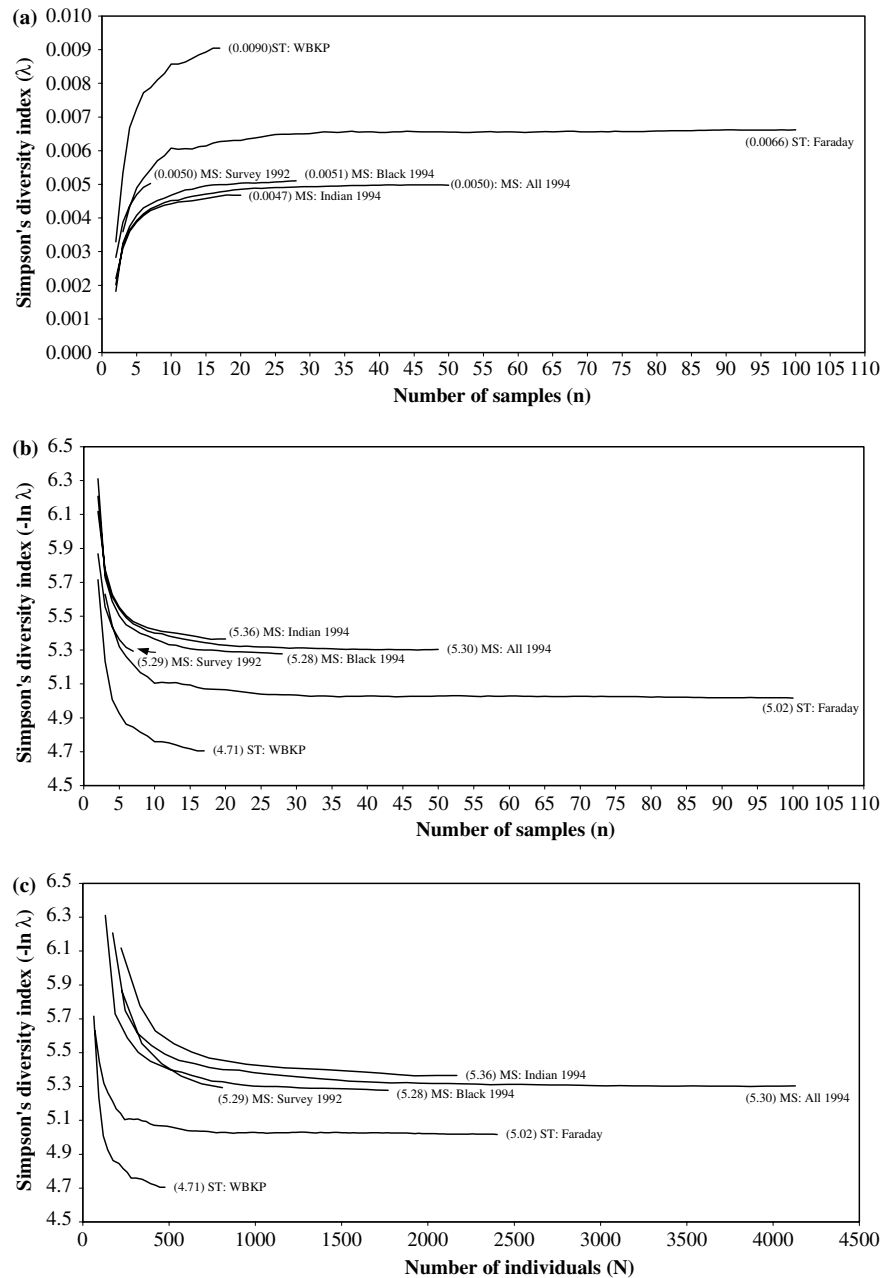
Dominance indices

Simpson's index (λ) proposes that diversity is inversely related to the probability that two individuals picked at random from a sample belong to the same species. Simply stated, if the probability (λ) is high (λ approaches 1) that both individuals belong to the *same* species, then the diversity of the sample is low and *vice versa* (Pielou 1975; Ludwig and Reynolds 1988; Krebs 1989) (e.g., Figure 5a). However, because λ decreases as diversity increases, Simpson's index is usually expressed as $1 - \lambda$ (the probability that two individuals chosen at random are *different species*) or $1/\lambda$ (also known as Hill's number N_2 , which is a measure of the number of very abundant species in a sample). A rarely cited function of λ is $-\ln \lambda$ (Pielou 1975) (Figure 5b, c), and is preferred by the authors as a diversity indicator. The function does not represent a probability, but a single diversity statistic that increases as diversity increases.

Indian shop traders, followed by the total sample of shops (*MS: All*) and Black traders, have the largest diversity of plants for sale, especially compared to the street traders (Figure 5c). As with the other indices discussed so far, there is a distinction between the diversity of plants for sale by the *muti* shops and street traders. There is, therefore, a higher probability that two plants selected at random from different street trader stalls belong to the same species than for two plants selected from different *muti* shops (Figure 5a). As a result, there is a greater dominance of certain species sold by the street traders.

When $\lambda = 1$, most individuals from a sample are concentrated in a single species, and therefore dominance of species within the sample is high. Values for λ for the ethnobotanical samples investigated in this paper are relatively low ($\lambda < 0.008$), and therefore the overall dominance of species is relatively low and diversity is relatively high. By comparison, Hanazaki et al. (2000) calculated λ to be between 0.015 and 0.06 for plants used by native inhabitants from the Atlantic Forest coast in south-eastern Brazil. The overall diversity of plants used by the community investigated in Brazil is, therefore, lower and dominance of plants is higher when compared with the South African study.

The minimum sample size necessary for evaluating Simpson's index is between 15–20 for street traders and 20+ for *muti* shops (Figure 5b). The graph



shows that the sample '*MS: Survey 1992*' is too small to assess the index, and therefore comparisons of diversity with the other samples. Additionally, curves for the street traders show that despite the smaller number of individuals sold

Figure 5. The cumulative diversity curves of Simpson's diversity index (λ) for ethnospecies traded on the Witwatersrand for samples (a, b) and individuals (c). (a) plots the standard form of the index, namely λ , the probability that two individuals will belong to the same species. b and c plot $-\ln \lambda$, a rarely cited form of the index recommended by Pielou (1975) that expresses λ as a diversity statistic that increases as diversity increases. The overall value of the index for the randomly pooled samples is labelled in brackets at the end of each curve. (MS: *Muti* shops; ST: Street traders; WBKP: Western Boundary Kruger National Park).

per street trader, that dominance of individuals related to their relative abundance is established sooner (indicated by the point at which the curve levels off). This confirms the Shannon index results, i.e., that the probability of encountering the same species is higher between street traders. Additionally, there are fewer abundant species sold by the street traders compared to the *muti* shops (as demonstrated by the curve for N2 [or $1/\lambda$] in Figure 6).

The Berger–Parker index is a dominance measure that expresses the proportional abundance of the most abundant species ($d = N_{\max}/N$) (Table 3). The index is independent of S , but is subject to bias caused by fluctuations in the abundance of the commonest species (Magurran 1988). Like Simpson's index (λ), diversity increases and dominance decreases as d decreases. The results are concordant with Shannon and Simpson's index, except for the 'MS: Survey 1992' sample which shows a higher diversity than for the Indian traders. However, it was established in Figure 5a that the sample size was too small in this sample to assess species dominance and hence the result is disregarded in

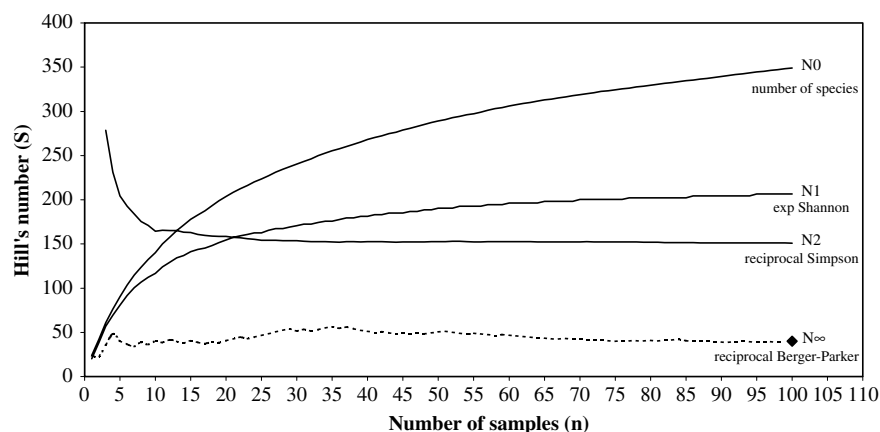


Figure 6. The cumulative diversity curves for Hill's numbers (N_0 , N_1 , N_2 and N_∞) for ethnospecies traded in the Faraday Street market. The curves for N_0 , N_1 , and N_2 were based on successively pooled and randomly ordered samples, whilst N_∞ has been estimated from non-randomised values directly from the dataset because the requisite parameters for a randomised and cumulative curve were not available on the software. The final value for N_∞ is shown as \diamond because it does not depend on the order of the randomised data matrix. The unit on the y-axis is number of plant ethnospecies.

this instance. It was not possible to plot the curves for the performance of this index as *EstimateS* does not calculate the requisite parameters.

McIntosh's index (D) is a measure of diversity independent of N . As dominance increases (related to the increase in abundance of species in the sample), the value of D appears to decrease. However, the discriminant ability of this index in samples of different sizes is poor and, coupled with the inability to graph the index, the performance and usefulness of this index was difficult to evaluate.

Hill's diversity numbers

Hill's numbers are the easiest diversity statistic to interpret. The diversity numbers are in units of *number of species* and measure what Hill calls the *effective number of species* present in a sample (Hill 1973; Ludwig and Reynolds 1988). The numbers are mathematically related to the Shannon, Simpson and Berger–Parker indices (Table 3). As the number of species (N_0 or S) increases, less weight is placed on 'rare' species, and lower values are obtained for N_1 , N_2 and N_∞ since they measure the number of *abundant*, *very abundant* and *most abundant* species in a sample respectively (Table 3). 'Rare' species in an ethnobotanical context are those species with low incidences/abundances in the sample (i.e., low values of n).

Hill's numbers may be plotted, for example the Faraday Street trader sample in Figure 6. Like the other indices previously described, the point at which the curve levels off gives an indication of the minimum viable sample size needed to assess the index. The minimum sample size required for the street traders is around 30 (Figure 6), compared to ± 20 for shop traders (not shown graphically). The distinctive high values for the first part of the N_2 curve (derived from Simpson's index) is because N_2 is weighted in favour of the commonest species. Because incidence equals abundance in these samples, adding new or 'rare' species to the sample (as n or N increases), decreases the relative abundance of the commonest species until the value stabilises when fewer new species are added. N_1 , on the other hand, shows a steady increase until the curve reaches an asymptote – this is because the function is derived from the Shannon index, and therefore weighted in favour of species richness. As the number of new species increases, the value of the curve increases until it eventually levels off when very few new species are added to the sample as n increases.

It was not possible to plot N_∞ for the randomised and successively pooled samples because the requisite parameters were not available on the *EstimateS* software. Except for the final value of N_∞ (shown as ♦ in Figure 6 for the total pooled sample), the projection of the curve has been estimated from non-randomised values from the data matrix for Faraday. The curves for the *muti* shop samples are similar in shape to the street trader data, but higher in value. The curves are not plotted because Table 3 adequately expresses the results.

Values for N_1 , N_2 and N_∞ are higher for the *muti* shops than for the Faraday Street market (Table 3), indicating that if species abundance equates

with plant popularity, then there is greater dominance of a few popular (abundant) species in the street market. In other words, the street traders sell a smaller number of ethnospecies that have very high occurrences within the market. *Muti* shops, on the other hand, sell a larger range of species with equally high abundances.

This pattern is related to factors already discussed in this paper, i.e., that street traders have neither the space nor the financial capacity to sell large numbers of species. They, therefore, sell a smaller range of ethnospecies that have assured commercial value and are likely to have a higher restocking potential. Additionally, if one trader does not have a certain plant that the customer is looking for, then there are at least 160+ other traders in the market that might sell the plant. *Muti* shop traders, on the other hand, have larger trading spaces and financial freedom and can afford to stock a large range of species in their shop – i.e., they are ‘one-stop-shops’. The number of species represented by N_2 and N_∞ are indicators of the number of ethnospecies within the markets that are candidates for more immediate conservation action, assuming that high incidence is an indicator of potential risk.

Fisher's alpha

Generated from a species abundance model, Fisher's alpha (α) is a constant used to fit the logarithmic series distribution model once the parameter x has been solved for iteratively. The index is also known as the log series alpha (Magurran 1988), and has been adjudicated as a good, if not one of the best, measures of species diversity by several authors (e.g., Magurran 1988; Krebs 1989; Hayek and Buzas 1997) even when the underlying species abundances do not follow a log series distribution.

Alpha is low when the number of species is low (Table 3, Figure 7), and therefore the smaller samples with fewer ethnospecies have smaller values of α (e.g., WBKP and MS: Survey 1992). The index is less affected by the abundance of the rarest or commonest species than either H' or λ , respectively (Magurran 1988), and depends more on the number of species of intermediate abundance. According to Hayek and Buzas (1997), Fisher's α is a number close to the number of species we expect to be represented by one individual – this would account for the high values of α in the initial part of the curves in Figure 7. Because the incidence of species sold at trader's shops/stalls equals abundance, the samples all initially have very high numbers of ethnospecies represented by one individual due to the nature of the sampling methods.

Hayek and Buzas (1997) recommend that α is used as a diversity measure when the parameter x of the log series model is $1 \geq x \geq 0.61$ because when $x \leq 0.61$ then $\alpha > S$ and the statistic becomes unacceptable and biologically meaningless (where $x = N/[N + \alpha]$). Greater confidence in the true value of α occurs when x is close to 1 and N is large. The point on the curves at which $x \geq 0.61$ is marked with *, and the final value of x is also shown (Figure 7). Eventhough α is a constant, it would appear from the data that α increases with N and S . Hayek and Buzas (1997) explain this apparent paradox in the

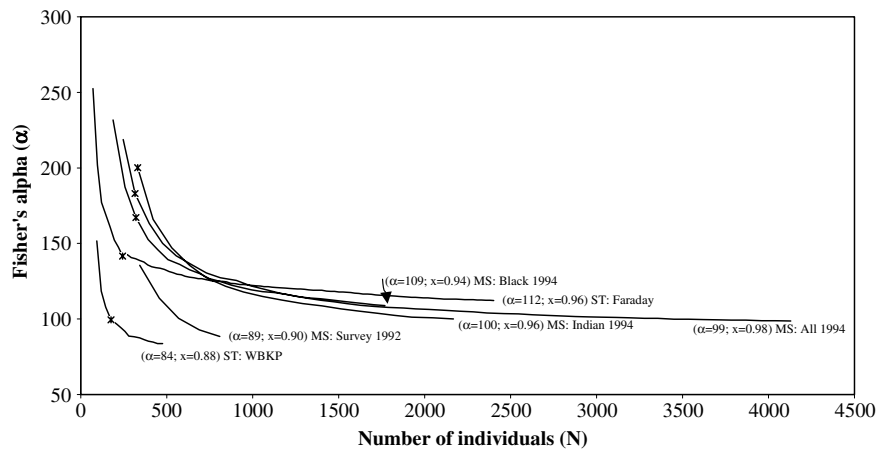


Figure 7. The cumulative diversity curves of Fisher's α for ethnospecies traded in the Witwatersrand. The overall value of α and x (a parameter required to fit the log series model) for the randomly pooled samples is labelled in brackets at the end of each curve. The point marked * on the curves is where $x \leq 0.61$. (MS: *Muti* shops; ST: Street traders; WBKP: Western Boundary Kruger National Park).

following way: if the data fits the log series model *exactly*, then α is a constant independent of N . However, data rarely fits the models exactly in “Mother nature”, and therefore an increase in α is observed as S and N accumulate. The results in Figure 7 show the Faraday Street traders to have greater species diversity than the other samples.

Evenness

Measures of evenness (or equitability) attempt to quantify the unequal representation of species against a hypothetical sample in which all species are equally abundant (Krebs 1989), i.e., the ratio of observed diversity to maximum possible diversity. Hence, evenness may be referred to as *relative diversity* or *homogeneity* (Brower and Zar 1977; Zar 1984). A low evenness means a high dominance in the use (or presence) of a few species (Begossi 1996). When all species are equally abundant, an evenness index would be at a maximum (of 1.0) and decrease towards zero as the relative abundances of the species diverge away from evenness (Ludwig and Reynolds 1988). Different measures of evenness have been proposed, most of which are expressions of Hill's numbers (Table 4).

All the indices gave different values but consistent rankings for the samples and subsamples (Table 4). Which index should, therefore, be chosen as a representative measure of how evenly distributed are the species sold by the Witwatersrand traders? Magurran (1988) recommends the use of the Brillouin

evenness index, but the computation of very large factorials made it impossible to calculate. E1, also called the Shannon J' or Pielou's J , is probably the most common evenness index in use but is strongly affected by species richness, and the addition of rare species (or singletons) can greatly change the value of E1 (Ludwig and Reynolds 1988). Hayek and Buzas (1997) recommend the use of E1 and E2 (also known as the Buzas and Gibson E). Ludwig and Reynolds (1988) further describe E3–E5, but consider E1–E3 to be of limited value because they are highly sensitive to the number of species in the sample. A general problem with all measures of evenness, however, is that they assume that the total number of species that could possibly be sampled is known (Krebs 1989). Since observed species numbers must always be less than true species richness, the evenness ratios are always *overestimated*, with the possible exception of E4 and E5.

E4 and E5 remain relatively constant with sampling variations and hence tend to be independent of sample size (Ludwig and Reynolds 1988). This is because E4 and E5 are computed as ratios where S is in both the numerator and the denominator, thus effectively cancelling the impact of the number of species in the sample (Ludwig and Reynolds 1988). However, E4 and E5 are not totally unaffected by the large number of singletons found in small samples, including the samples collected in the initial stages of research at a site before an adequate sample size is accumulated. Figure 8a shows how $E5 > 1.0$ until about sample 17 in the *muti* shops and about sample 22 in the street traders. An explanation for this feature of the index is as follows: initially $N2 > N1$ (Figure 6) because 'rare' species and singletons are to begin with very abundant in the ethnobotanical samples, thereafter declining in numbers as samples accumulate and the more dominant species become evident in the sample. This feature of the index is useful for determining the minimum viable sample size required for assessing evenness. E5 for samples *MS: White* and *MS: Survey 1992* is never less than 1.0 (Table 4), and therefore their evenness cannot be compared with the other samples. The results demonstrate that evenness is higher in the sample of Indian shop traders, followed by Black, All and Faraday Street, and therefore there is greater dominance in the sale of few species within the street market (Table 4). This result is consistent with the observations described earlier, namely that a high sample diversity means that it is more difficult to correctly predict a species chosen at random from a sample, and therefore the dominance of species is lower and evenness higher.

The values of E2 and E3, as well as E4 and E5, are similar, and therefore either may be used. However, Ludwig and Reynolds (1988) recommend the use of E5 as a measure of evenness because it is the least ambiguous. The authors also suggest calculating E1 because it is more widely used as a comparative index (e.g., Begossi 1996). The performance of E1 as an index is shown in Figure 8b for the sample of Faraday Street traders and shops (*MS: All*). The sensitivity of E1 to the addition of 'rare' species (singletons) is evident in the first part of the curve. When the Witwatersrand results are compared with six ethnobotanical samples from South America (Begossi 1996), it is evident that

Table 4. Species evenness indices calculated for six data sets sampled from Witwatersrand traders of traditional medicine.

Index/Measure	1992 <i>Muti</i> Shop Survey ($n = 7$) $N = 809$	1994 <i>Muti</i> Shop Survey ($N = 50$)	2001 Street Trader Survey ($n = 100$) $N = 2402$
		'White-owned' ($n = 2$) $N = 193$	'Indian-owned' ($n = 20$) $N = 2168$
		'Black-owned' ($n = 28$) $N = 1769$	Total ('All') shops ($n = 50$) $N = 4129$
E1 (Shannon J' or Pielou's J) (H'/H_{\max})	0.970	0.945	0.938
E2 ($N1/N0$, or Buzas & Gibson E)	0.853	0.727	0.699
E3 ($N1-1/N0-1$)	0.853	0.726	0.698
E4 ($N2/N1$)	1.137	0.937	0.903
E5 ($N2-1/N1-1$)	1.138	0.937	0.903
McIntosh's (E)	0.990	0.984	0.980
Brillouin (J)	Can't compute	Can't compute	Can't compute
			Can't compute

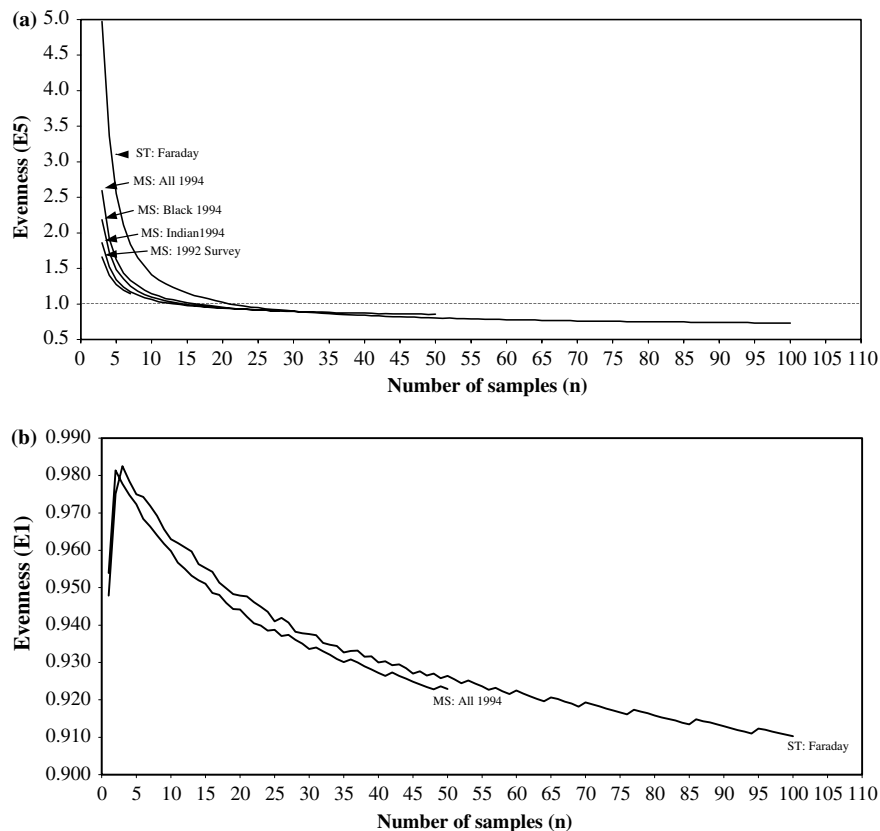


Figure 8. The cumulative evenness curves for (a) E5 and (b) E1 for ethnospecies traded in the Witwatersrand. (MS: *Muti* shops; ST: Street traders).

evenness is high overall for ethnobotanical samples ($E1 > 0.90$ on average), with very little overall dominance of certain species for use/sale. Values for E1 in the South American study range between 0.78 and 0.97 (average 0.91) compared to an average of 0.92 for the Witwatersrand traders.

Assessment of rare, intermediate and common ethnospecies

Indicator species are a useful adjunct to investigations of diversity (Magurran 1988). In ecology, they can provide an additional clue to how community structure is changing (Magurran 1988). In ethnobotany, indicator species are usually those in high demand by resource users and are at risk of over-exploitation and population decline. A question that frequently arises is: how does one objectively select criteria and categories for delimiting high risk species from those that are lower risk? Cunningham (2001) describes some of

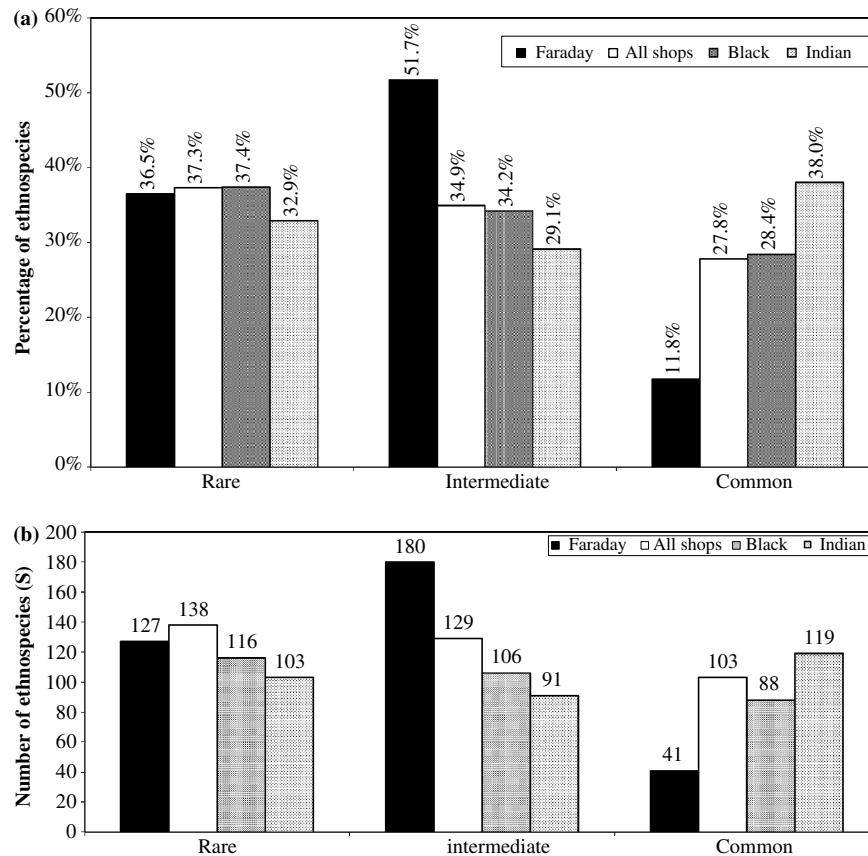


Figure 9. The percentage (a) and number (b) of species of rare, intermediate and common abundances within the Faraday street trader sample and the muti shop survey of 1994. The categories are derived from Hill's numbers N_0 , N_1 and N_∞ .

the categories used for choosing priority species for monitoring as “filters” which help to sift out species that are likely to be more vulnerable to over-harvesting. While complex and comprehensive models and methods exist for ‘filtering’ species (the authors are currently addressing this question in more detail in forthcoming papers), a simple (albeit crude) method obtained from Hill's numbers can be used as a first step in the process of prioritising species for monitoring.

Hill's numbers N_1 , N_2 and N_∞ measure the number of *abundant*, *very abundant* and *most abundant* species in a sample, respectively. In addition, N_0 equals S – the total number of species. The number of species of *rare*, *intermediate* and *common* abundances within a sample were defined in the following way: common $\approx N_\infty$; intermediate $\approx N_1 - N_\infty$; and rare $\approx N_0 - N_1$. ‘Rare’ species, as previously mentioned, are those species with low incidences/abun-

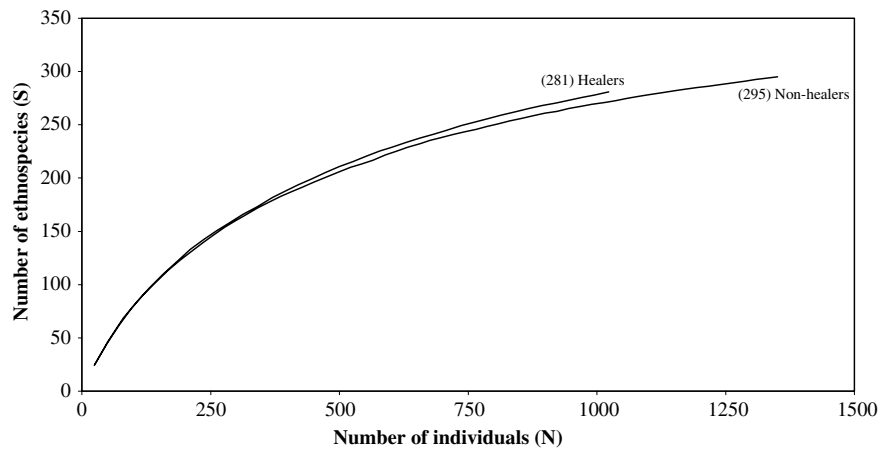


Figure 10. Species accumulation curves for ethnospecies sold by healer and non-healer traders in the Faraday street market.

dances in the sample (i.e., low values of n) and are not necessarily endangered. Rare species are, therefore, the remaining species in a sample that are not abundant (N_1). Because N_1 and N_2 represent species present in abundance in a sample (but not of 'most' abundance like N_∞), they were combined to produce the category 'intermediate'.

Figure 9 shows the number and percentage of ethnospecies categorised as rare, intermediate and common in abundance from the Faraday street market survey and the *muti* shop survey of 1994. The purpose of this paper, is not to discuss what those species actually are, but to derive a method to more easily delimit the species based on their abundance within a data matrix. The histograms in Figure 9 corroborate the results of the evenness and diversity measures, i.e., there is a higher evenness in the representation of species in the *muti* shops. Additionally, there are fewer common or abundant species sold by the street traders compared to the *muti* shops and, therefore greater dominance of a few species in samples of the former. Further, street traders sell a smaller range of species that are more prevalent in the market than other species. It is known from research conducted in the market (Williams 2003) that many of these species are currently threatened or have the potential to be so if current harvesting and utilisation trends continue.

An example of intra-sample diversity: healers and non-healer traders from the faraday market

The main use of diversity measures within this paper has been to compare inter- and intra-sample diversity and sampling effort with respect to traders of different ethnic groups selling plants within formal and informal markets. A

further example of the application of diversity indices to intra-sample differences is the comparison of plants sold by traders that are traditional healers and traders that are *not* traditional healers within the Faraday market. The data matrix for Faraday was subdivided into 'healers' and 'non-healers' and analysed accordingly. The results are shown in Table 5. The results make it clear as to why it is necessary to describe a sample in terms of more than one index of species richness, evenness and diversity as well as the necessity for a species accumulation curve (Figure 10).

Numerical species richness of the plants sold by healers is lower than non-healers (Table 5) – however, this is a function of the smaller subsample size. The curve for healers lies above the curve for non-healers (Figure 10), and therefore the subsample is comparatively more species rich. Additionally, Margalef's index for species richness indicates that species richness is almost the same for both subsamples, with the fractionally higher value for non-healers being the result of a larger subsample. Shannon's index of diversity is exactly the same for both subsamples, thereby underlining the importance of using additional indices to discriminate between sample diversity where sample sizes differ. Simpson's index and Fisher's α show plant diversity sold by the healers to be higher, and therefore there is lower dominance of species sold by the healers. Additionally, the probability of encountering the same species amongst the non-healer traders is higher – the low evenness values for the non-healer subsample corroborate this evidence. The greater dominance of a few species in the non-healer subsample can be accounted for in the following way: unlike healers, non-healers do not supplement their trading incomes with paid consultations by patients. With the high level of competition in the market, the non-healer traders cannot afford to keep too many species that have intermediate demand and commercial value (wholesale), and they therefore, tend to keep more of the species known to be in demand by customers. The plant knowledge traditional healers have, by contrast, allows more flexibility in the range of plants sold – some of which are added to mixtures and preparations sold to patients.

Conclusions

The true value of diversity measures will be determined by whether or not they are empirically useful (Magurran 1988). In non-ethnobotanical studies there are two main areas in which diversity measures have potential application, namely: conservation management, which concentrates almost exclusively on measures of species richness (underpinned by the idea that species-rich communities are better than species-poor ones); and, environmental monitoring, which makes extensive use of diversity indices and species abundance distributions (where, for example, the adverse affects of pollution will be reflected in a reduction in diversity or a change in the shape of the species abundance distribution) (Magurran 1988).

According to Begossi (1996), diversity measures can be used to evaluate the intensity of resources used by human populations, to allow comparisons among different populations in different environments, and to allow evaluations of sampling effort. Further, Begossi (1996) used diversity indices to help answer the following questions: does the diversity of plant use in an area represent the floristic abundance available?; are the same plants used by most individuals?; and, are there differences in the diversity of plant uses per category of user (e.g., gender/age)?

Diversity indices, in the broad sense of the term, therefore, have a useful role to play in the quantitative analysis of ethnobotanical data. Which measures are 'best' cannot be decided without first knowing why diversity should be of interest (Pielou 1975) or how it can help appraise the availability of plants used and/or traded commercially. The primary goal of this paper was to evaluate the performance of a large variety of indices in relation to samples of different sizes and trader profiles, and to examine the kind of useful information they provided. The primary criterion used for recommending certain measures is the value and economy that can be added to the description of plant availability/use, and therefore the degree of insight that can be acquired into interpreting relative abundances. Second, indices are recommended based on ease of calculation and the extent of use by other researchers so that comparisons may be made with other data sets similarly analysed (e.g., Begossi 1996; Hanazaki et al. 2000). The relative merits and shortcomings of some diversity measures have been previously described in Magurran (1988), and an awareness of their limitations is necessary (Ludwig and Reynolds 1988). Ultimately, the choice of index depends on the requirements of the researcher and the value that the index adds to the quantitative description and understanding of the resources under investigation.

A single index of diversity will most often not be sufficient to describe inter- and intra-sample diversity (Hayek and Buzas 1997). Additionally, to describe a sample only in terms of its diversity/heterogeneity index is to confound the two factors of species richness and evenness (Pielou 1975). It is, therefore, judicious to describe a sample in terms of richness, evenness *and* diversity. To this end, we recommend the use of the following measures: (1) species richness (S or N_0); (2) Margalef's index; (3) Shannon index (H'); (4) Simpson's index (both λ and $-\ln \lambda$); (5) Hill's diversity numbers N_1 , N_2 and N_∞ ; (6) Fisher's α ; and (7) evenness indices E_1 and E_5 . It is additionally essential that these indices are graphed as diversity accumulation curves so that the performance of an index may be comprehensively evaluated, and the minimum viable sample size can be determined. According to Gotelli and Colwell (2001), comparing richness without reference to a taxon sampling curve is problematic and graphing the results is necessary to detect differences in measured species richness (and diversity) related to the relative abundance shown in the species accumulation or species diversity curves. We recommend the use of *EstimateS* (Colwell 2001) as a basis for calculating the accumulation curves and the input values necessary for computing most of the other indices. Hill's numbers have an

additional beneficial use in the delimitation of species that are rare, intermediate and common in abundance within a sample, and this is a crucial first step in prioritising species for monitoring and/or remedial conservation action.

The second objective of the paper was to assess whether the survey sites were adequately sampled, and determine the minimum viable sample size on which a diversity measure should be based for the type of data collected. Rarefaction is a commonly used method for estimating species richness, and can be applied to evaluating sampling effort (Magurran 1988; Begossi 1996; Williams et al. 2000; Gotelli and Colwell 2001). However, the size of the sampling unit should be chosen according to factors other than richness because relative abundance affects the performance of the indices. When evaluating species diversity measures (including richness and evenness) a sample size of at least 20 *muti* shops and 35–40 street traders is necessary for the formal and informal sectors in the Witwatersrand, respectively, (Table 6). The actual number of traders surveyed depends on what aspect of diversity is being measured (Table 6). One reason for the necessity to sample more street traders than *muti* shops is because of the lower mean number of species sold per street trader, therefore, requiring additional sampling effort to increase the number of individuals

Table 5. Comparisons of selected measures of diversity between healer and non-healer traders in the Faraday street market.

Index/Measure	Healer traders ($n = 39$) $N = 1023$	Non-healer traders ($n = 60$) $N = 1351$
Species richness (S/N0)	281	295
Margalef	40.4	40.8
Shannon (H')	5.25	5.25
Simpson (λ)	0.0062	0.0067
Simpson ($-\ln \lambda$)	5.08	5.00
Fisher's α	127.9	116.4
Hill's N1	190.6	190.6
Hill's N2	161.1	148.6
Hill's N_∞	40.9	39.7
Evenness E1	0.931	0.923
Evenness E2	0.845	0.779

Table 6. The minimum viable sample size on which a species diversity measure should be based.

Index/Measure	<i>Muti</i> shops	Street traders
Species richness	At least 20	At least 40
<i>Diversity indices</i>		
Information theory (e.g., Shannon)	15–25	25–30
Dominance indices (e.g., Simpson)	20+	15–20
Hill's numbers	20+	30+
Fisher's alpha	30	35
Evenness	17	22
Summary: minimum viable sample size	20–30	35–40

recorded for the sampling curves to reach a horizontal asymptote. Overall, however, sampling effort was found to be more than adequate.

The third objective was to compare the species diversity of sites within the formal and informal sector, and thereby appraise plant availability within the ethnomedicinal trade in the region. As a result, inter- and intra-sample similarities and differences in the sale of plants were identified. Numerical species richness was found to be higher for *muti* shops than street traders, despite the smaller sample size (n). This is related to the large number of individuals (N) and ethnospecies sold per shop trader. Most of the diversity accumulation curves for the indices showed a distinction in the availability and relative abundance of plants sold by street traders compared to shop traders. Different trading factors, therefore, operate within the trading sectors to determine the plant diversity for sale. The graphs of the diversity curves are, therefore, essential for interpreting the different mechanisms operating within the different markets. The average degree of uncertainty in predicting the identity of species sold by the traders is higher in the formal sector, therefore, diversity is higher and dominance of a few species is lower. The higher dominance of certain plants sold by the street traders is confirmed by the lower evenness values of the samples. Intra-sample differences in the *muti* shops showed Indian traders to sell a larger diversity of plants compared to the Black traders, and therefore dominance of plants in the latter was higher. In general, all the indices gave different values but consistent rankings for the different sites.

The high diversity of plants sold within both the formal and informal sector in the study area is likely to be related to a number of factors. Cities (like Johannesburg) are more likely to have more culturally diverse populations, drawn in from many rural communities (Cunningham 2001). Diversity of species sold increases with increasing size of the marketing area, and therefore more species are sold in regional markets (such as the Witwatersrand), fewer in central markets and still fewer in minor or local markets (Cunningham 2001). Begossi (1996) suggests that local resistance to Western medicine may result in a greater demand for traditional medicines, thereby increasing the diversity of plants used. At least 12–15 million people are estimated to consult traditional healers in South Africa annually, and urbanisation has not precluded the use of traditional medicine. In one ‘township’ southwest of Johannesburg (Soweto), there were estimated to be at least 18,000 traditional healers. The Faraday Street market functions primarily as a wholesale market to the traditional healers in townships in the region (Williams 2003). The high diversity of traditional medicines sold in the region is, therefore, indicative of the high demand and the acceptability of traditional healing practices – which to some extent is related to the affordability of primary health care.

Ecologists have long known of species richness, diversity and evenness, and it is only recently that these measures have been applied to the quantitative analysis of ethnobotanical data (e.g., Begossi 1996; Williams et al. 2000). The methods add greater depth to the exploration and understanding of mechanisms and patterns operating within the field of indigenous plant use and trade.

While a quantitative approach to analysing ethnobotanical data might not always be possible, the approach is highly recommended.

References

- Begossi A. 1996. Use of ecological methods in ethnobotany. *Econ. Bot.* 50: 280–289.
- Botha J. 2001. Perceptions of Availability and Values of Medicinal Plants Traded on the Western Boundary of the Kruger National Park, South Africa. MSc dissertation, University of the Witwatersrand, Johannesburg.
- Botha J., Witkowski E.T.F. and Shackleton C.M. 2001. An inventory of medicinal plants traded on the western boundary of the Kruger National Park, South Africa. *Koedoe* 44(2): 7–46.
- Brower J.E. and Zar J.H. 1977. *Field and Laboratory Methods for General Ecology*. Wm. C. Brown Publishers, Dubuque, Iowa.
- Colwell R.K. and Coddington J.A. 1995. Estimating terrestrial biodiversity through extrapolation. In: Hawksworth D.L. (ed.), *Biodiversity Measurement and Estimation*. Chapman and Hall, pp. 101–118.
- Colwell R.K. 2001. EstimateS: Statistical estimation of species richness and shared species from samples. Version 6, User's guide and application published at: <http://viceroy.eeb.uconn.edu/estimates>.
- Cunningham A.B. 2001. *Applied ethnobotany: people, wild plant use and conservation*. People and Plants Conservation Manual, Earthscan.
- Dauskardt R.P.A. 1990. The changing geography of traditional medicine: urban herbalism on the Witwatersrand, Johannesburg. *GeoJournal* 22(3): 275–283.
- Dauskardt R.P.A. 1991. 'Urban herbalism': the restructuring of informal survival in Johannesburg. In: Preston-Whyte E. and Rogerson C. (eds), *South Africa's Informal Economy*. Oxford University Press, Cape Town, pp. 87–100.
- Gotelli N.J. and Colwell R.K. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* 4: 379–391.
- Hanazaki N., Tamashiro J.Y., Leitão-Filho H.F. and Begossi A. 2000. Diversity of plant use in two Caçara communities from the Atlantic Forest coast, Brazil. *Biodivers. Conserv.* 9: 597–615.
- Hayek L.C. and Buzas M.A. 1997. *Surveying Natural Populations*. Columbia University Press, New York.
- Höft M., Barik S.K. and Lykke A.M. 1999. Quantitative ethnobotany: applications of multivariate and statistical analysis in ethnobotany. People and Plants working paper 6. UNESCO, Paris.
- Hill M.O. 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology* 54(2): 427–432.
- Johns T., Mhoro E.B., Sanaya P. and Kimanani E.K. 1994. Herbal remedies of the Batemi of Ngorongoro District, Tanzania: a quantitative appraisal. *Econ. Bot.* 48: 90–95.
- Krebs C.J. 1989. *Ecological Methodology*. Harper and Row Publications, New York.
- Lowrey T.K. and Wright S., eds. 1987. *The Flora of the Witwatersrand, Volume I, The Monocotyledonae*. Wits University Press, Johannesburg.
- Ludwig J.A. and Reynolds J.F. 1988. *Statistical Ecology – a Primer on Methods and Computing*. John Wiley and Sons, Toronto.
- Luoga E.J., Witkowski E.T.F. and Balkwill K. 2000. Subsistence use of wood products and shifting cultivation within miombo woodland of eastern Tanzania, with some notes on commercial uses. *S. Afr. J. Bot.* 66(1): 72–85.
- Magurran A. 1988. *Ecological Diversity and its Measurement*. Princeton University Press, Princeton.
- Murali K.S., Shankar U., Shaanker R.U., Ganeshiah K.N. and Bawa K.S. 1996. Extraction of non-timber forest products in the forests of Biligiri Rangan Hills, India. 2. Impact of NTFP extraction on regeneration, population structure, and species composition. *Econ. Bot.* 50(3): 252–269.

- Phillips O. and Gentry A.H. 1993a. The useful plants of Tambopata, Peru: I. Statistical hypotheses tests with a new quantitative technique. *Econ. Bot.* 47: 15–32.
- Phillips O. and Gentry A.H. 1993b. The useful plants of Tambopata, Peru: II. Additional hypothesis testing in quantitative ethnobotany. *Econ. Bot.* 47: 33–43.
- Pielou E.C. 1975. *Ecological Diversity*. John Wiley and Sons, New York.
- Prance G.T., Balée W., Boom B.M. and Carneiro R.L. 1987. Quantitative ethnobotany and the case for conservation in Amazonia. *Conserv. Biol.* 1: 296–310.
- Williams V.L., Balkwill K. and Witkowski E.T.F. 1997. Muthi traders on the Witwatersrand, South Africa – an urban mosaic. *S. Afr. J. Bot.* 63(6): 378–381.
- Williams V.L., Balkwill K. and Witkowski E.T.F. 2000. Unravelling the commercial market for medicinal plants and plant parts on the Witwatersrand, South Africa. *Econ. Bot.* 54(3): 310–327.
- Williams V.L., Balkwill K. and Witkowski E.T.F. 2001. A lexicon of plants traded in the Witwatersrand umuthi shops, South Africa. *Bothalia* 31(1): 71–98.
- Williams V.L. 1992. An investigation of the herbal medicine or ‘muti’ trade on Witwatersrand. BSc (Hons) Dissertation, Department of Geography and Environmental Sciences, University of the Witwatersrand, Johannesburg.
- Williams V.L. 2003. Hawkers of health: an investigation of the Faraday Street traditional medicine market in Johannesburg, Gauteng. Report to the Gauteng Directorate of Nature Conservation, DACEL, Johannesburg.
- Wong J., Thornber K. and Baker N. 2001. Resource assessment of non-wood forest products: experience and biometric principles. NWFP Series 13, FAO, Rome.
- Zar J.H. 1984. *Biostatistical Analysis*. Prentice-Hall, Inc., Englewood Cliffs.

The use of incidence-based species richness estimators, species accumulation curves and similarity measures to appraise ethnobotanical inventories from South Africa

Vivienne L. Williams · Edward T. F. Witkowski · Kevin Balkwill

Received: 2 December 2003 / Accepted: 20 February 2006
© Springer Science+Business Media B.V. 2006

Abstract The incorporation of suitable quantitative methods into ethnobotanical studies enhances the value of the research and the interpretation of the results. Prediction of sample species richness and the use of species accumulation functions have been addressed little in applied ethnobotany. In this paper, incidence-based species richness estimators, species accumulation curves and similarity measures are used to compare and predict species richness, evaluate sampling effort and compare the similarity of species inventories for ethnobotanical data sets derived from the trade in traditional medicine in Johannesburg and Mpumalanga, South Africa. *EstimateS* was used to compute estimators of species richness (e.g. Jackknife), rarefaction curves, species accumulation curves and complementarity. Results showed that while the Michaelis–Menten Means estimator appeared to be the best estimator because the curve approached a horizontal asymptote, it was not able to accurately predict species richness for one of the data sets when two of its subsamples were individually tested. Instead, the first-order Jackknife estimator best approximated the known richness.

Keywords Traditional medicine trade · Quantitative ethnobotany · Species accumulation curves · Rarefaction · Richness estimation · Complementarity · Sampling effort · *EstimateS*

Introduction

Species richness, or the number of species in a sample of a specified size, is an instantly comprehensible expression of species diversity (Magurran 1988). Biologists often need to know how many species there are within an area or, they need to compare the number of species found in different areas (Colwell et al. 2004).

V. L. Williams (✉) · E. T. F. Witkowski · K. Balkwill
School of Animal, Plant and Environmental Sciences, University of the Witwatersrand,
Private Bag 3, Wits 2050, South Africa
e-mail: vivwill@planetac.co.za

Complete enumeration of species richness within an extensive study area is, however, generally not feasible and consequently a number of different methods for estimating total species richness from a sample have been devised (Chiarucci et al. 2003). Traditionally, species accumulation functions have been used for plotting species versus sampling effort to estimate visually whether a curve has asymptotically approached some ceiling (and hence whether sampling has been sufficient), as well as estimate the total richness of a site (Soberón and Llorente 1993; Gotelli and Entsminger 2001). Colwell and Coddington (1995), Gotelli and Colwell (2001) and Colwell et al. (2004) have explored the measurement, estimation and comparison of species richness through extrapolation and the use of taxon sampling curves, including assigning confidence intervals to the estimates. While the methods for estimating species richness have been applied to the approximate assessment of species richness in faunal and floral groupings at comparative sites (e.g. by Bunge and Fitzpatrick 1993; Soberón and Llorente 1993; Colwell and Coddington 1995; Toti et al. 2000; Gotelli and Colwell 2001; Chiarucci et al. 2003; Sørensen 2004), the benefits of such an approach to applied quantitative ethnobotany have yet to be fully explored.

Ethnobotanical surveys are often the first step in identifying useful species that are a resource management priority (Cunningham 2001). A challenge for ethnobotanical studies has been in knowing how complete the plant inventories are, i.e. how many more species would have been documented with increased sampling effort? Furthermore, how does a researcher objectively decide on the level of sampling effort required to satisfactorily complete a survey or draw conclusions on the adequacy of the effort already executed? In two of the few studies using species accumulation functions for ethnobotanical data, Begossi (1996) and Hanazaki et al. (2000) used rarefaction curves to evaluate sampling effort and explore differences in plant use per category of user (e.g. age and gender) within different communities, mainly in South America. These methods have the potential to be broadened to other ethnobotanical studies and resource inventories to include comparisons of species richness between sites, as well as estimate the number of species expected had a complete census of all the plants used/traded been possible. Plant inventories for ethnobotanical studies are usually incidence-based (i.e. presence or absence). The aim of this paper is to assess the usefulness of incidence-based species accumulation curves, species richness estimators and similarity measures with data collected from the medicinal plant trade in the Witwatersrand region (including Johannesburg), South Africa.

Species accumulation curves

A *species accumulation curve* is the graph of the cumulative number of observed species as a function of some measure of sampling effort (Colwell et al. 2004). If the species are randomly and sequentially recorded one after another within a sampling area, then the resulting accumulation curves are *individual-based*. If, however, the survey area is subdivided into smaller sampling units (such as quadrats) and the total number of species is accumulated as a result of successively sampling additional quadrats, then the accumulation curves are *sample-based* (Gotelli and Colwell 2001). Sequential or successive pooling of the individuals or samples will not, however, produce a smoothed curve and the shape very much depends on the sample order.

Rarefaction is used to produce a smoothed curve that is the statistical expectation of the corresponding accumulation curve (Gotelli and Colwell 2001; Colwell et al. 2004). Rarefying a sample estimates its expected species richness at different values of n samples or N individuals from the pooled total species richness after randomizing the sample order. Rarefaction curves can be individual- or sample-based, depending on the sampling methodology (Colwell 2001). In ethnobotany, the quadrat as a sampling unit would be replaced with a more appropriate sampling entity such as herbalists' shop, street traders' stall, or an informant within a community. Comparisons of the species accumulation curves with the rarefaction curves can be used to assess the homogeneity of the samples. In addition, the 95% confidence intervals of the rarefaction curves calculated by *EstimateS* (Colwell 2005) can be used to determine whether species richness for data sets is significantly different.

Estimating species richness

There are three approaches to estimate species richness from samples, namely (1) fitting parametric models of relative abundance (e.g. log normal distribution), and using the shape of the species abundance distribution to deduce the total species richness, (2) non-parametric estimators, and (3) extrapolation of species accumulation or species-area curves (Magurran 2004).

Parametric methods

Two parametric models of relative abundance with the greatest potential for estimating species richness are the log normal and log series distributions (Magurran 2004). When species fit a log normal model, the relative abundance of the data takes the form of a bell-shaped normal distribution (Krebs 1989). It is possible to estimate the theoretical number of species (S^*) in the community/assembly by extrapolating the bell-shaped curve beyond the class of minimal abundance (i.e. over the 'hidden' portion of the curve to the left of the 'veil line') and measuring the area under the curve (Krebs 1989; Colwell and Coddington 1995). Most log normal distributions are, however, of the truncated variety (Magurran 1988, 2004), i.e. part of the bell-shaped curve is obscured, and the truncation point is at the veil line. Pielou (1975) devised a method for fitting a truncated log normal distribution that assumes that the position of the veil line can be recognized, and that the area under the remaining part of the curve can be used to estimate S^* (Magurran 1988).

Data that fit a log series distribution show a 'hollow' curve. In other words, comparatively few species are common and comparatively large numbers of species are rare (or, infrequently recorded). Additionally, the modal class is always the singletons no matter how large the sample (Colwell and Coddington 1995). The estimate of species richness is always non-asymptotic; hence, the number of estimated species will increase as the number of individuals (N) increases (Magurran 2004).

Non-parametric methods

Non-parametric methods for estimating species richness from samples were adapted from mark-recapture applications for estimating population size. They also

require no assumptions about community structure (Colwell and Coddington 1995). *EstimateS* by Colwell (2005) computes seven non-parametric estimators of species richness. The estimators are homologous in that richness is estimated from the preponderance of infrequently occurring species in each sample, i.e. the higher the proportional abundance of rare species, the greater the probability of encountering more new species with increased sampling effort. The method for determining the best estimator involves visually estimating whether a plot of the estimated number of species for each successively accumulated sampling unit reaches a horizontal asymptote and reduces bias, following which conclusions are drawn regarding the adequacy of the size of the data set and the performance of the estimator. Sampling, or a census, is essentially complete if all species are observed “multiple” times or are “not rare” (Colwell and Coddington 1995). It is essential to note that richness estimators do not extrapolate beyond the last sample to estimate richness at an asymptote (Soberón and Llorente 1993). Instead, richness estimators predict richness, including species not discovered in the sample, from the proportional abundances of species within the total sample (Soberón and Llorente 1993; Colwell 2005).

Extrapolating species accumulation curves

Species accumulation curves illustrate the rate at which new species are found, but the total species richness of a site will not be directly revealed by the curves if sampling is not thorough (Magurran 2004). If sampling is incomplete, then the curves will show an upward trend illustrating the extent to which sampling effort might have to be increased to accumulate more species. The curves can also be used as a means of estimating species richness, most commonly by fitting functions such as the asymptotic Michaelis-Menten algorithm (Colwell et al. 2004), or non-asymptotic estimators such as log-linear models (Colwell and Coddington 1995) (not investigated in this paper because the non-parametric estimators are reported to perform better). Extrapolation of the curves is also possible, but there is no firm consensus on how to do so (Magurran 2004). Colwell et al. (2004) have explored the challenges of extrapolating species accumulation curves to estimate species richness, and the means to do so.

Complimentarity and similarity

Complimentarity is the measure of biotic distinctness or dissimilarity between species inventories that were compiled at sites located randomly within a habitat mosaic or along a gradient (Colwell and Coddington 1995). The greater the degree of distinctiveness, the more complimentary the two sites are, and hence the higher the combined species richness is likely to be between the sites. The measure captures the sense that species richness can be interpolated between site gradients depending on the degree of distinctiveness. Undersampled sites can, however, result in complimentarity and species richness being overestimated because the distinctness and the number of species shared between samples would be lower. In ethnobotanical studies, complimentarity can be used to measure the dissimilarity of species inventories between different markets and/or user groups and thereafter be used to interpolate the number of species that might be traded or in use between certain

sites or user groups. By contrast, *similarity* is the degree to which the species composition of samples is alike. The most widely adopted measures are the classic Jaccard and Sørensen indices for incidence-based data, and the Bray–Curtis (equal to the ‘Sørensen quantitative’ index) and Morisita–Horn indices for abundance-based data (Magurran 1988; Colwell 2005).

Study area

This case study is focused on the Gauteng Province of South Africa, and specifically the region called the Witwatersrand, which includes Johannesburg in its centre (Fig. 1). The Witwatersrand is an extensively urbanized axis of approximately 100×40 km, and has a large urban Black population with an active trade in traditional plant medicines. Johannesburg forms the hub of the regional informal plant trade, mainly because of the concentration of the labour force and the strong links of that labour force with rural communities (especially to the province of KwaZulu-Natal to the south-east) and traditional activities, including the utilization of traditional medicine. There were estimated to be 18,000 traditional healers in Soweto, a Johannesburg township, in the mid-1990s.

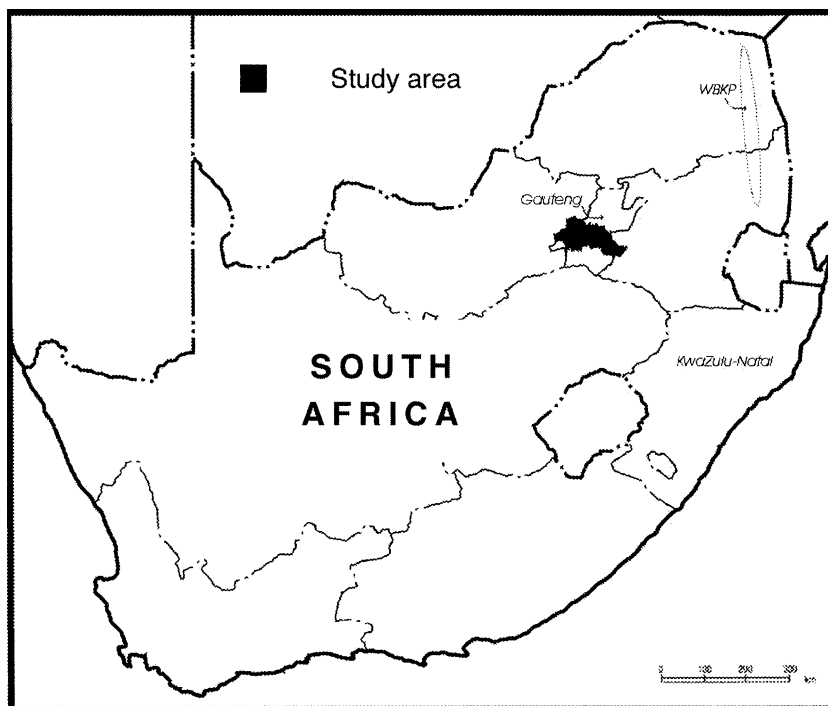


Fig. 1 The Witwatersrand study area (in black) within the province of Gauteng, South Africa. The Witwatersrand includes the metropolis of Johannesburg. One-hundred and fifty traders of traditional medicine were interviewed in the region between 1994 and 2001. ‘WBKP’ represents a region within which a survey of 17 traders was conducted by Botha (2001) on the western boundary of the Kruger National Park. The province of KwaZulu-Natal is the harvesting source of most plants sold in the Witwatersrand

The trade in medicinal plants is divided into two sectors, namely formal business represented by ‘*muti*’ shops or herbal pharmacies, and informal markets represented by street traders. In 1994, there were more than 200 *muti* shops on the Witwatersrand (Williams et al. 1997), the majority of which were owned by Black traditional healers (52%) and Indian merchants (25%). A study conducted in 2001 revealed there to be more than 160 street traders in a market called ‘Faraday’ (Williams 2003). Ninety-seven percent of the street traders were migrants to Gauteng, of which 90% regarded KwaZulu-Natal as ‘home’ (Williams 2003). For the most part, the trade of medicinal plants in Johannesburg meets the needs of the Zulu sector of the population, despite the multicultural nature of the metropolis (Williams et al. 1997). Customers to this market are primarily traditional healers from the region buying plants to restock their home-based healing practices.

Methods

The species accumulation functions were reviewed using seven sets of ethnobotanical data derived from inventories compiled from three surveys of medicinal plants in trade. The studies were conducted with formal *muti* shop traders and informal street traders of traditional medicine. The first survey in 1994 was for 50 *muti* shops on the Witwatersrand (‘*All shops*’) (Williams et al. 2000). This sample was subdivided into three smaller subsamples based on the ethnicity of the shop owner, namely Black ($n = 28$ shops), Indian ($n = 20$) and White ($n = 2$; data set was not included in the analyses due to its size). The second survey was conducted in 2001 with 100 street traders in the Faraday street market in Johannesburg (Williams 2003). The third survey, conducted by Botha (2001) with 17 informal street traders, inventoried plants traded on the western boundary of the Kruger National Park (‘*WBKP*’), Mpumalanga, South Africa. The seventh data set is a combination of all the ethnospecies inventoried during the 1994 and 2001 Witwatersrand studies and is called ‘*WRand Total*’ ($n = 150$). This data set was only used in complementarity and similarity analyses because of the variations in sampling technique and the number of individuals per sample. All inventories recorded the presence of plants in a shop or at a stall, hence the data are incidence-based (qualitative) not abundance-based (quantitative).

‘*Ethnospecies*’ was used as a proxy for species during the analyses (except where otherwise specified). The term ‘ethnospecies’ (Hanazaki et al. 2000) takes into account the folk or common names given to one or several species quoted during resource inventory interviews. The ethnospecies name ‘*iMphepho*’, for example, applies to at least six species of *Helichrysum*, but it was not possible to ascertain which of the six the traders sold when the survey was conducted. Ethnospecies names are best used where uncertainty exists as to the correct identification of a species. Therefore, rather than duplicate the incidence of species occurrences, the data were quantified based on the number and frequency of occurrence of ethnospecies sold by the traders. Where more than one ethnospecies name applied to a single species, only one name was used during the analyses. Later in the paper, the ratio of ethnospecies to number of species is calculated for the specific surveys in order to extrapolate the estimated number of ethnospecies to actual species (e.g. Table 3). In the Faraday study, for example, 100 ethnospecies names \approx 136 plant species.

Public-domain software called *EstimateS* (Version 7.5, Colwell 2005) was used to calculate and evaluate the performance of six species richness estimators appropriate

for incidence-based data, namely: ICE, Chao 2, first order jackknife, second-order jackknife, bootstrap and Michaelis–Menten Means (Table 1). The Michaelis–Menten Runs estimator was considered, but eventually excluded from the analyses because of the very high and erratic richness estimates at the beginning of the curve. *EstimateS* was also used to compute the expected species accumulation curves (sample-based rarefaction) and 95% confidence intervals. The sample order was randomized 100 times to compute the mean estimator and expected species richness for each sample accumulation level. In doing so, a smoothed curve was generated. Species richness was also estimated using the log normal and log series parametric models of relative abundance. The formulae in Magurran (1988), Krebs (1989) and Ludwig and Reynolds (1988) were used to fit the data to the distributions.

Percentage complementarity and the classic Jaccard and Sørensen indices of similarity for incidence-based data were used to compare the distinctiveness of the data sets and calculate the combined species richness for two data sets. Complementarity was calculated using the formula in Colwell and Coddington (1995) and variables computed by *EstimateS*. The Jaccard and Sørensen indices were also computed by *EstimateS*. Complementarity coefficients vary from 0 (least complementary and most identical) to 1 (most complimentary and least identical), whereas similarity measures range from 0 (least similar) to 1 (most similar). Alternatively, the measures are expressed as a percentage.

Results and discussion

Species accumulation and rarefaction curves

Non-randomized species accumulation and sample-based rarefaction curves were plotted for species inventoried in the Faraday Street market and *muti* shops (Fig. 2). The 95% confidence intervals for the rarefaction curves were also plotted. The

Table 1 Brief descriptions of the species richness estimators used that are computed by *EstimateS* (Version 7.5) (From Colwell and Coddington 1995; Colwell 2005)

Name	Richness estimators	Description
<i>Non-parametric estimators</i>		
ICE	Incidence-based coverage estimator	Based on species found in 10 or fewer sampling units
Chao 2	Chao 2 richness estimator	Incidence-based; based on the presence of singletons and doubletons
Jack 1	First-order Jackknife richness estimator	Based on the number of species occurring in only 1 sample
Jack 2	Second-order Jackknife richness estimator	Based on the number of species that occur in only 1 sample, as well as the number that occur in exactly 2 samples
Bootstrap	Bootstrap richness estimator	Based on the proportion of the samples containing each samples
<i>Asymptotic extrapolation of the species accumulation curve</i>		
MMMean	Michaelis–Menten Mean richness estimator	Computes estimates once for each sample pooling level

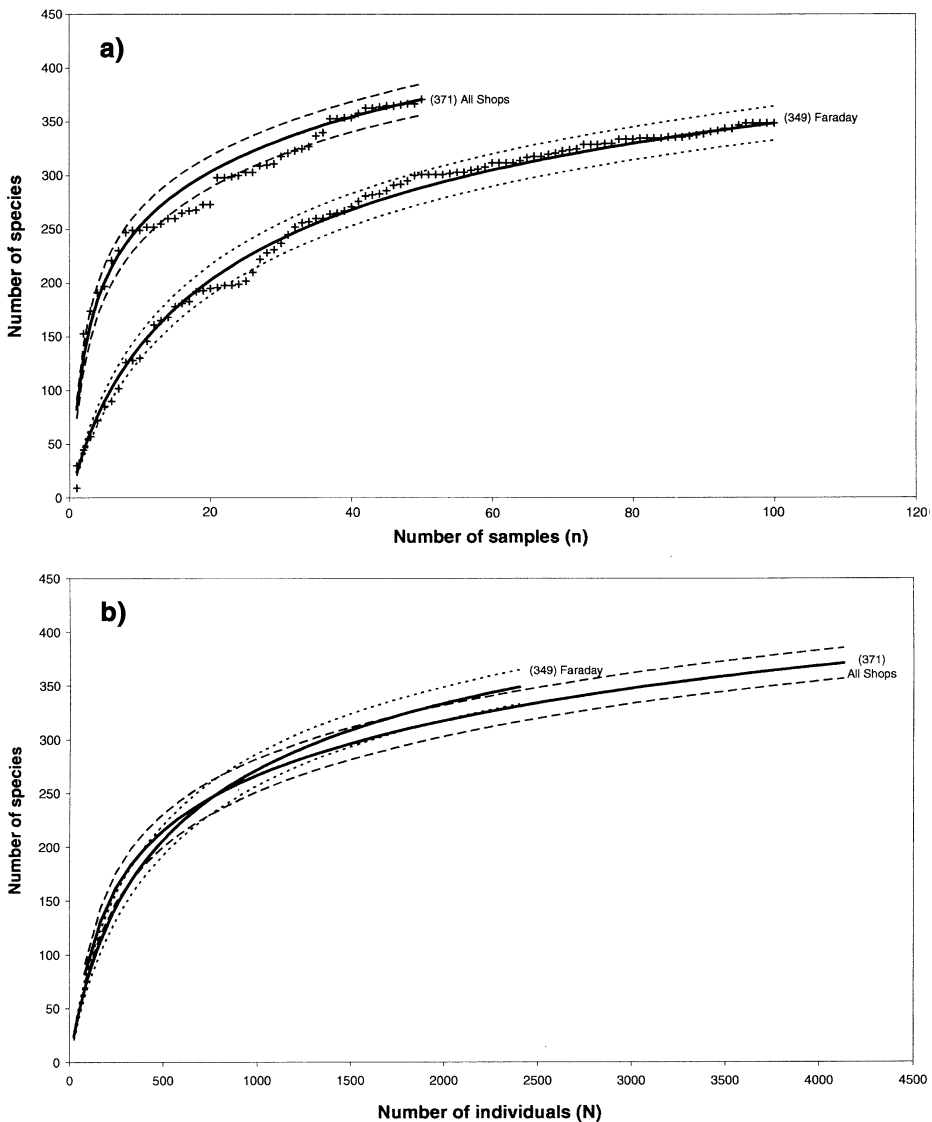


Fig. 2 Comparing sample-based rarefaction curves (solid-lines), 95% confidence intervals (dashed lines) and the non-randomized species accumulation curves (shown as + in 2a) for species recorded in the stalls of 100 street traders in the Faraday market in 2001 and in 50 *muti* shops in 1994. The x-axes are scaled (a) by number of samples, and (b) by number of individuals

x-axes are scaled by accumulated number of samples (i.e. the number of stalls or shops inventoried) (Fig. 2a) and, the number of individual plants accumulated as the samples were pooled (Fig. 2b). The graphs differ in that the number of stalls sampled in Faraday was twice that of the number of shops sampled, however the mean number of ethnospecies sold per shop trader was approximately 3.5 times the number of ethnospecies sold per stall in the Faraday market (means and standard deviation equal 83 ± 46 and 24 ± 12 ethnospecies per trader for *All shops* and

Faraday respectively). Numerical species richness per trader is thus greater for the shop traders, and the initial steep gradient of the curve shows that species are accumulated more rapidly for comparatively smaller sample numbers. For example, the number of ethnospecies expected by the rarefaction curve after 25 samples is 320 for *All shops* and 224 for *Faraday*—a difference of 96 ethnospecies (Fig. 2a). Figure 2a also shows that *muti* shops have consistently higher number of ethnospecies for similar levels of sampling effort compared to stalls at the Faraday market.

In ecology, Fig. 2a would represent a comparison between the densities of species for two data sets (i.e. the number of species per sampling unit), and Fig. 2b a comparison of species richness because the effect of density is removed when the individuals are pooled. When comparing sample-based rarefaction curves and 95% confidence intervals for two data sets derived from tropical rain forest saplings, Colwell et al. (2004) concluded that where the confidence intervals don't overlap, the differences in density and/or species richness are not significant at $P < 0.05$. If one uses this line of reasoning when comparing the differences in the mean number of ethnospecies per sample (Fig. 2a), then the differences are ostensibly significant because the confidence intervals do not overlap for the two data sets. However, when the x -axis is rescaled by individuals (Fig. 2b), then species richness for the two data sets is not significantly different. Gotelli and Colwell (2001) recommend that when comparing and estimating species richness between data sets and sample sites, the sample-based rarefaction curves are rescaled by individuals to adjust for the differing densities of individuals (Colwell et al. 2004).

The order in which the samples (or individuals) are included in a species accumulation curve influences its overall shape (Magurran 2004). Despite the non-randomized species accumulation curve not being smoothed (represented by ++ in Fig. 2a), the samples are relatively homogenous and there is little variation in the number of ethnospecies per sample—as indicated by how closely the species accumulation curve approximates the rarefaction curve (especially for *Faraday* data). The traders therefore consistently sell similar numbers of species.

The ability of a species accumulation curve to reach an asymptote is seen as an indicator of sampling sufficiency (Heck et al. 1975). While none of the curves actually reaches an asymptote, they appear to be approaching one and the rate of accumulation of species as successively more samples are added to the total is decreasing.

Species richness estimates

The procedure for fitting parametric models of relative abundance to a data set involves calculating the number of species expected in an abundance class and comparing it with the number of species actually observed. By summing the number of species expected (S^*), an estimate for species richness is obtained. The *Faraday* data fit both the log normal and truncated log normal models (Table 2), and the expected ethnospecies richness is 367 and 393 respectively. However, the log normal and truncated log normal models do not fit the other data sets and the expected ethnospecies richness for the *Black-owned* shops is unrealistically high (≈ 1016 ethnospecies for the log normal).

The log series model does not fit any of the data sets and there is an underestimation of the expected ethnospecies and species richness (Tables 2 and 3). The log

Table 2 Species richness estimates and other summary values for *ethnospecies* sold by shop traders and traders in the Faraday Street market

	Faraday street market	Muti shops		
		All shops	‘Black-owned’	‘Indian-owned’
No. of samples (<i>n</i>)	100	50	28	20
No. of individuals (<i>N</i>)	2402	4129	1769	2168
Observed <i>ethnospecies</i> richness	349	371	310	312
<i>Estimated ethnospecies richness</i>				
Chao 2	427.4±19.4	436.0±20.0	364.8±15.0	358.4±13.6
ICE	408.6	435.0	368.8	361.2
Jack 1	432.2±10.0	442.5±13.2	379.4±17.5	371.9±16.3
Jack 2	472.8	474.1	406.9	395.3
Bootstrap	388.0	404.4	343.4	340.3
MMMean	394.3	379.2	352.4	333.9
<i>Expected ethnospecies richness</i>				
Log normal	393.2*	698.5	1015.5	437.4
Truncated log normal	367.2*	386.5	320.1	319.9
Log series	(341.8)	(346.7)	(293.4)	(288.1)
No. of singletons	84	73	72	63
No. of doubletons	43	41	44	39

* $P \approx 0.21$

Each richness estimate represents the mean (and SD for some estimators) for 100 randomizations of the sample order (with the exception of the log normal and log series distributions). Number of samples (*n*) is the number of street market stalls or shops inventoried. Number of individuals (*N*) is the total number of individual plants recorded. Note: figures are for numbers of *ethnospecies* (i.e. common names) *not* species. Values in brackets estimated by the log series model are underestimates of the observed richness. Richness estimates for the subset of Black- and Indian-owned shops (shaded area) should ideally approximate the observed *ethnospecies* richness of 371 for the entire ‘All shops’ sample (shaded)

Table 3 Estimated *species* richness estimated based on multiplying the values in Table 2 by the ratio of observed *ethnospecies* to *species*

	Faraday street market	Muti shops		
		All shops	‘Black-owned’	‘Indian-owned’
Observed <i>ethnospecies</i> richness	349	371	310	312
Probable no. of corresponding species	≈475	≈508	≈425	≈427
Ratio of <i>ethnospecies</i> to species	1 to 1.36		1 to 1.37	
<i>Estimated species richness</i>				
Chao 2	581	587	500	491
ICE	556	596	505	495
Jack 1	588	606	520	510
Jack 2	643	650	557	395.3
Bootstrap	528	554	470	542
MMMean	536	520	483	457
Log normal	535	957	1391	599
Truncated log normal	499	530	320.1	438
Log series	(465)	(475)	(439)	(395)

For example, the 349 *ethnospecies* inventoried at the Faraday market corresponds with ≈475 species i.e. 1:1.36

series model, if it fits the data well, is reported to accurately predict the number of new species likely to be found in larger samples (Colwell and Coddington 1995). However, species richness will be underestimated if the data do not conform to the log series distribution (Magurran 2004). Furthermore, data that fit a log series have an implied low evenness because of the presumption that there is a large proportion of ‘rare’ species (i.e. singletons) and a small proportion of abundant species in the sample (Magurran 1988). Plant data frequencies recorded during the two surveys show the evenness to be very high (in the region of 0.91 to 0.99 for the Shannon J') (Williams et al. 2005). Therefore, these data won’t fit the log series distribution and the number of species not recorded during the sampling of the shops and stalls cannot be realistically predicted. An alternative is to use non-parametric methods to estimate species richness.

The best way to test the performance of a species richness estimator is to use data sets from sites where the actual species richness is known (Toti et al. 2000). Unfortunately, we cannot always use this approach successfully in ethnobotanical studies of large, regional ethnomedicinal markets such as Johannesburg because the plants sold in the markets originate from a wide geographical area of high species richness. Some plants sold in the Johannesburg markets are known to have been harvested in Zambia, Malawi and Mozambique. Furthermore, species entering commercial trade represent a ‘short-list’ of a far greater diversity of species used in rural areas (Cunningham 2001). The total number of species used for traditional medicine is thus much greater than a survey of a commercial market would ever capture.

There are, however, other ways to assess the usefulness of an estimator (Colwell and Coddington 1995; Toti et al. 2000). First, is to observe how rapidly the estimation curves approach an asymptote as sample size increases. Second, is to look for consensus among the estimators. And third, to compare the estimator curves to subjective visual extrapolations of the possible asymptotes of an observed species richness accumulation curve. A good estimator (1) should reach, or closely approach, a stable asymptote sooner (i.e. with fewer samples) than the observed species accumulation curve would, (2) yield an estimate that is not vastly different from the other estimators, and (3) should give estimates that are close to reasonable visual extrapolations of the asymptote of the observed species accumulation curve (Toti et al. 2000). However, not all curves will reach an asymptote, and if part of the reason for estimating species richness is to know the degree to which a site is ‘un-sampled’ and hence how many species remain to be discovered with further sampling, then another method for checking the efficiency of an estimator is to apply the ‘stopping rule’ i.e. the point beyond which further sampling is no longer necessary or too costly (Magurran 2004). Asymptotes are generally used to establish this point, but Magurran (2004) also proposes that a sample is subdivided into two parts and that the richness of each is estimated separately. If the estimates give answers that are consistent with the one obtained for the combined sample, then ample data have been collected (Magurran 2004). In the same way, the estimator that best predicts the species richness of the total sample from the subsamples can be adjudicated as the better estimator.

The Michaelis–Menton Mean (MMMean) estimator appears to be the only estimator that approaches an asymptote sooner than the expected accumulation curve, for all data sets (Fig. 3a–d). This result is probably to be expected given the asymptotic nature of the estimator (Magurran 2004). The remainder of the

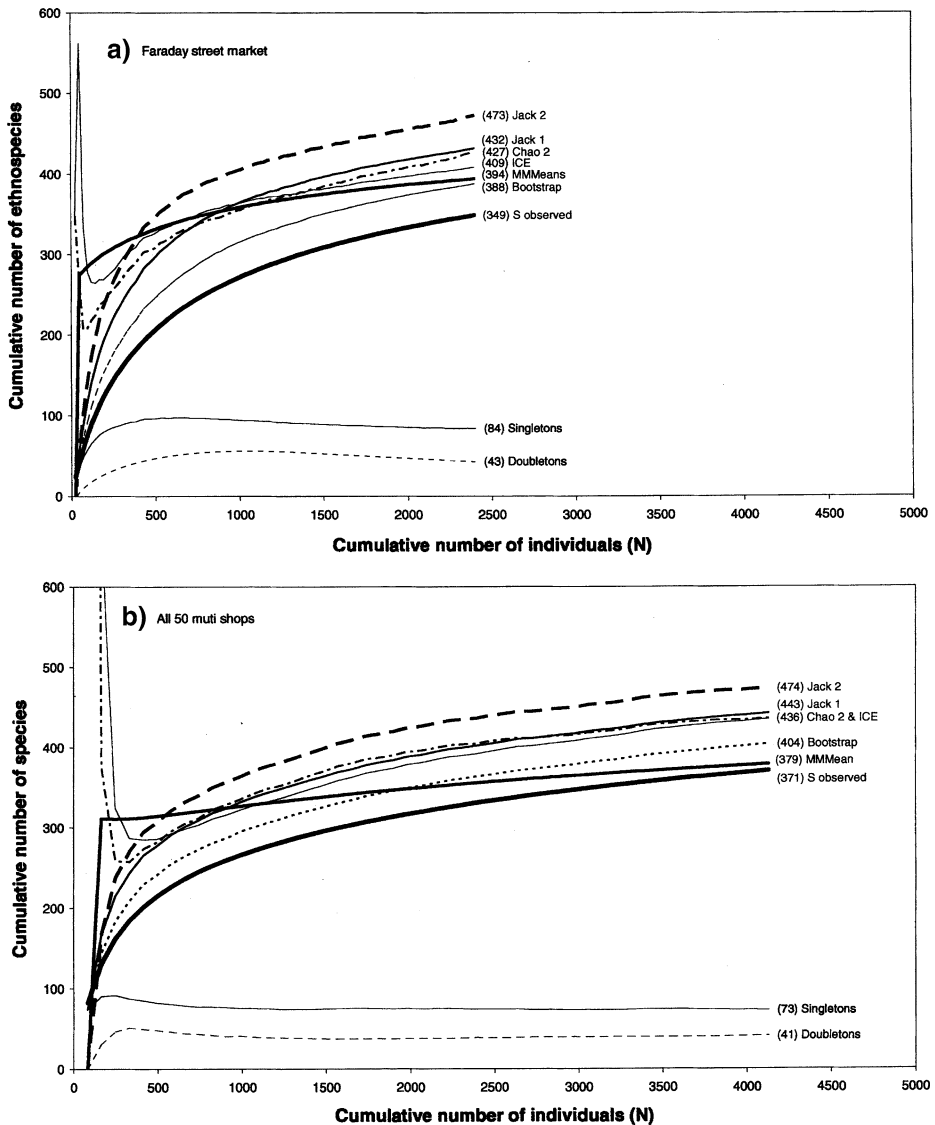


Fig. 3 The performance of the incidence-based species richness estimators compared with the expected species accumulation curve (S observed) for four data sets from the Faraday street market (a) and multi shops (b–d). The cumulative number of singletons and doubletons are also plotted. The curves above the species accumulation curve show the estimated species richness (final value in brackets) based on successively larger numbers of samples. The estimators used are Jack 1 & 2, Chao 2, ICE, MMMean and Bootstrap. The values were generated by *EstimateS* (Colwell 2005). For all curves, each point is the mean of 100 estimates based on 100 randomizations of the sample accumulation order. All graphs are scaled to the same axis values

estimators tend to rise in parallel with the accumulation curve and do not reach an asymptote. The first- and second-order Jackknife estimators climb more steeply than the other estimators, and Chao 2 and ICE over-estimate species richness at the beginning of the sample (because they are strongly affected by the high number of

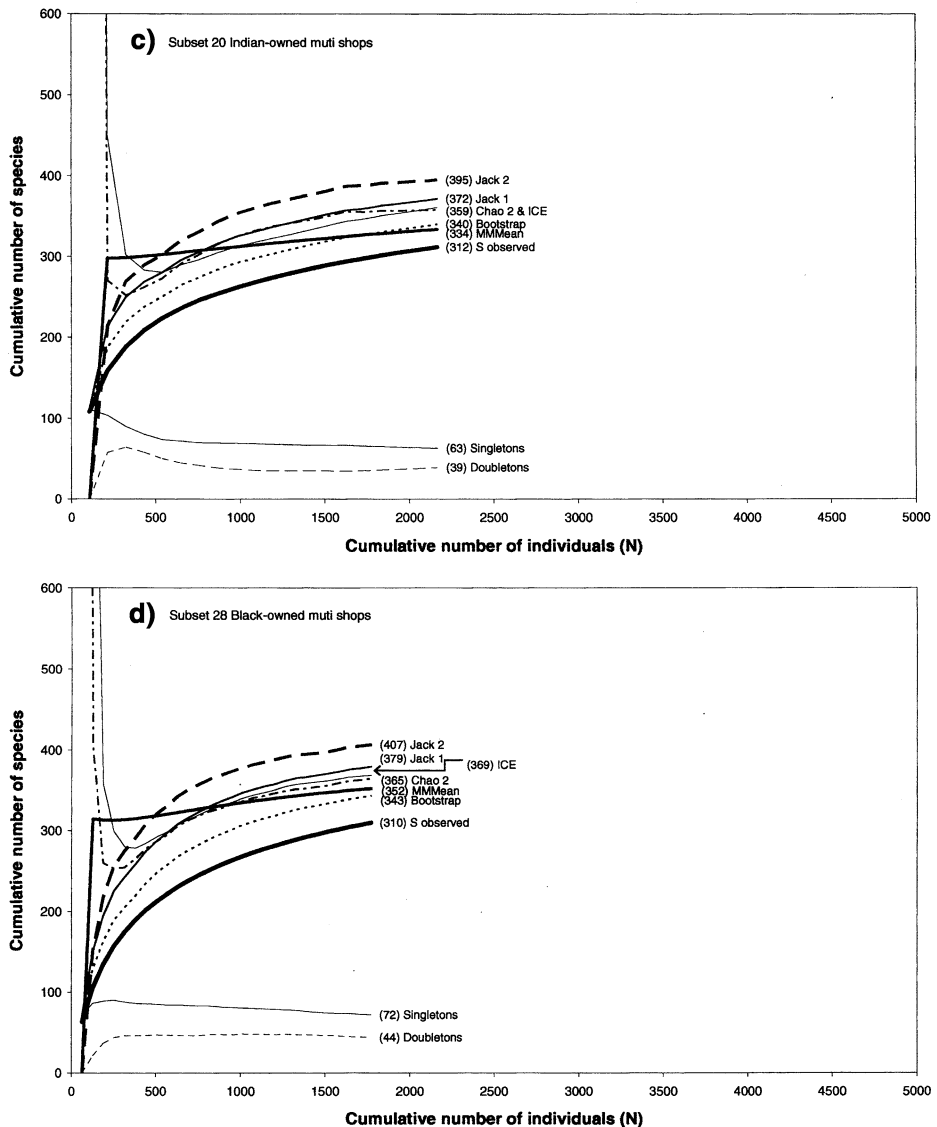


Fig. 3 continued

singletons and rare species present in the first few samples). Plots of singletons and doubletons rise quickly and then tend to level off and decrease very slightly as the sample size increases.

The difference between the highest and lowest ethnospecies estimate is 84.9 and 94.9 for *Faraday* and *All shops* respectively, and between 54.5 and 61.4 for the *Black-* and *Indian-owned* shops subset respectively (Table 2). Interestingly, the observed richnesses range from 74 to 79% of the maximum estimated value. The second- and first-order Jackknife estimators (Jack 2 and Jack 1) consistently generate the highest estimates, and either Bootstrap or MMEan the lowest estimates for the samples.

Despite the smaller number of species observed in the Faraday market, the species richness estimators predict that relatively more species are yet to be sampled in the Faraday market compared to the shops. This is due to the higher number of singletons recorded during the survey. Species richness estimators use the frequency of rare species to predict total sample species richness. The more singletons there are in a sample, the more species are predicted to be present in the sampling ‘universe’ that remain to be captured during a survey.

Overall, results indicate that based on the rate at which the estimators reach an asymptote, the MMMeans is the best estimator of species richness for the four ethnobotanical samples. However, the MMMeans estimates for species sold by the subset of *Indian*- and *Black-owned* shops (334 and 352 ethnospecies respectively) is less than the observed richness for the total number of ethnospecies inventoried in the 50 *muti* shops (i.e. 371 ethnospecies), of which the subsets are part (shaded area Table 2, Figure 3). Therefore, MMMeans was not able to correctly estimate the total richness of the total sample (*All shops*) based on the species present in either of the two data subsets. Instead, Jack 1 better predicts the species richness for the total sample (372 and 379 estimated ethnospecies for *Indian*- and *Black-owned* shops respectively, compared to 371 observed for *All shops*). Similarly, Jack 1 estimates for *Faraday* (Fig. 3a) and *All shops* (Fig. 3b) are close to those observed if the ethnospecies inventories of both samples are amalgamated to form a total species list for the region. When the two samples are combined (*WRand* in Table 4), the observed ethnospecies richness is 436 for 150 samples, which is similar to the 432 and 443 ethnospecies richness estimated by Jack 1 to be potentially present in either *Faraday* or *All shops* respectively. So, while MMMeans reaches an asymptote sooner and *appears* to be the best estimator for the data based on asymptote criteria, Jack 1 seems to provide the best lower-bound estimate of species richness for the individual data sets. If Jack 1 is the lower-bound estimator, then it is possible that Jack 2 estimates approach an upper limit for the number of species likely to be found in *commercial trade* in the region. The number of species actually used is far greater, but not all of them will find their way to the markets. While Chao 2 and ICE sometimes give estimates similar to that of Jack 1, the estimators are initially unstable and are therefore discounted.

Table 4 Ethnospecies richness and percentage complementarity between four data sets, including a study by Botha (2001) on medicinal plants traded on the western boundary of the Kruger National Park (*WBKP*)

	Faraday (n=100)	All Shops (n=50)	WRand Total (n=150)	WBKP (n=17)
Observed ethnospecies richness	349	371	436	185
<i>Complimentarity</i>				
All Shops	34.8% (284)			
WBKP	76.7% (101)	77.3% (103)	78.2% (111)	
<i>Combined richness for both sites</i>				
All Shops	436 (≈595)			
WBKP	453 (≈616)	473 (≈648)	510 (≈700)	

‘WRand Total’ is the combined data set of the Faraday and All Shop surveys. Values in brackets after the percent complementarity are the number of ethnospecies common on both data sets. Values in brackets after the combined ethnospecies richnesses are approximate numbers of corresponding species

A question of sampling sufficiency and the completeness of the inventories is raised because of the failure of most estimators (with the exception of MMMean and the *All shops* curve), and the species accumulation curves, to reach an asymptote. Curves that have an upward trend tend to indicate that an increase in sampling effort would yield more species, as is predicted by the estimators and the difference between the observed and estimated number of ethnospecies. As regards the species accumulation curves, however, the rate of accumulation of new species is also necessary to make comparisons (Soberón and Llorente 1993), especially for certain ethnobotanical samples where the curves may never reach an asymptote because the commercial trade functions within a potentially very large regional floral diversity. Data for *Faraday* indicate that the rate of accumulation of new ethnospecies is 0.9 per trader, whereas the rate for *All shops*, *Black-* and *Indian-owned* shops is 1.5, 2.7 and 3.5 respectively. These figures indicate the sampling effort was satisfactory, and that there would be no real benefit to interviewing more traders and inventorying their medicinal plant stock if, overall, <2 new ethnospecies will be recorded per trader. Consequently, the sample size required for the species accumulation curves to reach an asymptote might be too large to be practical.

Heck et al. (1975) reported that in some situations it might be satisfactory to collect 50%–75% of the total number of species known to occur in a given area, as long as the most common species are recorded. While the total number of species in commercial medicinal plant trade is not precisely known, observed richness is $\geq 74\%$ of the maximum richness predicted by the estimators for all the samples and the most commonly traded species were captured by the survey. However, species represented as singletons/doubletons in an ethnobotanical sample derived from a commercial medicinal plant market are not necessarily ‘rare’ or ‘scarce’ in the wild. Instead, they often represent species that have little or no commercial value and/or have been speculatively harvested and sold to traders. As a result, the species richness estimators are, in principle, estimating total richness and the number of species that remain to be discovered from the preponderance of species with less commercial value.

Complimentarity and similarity

Values for the complimentarity and similarity indices show there to be a low level distinctness and high degree of similarity between the ethnospecies sold in the *Faraday* market and *muti* shops (Tables 4 and 5). About 79% of the species sold in

Table 5 Jaccard and Sørensen’s coefficients of similarity for ethnospecies sold, expressed as percentages

	Faraday (%)	All shops (%)	WRand total (%)	WBKP
<i>Classic Jaccard index</i>				
All shops	65.1			
WRand total	80.0	85.0		
WBKP	23.3	22.7	21.7	
<i>Classic Sørensen index (incidence-based)</i>				
All shops	78.9			
WRand total	88.9	91.9		
WBKP	37.8	37.0	35.7	

the market and shops are common to both samples (Sørensen index, Table 5), and the combined ethnospecies richness is 436 (≈ 595 species) (Table 4). The number of ethnospecies common to both samples is 284 (Table 4). Given that most shop traders purchase plants from the wholesale Faraday market and are in adjacent trading areas, the number of shared taxa and percentage similarity is expected to be high.

By contrast, comparing plants sold in the Witwatersrand and Mpumalanga markets (WBKP) (about 350 km apart) show a low degree of similarity and high degree of distinctness. Plants sold by the vendors in WBKP were 37% and 38% similar to the plants sold in *Faraday* and *All shops* respectively (Sørensen index, Table 5), and the combined ethnospecies richness was 453 (≈ 616 species) for WBKP and *Faraday*, and 473 (≈ 648 species) for WBKP and *All shops* (Table 4). When one combines the *Faraday* and *All shop* samples to produce an amalgamated data set (called 'WRand Total'), and compare it with WBKP, the Jaccard index of similarity is 36% and the shared number of ethnospecies is 510 (or ≈ 700 species). The number of ethnospecies common to both samples is relatively low (111 ethnospecies) and the two data sets are thus distinctly different. Given that $\leq 3\%$ of the plants sold in the Witwatersrand are harvested by gatherers in Mpumalanga, the dissimilarity of the two markets in terms of species richness is to be expected.

The resulting combined species richness for both sites (WRand and WBKP) (≈ 700 species) raises expectations for a high level of species richness commercially traded in and between different markets in South Africa. While no inventory data were available to compare the taxonomic similarities of species commercially traded in markets in KwaZulu-Natal (e.g. the Durban markets), the results would probably have shown the percent similarity to be greater than that of WBKP (i.e. $>40\%$) given that two-thirds of plants sold in the Johannesburg markets originate from KwaZulu-Natal. A factor that would additionally influence these results, however, would be the demand for medicinal plants harvested outside of South Africa (e.g. in Swaziland, Malawi and Mozambique). There are known to be noteworthy trade links between harvesters and markets for species found elsewhere in southern Africa (Williams 2005).

Conclusion

To enhance the value of ethnobotanical studies, the incorporation of suitable quantitative methods into data collection, processing and interpretation improves on the traditional compilation-style approach to the discipline (Höft et al. 1999) and, creates an opportunity for cogent arguments that advance scientific and practical knowledge. In the case of species accumulation functions, richness estimators and similarity measures, the methods broaden the interpretation of species inventories in several ways.

First, plotting species accumulation and rarefaction curves enables comparisons of species density per sample (i.e. the number of species per trader) and hence species richness at similar levels of sampling effort. When the axes are rescaled to individuals (N), the effect of density is removed and overall species richness between data sets can be compared. In addition, the 95% confidence intervals help assess whether the differences in species richness are significant. The shape of the curve, the rate of accumulation of new species and the degree to which curves approach a horizontal asymptote illustrate sampling sufficiency. It was clear from the results that *muti* shops stocked significantly more species than street traders did and that considerably

more sampling was required in the Faraday market before an equivalent level of species richness could be reached. Overall, the species richness was high. Even though the curves did not reach an obvious asymptote, the rate of accumulation of new species with increased sampling effort had levelled off enough for sampling effort to be considered satisfactory.

A well censused site is reportedly one where the species accumulation curve reaches a stable asymptote (Gotelli and Colwell 2001). However, given that the diversity of plants sold in traditional medicine markets is high (Williams et al. 2005) and the potential area from which plants can be harvested is extensive and continuously expanding (≈ 3 million km² including South Africa's immediate neighbours but excluding Namibia), there is some doubt as to whether the curve would *ever* reach a stable asymptote, especially if it hasn't done so after sampling 50 to 100 traders. The same doubt would apply to other ethnobotanical data sets, especially those derived from large, regional plant markets.

Second, is the investigative benefit of species richness estimation. A challenge with ethnobotanical species inventories and using richness estimators is that there are rarely survey scenarios where the actual species richness is known a priori. In the case of species traded in large regional markets such as Johannesburg, species are sold that were not only harvested in South Africa, but also in other southern African countries. Hence, it is difficult to test the performance of the richness estimator by comparing the results with data from sites where the actual species richness is known, because the total number of species that could be used cannot be realistically enumerated. That said, the estimators are still useful for predicting the total species richness of a site/assembly from a pooled set of samples.

The first-order Jackknife (Jack 1) correctly estimated the total ethnospecies richness of the *All shop* sample from the sub-samples of *Indian-* and *Black-owned* shops within 1–9 ethnospecies respectively. When the *Faraday* and *All shop* species inventories were combined ('WRand'), then the total number of ethnospecies was also similar to the Jack 1 estimate. Therefore, if Jack 1 is the best lower-bound estimator (because it predicted WRand richness), then the lower-bound estimate for the number of ethnospecies in commercial trade in the Witwatersrand ranges between 432 and 442 (equivalent to 588 and 606 species respectively). This indicates that the surveys captured at least 81% of the total number of species likely to be in commercial trade in the region. Alternatively, if Jack 2 is taken as the next best estimate to account for species not captured by either of the surveys, then total ethnospecies richness for the Witwatersrand traditional medicine trade is in the region of 473 ethnospecies or >643 species. The surveys would have thus accounted for $\approx 74\%$ of the total richness. This is more than adequate given Heck et al. (1975) recommended that the target should be 50%–75%. For the data sets investigated, the Jack 2 probably give an upper-bound approximation for the number of species most likely to be commercially traded in the region, including species that are speculatively harvested.

The least useful of the estimators are the species abundance models. Parametric models were originally developed to quantify abundance patterns that were evident in certain ecological communities and to test hypotheses about their underlying organization, including resource partitioning (Ludwig and Reynolds 1988; Magurran 1988). The generally poor fit of the data sets to the species distributions predicted by the models makes the rationale for using the models to estimate richness for ethnobotanical samples questionable. The relative abundances of ethnospecies in

ethnobotanical samples are not expected to have characteristics inherent in natural populations. Thus, the underlying assumptions that the models make a priori about the character of the data make the validity of fitting the models even more debatable. The log normal model is known to fit most 'large' data sets reasonably well (Hayek and Buzas 1997), and it is therefore probable that the good fit observed in the Faraday data set is due to the large number of samples. However, the definition of what constitutes an 'adequately large data set' is vague. The data sets investigated were large, but it was not possible to determine from the abundance models whether the richness estimates were upper- or lower-bound predictions. Therefore, collecting data adequate for fitting parametric models will not satisfactorily resolve the time and effort cost to benefit ratio for further sampling—which is one of the objectives when estimating species richness.

Third, is the usefulness of complementarity and similarity measures. They are the easiest of the measures to calculate and indicate the degree to which plant inventories compiled in different local and regional markets are distinct or similar. The percent similarity also infers the overall richness of species in commercial trade, especially between markets located in other parts of the country. If the WBKP sample has an observed ethnospecies richness of 185 and is only 36% similar to the combined WRand richness of 436, then potentially many more species exist along a 'trade gradient' between the two provinces. It would be similarly useful to compare the similarity of species sold in the markets of KwaZulu-Natal (e.g. Cunningham 1988) and the Eastern Cape (e.g. Dold and Cocks 2002), where the richness of species in trade is known to be high.

When using incidence-based species richness measures with ethnobotanical data, the least one should do is plot the sample-based rarefaction curves to compare the richness and density of the data sets at similar levels of sampling effort and assess the relative value of additional sampling. The benefits of estimating species richness, however, requires further consideration. Overall, the incidence-based measures described in this paper add value to the description of ethnobotanical species inventories and there is merit in applying these techniques to other case studies.

References

- Begossi A (1996) Use of ecological methods in ethnobotany. *Econ Bot* 50:280–289
- Bunge J, Fitzpatrick M (1993) Estimating the number of species: a review. *J Am Stat Assoc* 88:364–373
- Botha J (2001) Perceptions of availability and values of medicinal plants traded on the western boundary of the Kruger National Park, South Africa. MSc dissertation, University of the Witwatersrand, Johannesburg
- Chiarucci A, Enright NJ, Perry GLW, Miller BP, Lamont BB (2003) Performance of nonparametric species richness estimators in a high diversity plant community. *Divers Distrib* 9:283–295
- Colwell RK (2001) EstimateS: statistical estimation of species richness and shared species from samples. Version 6. User's guide and application published at: <http://www.viceroy.eeb.uconn.edu/estimates>
- Colwell RK (2005) EstimateS: statistical estimation of species richness and shared species from samples. Version 7.5. User's Guide and application published at: <http://www.purl.oclc.org/estimates>
- Colwell RK, Coddington JA (1995) Estimating terrestrial biodiversity through extrapolation. In: Hawksworth DL (ed), *Biodiversity, measurement and estimation*. Chapman and Hall, pp 101–118
- Colwell RK, Mao CX, Chang J (2004) Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* 85(10):2717–2727

- Cunningham AB (1988) An investigation of the herbal medicine trade in Natal-Kwazulu. Investigational Report No. 29, Institute of Natural Resources, University of Natal, Pietermaritzburg
- Cunningham AB (2001) Applied ethnobotany: people, wild plant use and conservation. People and Plants Conservation Manual, Earthscan, London
- Dold AT, Cocks ML (2002) The trade in medicinal plants in the Eastern Cape Province, South Africa. *South African J Sci* 98(11–12):589–597
- Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol Lett* 4:379–391
- Gotelli NJ, Entsminger GL (2001) EcoSim: null models software for ecology. Version 7.0. Acquired Intelligence Inc. and Kesey-Bear. <http://www.homepages.together.net/~gentsmin/ecosim.htm>
- Hanazaki N, Tamashiro JY, Leitão-Filho HF, Begossi A (2000) Diversity of plant use in two Caçara communities from the Atlantic Forest coast, Brazil. *Biodiv Conserv* 9:597–615
- Hayek LC, Buzas MA (1997) Surveying natural populations. Columbia University Press, New York
- Heck KL, van Belle G, Simberloff D (1975) Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology* 56:1459–1461
- Höft M, Barik SK, Lykke AM (1999) Quantitative ethnobotany: applications of multivariate and statistical analysis in ethnobotany. People and Plants working paper 6. UNESCO, Paris
- Krebs CJ (1989) Ecological methodology. Harper and Row, New York
- Ludwig JA, Reynolds JF (1988) Statistical ecology—a primer on methods and computing. John Wiley and Sons, Toronto
- Magurran A (1988) Ecological diversity and its measurement. Princeton University Press, Princeton
- Magurran A (2004) Measuring biological diversity. Blackwell Publishing, Oxford
- Pielou EC (1975) Ecological diversity. Wiley InterScience, New York
- Soberón J, Llorente J (1993) The use of species accumulation functions for the prediction of species richness. *Conserv Biol* 7:480–488
- Sørensen L (2004) Composition and diversity of the spider fauna in the canopy of the montane forest in Tanzania. *Biodiv Conserv* 13:437–452
- Toti DS, Coyle FA, Miller JA (2000) A structured inventory of Appalachian grass bald and heath bald spider assemblages and a test of species richness estimator performance. *J Arachnol* 28:329–345
- Williams VL (2003) Hawkers of health: an investigation of the Faraday Street traditional medicine market in Johannesburg, Gauteng. Plant Ecology and Conservation Series No. 15, University of the Witwatersrand (Report to the Gauteng Directorate for Nature Conservation, DACEL), p 215
- Williams VL, Balkwill K, Witkowski ETF (1997) *Muthi* traders on the Witwatersrand, South Africa – an urban mosaic. *S Afr J Bot* 63(6):378–381
- Williams VL, Balkwill K, Witkowski ETF (2000) Unravelling the commercial market for medicinal plants and plant parts on the Witwatersrand, South Africa. *Econ Bot* 54(3):310–327
- Williams VL, Witkowski ETF, Balkwill K (2005) Application of diversity indices to appraise plant availability in traditional medicinal markets in Johannesburg, South Africa. *Biodiv Conserv* 14:2971–3001
- Williams VL (2005) Market networks for medicinal bark in southern Africa. In: Trees for health—forever: Implementing sustainable medicinal bark use in southern Africa. Regional Workshop, Johannesburg, South Africa, 1–3 November 2005. DFID-FRP Project R8305, UK