Transactions of the Kansas Academy of Science
Vol. 116, no. 3-4
p. 149-155 (2013)

A statistical analysis of medicinal plants: A case study of plant families in Kansas and the Great Plains

Kelly Kindscher, Steve Corbett[1], Katrina McClure

Kansas Biological Survey, The University of Kansas, Lawrence, Kansas 66047
1. Corresponding author: scorbett@ku.edu

We analyzed ethnobotanical data on Kansas plants to evaluate differences among families for medicinal uses by Native Americans. We compared three different statistical methods used in previous ethnobotanical studies for use in our Native Medicinal Plant Research Program, which seeks to determine which plant families are more likely to contain species with medicinal compounds. The three approaches were linear regression, binomial approach, and Bayesian analysis. All were useful for determining medicinal plant use differences among plant families, although regression analysis was most useful for our purposes. Asteraceae and Fabaceae are abundant in the open sun environments of Great Plains grasslands and contain high numbers of medicinal plants. In contrast, although grasses (Poaceae) and sedges (Cyperaceae) also are species-rich and ecologically abundant, each is underrepresented as being used by Native Americans as medicine, which can be explained at least partially by their paucity of secondary compounds.

Introduction

Researchers the past 35 years have used statistics to assess the manner in which plants were selected for use as medicine by cultural groups, most notably indigenous (Moerman 1979a; Bennett and Husby 2008; Weckerle et al. 2011). Questions and ideas driving this type of research are multifaceted, ranging from anthropological questions concerning the proportion of medicinal species used locally, and pharmacological questions largely interested in using ethnographic tools to screen plants for chemical analysis. Although cross-cultural regional analyses have been noted as having the greatest potential for pharmacologically relevant findings, few regional analyses are available (Saslis-Lagoudakis et al. 2011).

The number of species in a region should correlate with the number of medicinal plants identified by a culture if random selection of plants as medicines is the underlying mechanism behind medicinal plant recognition. In such circumstances, any therapeutic value is assumed to be due to the placebo effect, as opposed to any real medicinal efficacy (Moerman, 1979a). Following this premise, a placebo hypothesis was first posited by Moerman (1979a, 1979b), by which the proper computational methods could elucidate the selection pattern of plant species for medicines. A non-random selection pattern of species used that deviates from that expected, refutes the idea that plants were chosen as medicines due to proximity or familiarity alone. A non-random selection pattern also provides evidence for folk beliefs and potential efficacy.

This research approach has broad implications. First, one can use statistical analysis to assess randomness of plant selection, including techniques such as regression analysis, the binomial method, and the Bayesian approach (Moerman 1979a; Bennett and Husby 2008; Weckerle et al. 2011). This assessment allows judgments to be made regarding whether plants were selected by random chance or due to some stated effect that made them useful to people in treating various illnesses, ailments, diseases, or infections. Second, by using all three statistical methodologies, we can evaluate the efficacy of each type of analysis for the research conducted by the Native Medicinal Plant Research Program, using data from our Prairie Ethnobotany Database of Great Plains plants.

The Native Medicinal Plant Research Program is a collaborative project of the Department of Medicinal Chemistry and the Kansas Biological Survey, both housed at the University of Kansas (see: http://nativeplants.ku.edu). The mission of the program is to promote public understanding of medicinal uses of native Kansas plants and to provide scientific validation of traditional ecological knowledge, that is, we are demonstrating that the vast majority of plants that Native Americans used for medicine had active medicinal constituents. The program began in 2009 as a broad-based search for medicinal compounds in plants native to the Great Plains in the U.S. and has had substantial success (Kindscher et al., 2012; Zhang et al., 2011; Zhang et al., 2012; and Arraya et al. 2012). Plants are collected in the field and targeted species are identified based on ethnobotanical data and relationships to known medicinal plants.

The Native Medicinal Plant Research Program developed the Prairie Ethnobotany Database (PED) to house ethnobotanical information on plants of the Great Plains. The PED contains data similar to that found in the University of Michigan – Dearborn's Native American Ethnobotany database (http://herb.umd.umich.edu/), developed by Dan Moerman – and contains some information directly from that database. Additional data were added from the existing published literature (including Kindscher 1987, 1992; Kindscher and Hurlburt 1998) and from unpublished historical ethnographies. The database currently contains information on over 21,000 medicinal, edible, dye, fiber, and other uses for nearly 1,450 plant species.

Methods

The statistical methodologies followed Moerman (1979a), Moerman et al. (1999), Bennett and Husby (2008), and Weckerle et al. (2011), and included regression analysis, binomial analysis, and Bayesian analysis, respectively. Bennett and Husby (2008) also conducted a goodness of fit test on the contingency table to examine all plants, an approach also used here. Data on plants identified as being native to Kansas were extracted from the Prairie Ethnobotany Database. We chose Kansas as an appropriate subset not only because we work here, but also because it is a relatively homogeneous bio-geographical area for which we have good data. Identification of plants as native to Kansas was based on lists from the R. L. McGregor Herbarium at the University of Kansas and from the USDA Plants Database (http://plants.usda.gov/). Only those plants identified as being "medicinal" in the ethnographic literature were retained for the analysis. Counts of the number of medicinal plants per family were completed, with a mean expected frequency computed from the number of medicinal plant species per family. This mean expected frequency was used to calculate the expected ranges required of the various statistical approaches. Using the expected frequencies, these statistical methodologies would indicate which families are significantly over- or under-represented with regard to the number of ethnobotanically medicinal species compared to that expected by random chance alone.

Results

A total of 151 plant families, comprising the 2,316 species of Kansas plants (Freeman 2006) were analyzed, of which 426 were identified as having medicinal properties. This included both native and non-native species. The mean frequency of medicinal plant species per family

Table 1. Results from the three different analytical approaches.

REGRESSION ANALYSIS: Families showing the most extreme deviation from the expected, based on residuals (Standard deviation of the residual = 4.25; Standard deviation of the standardized residual = 1.14).								
Family	**Total #**	**Medicinal**	**Fit**	**Expected**	**SE Fit**	**Residual**	**St Resid**	**%**
Asteraceae	322	75	52.5	59.2	2.8	22.5	6.9	0.23
Ranunculaceae	43	23	7.3	7.9	0.4	15.7	3.7	0.53
Fabaceae	154	39	25.3	28.3	1.3	13.7	3.4	0.25
Lamiaceae	66	22	11.0	12.1	0.6	11.0	2.6	0.33
Apiaceae	46	15	7.8	8.5	0.4	7.2	1.7	0.33
Cyperaceae	142	7	23.3	26.1	1.2	-16.3	-4.0	0.05
Poaceae	261	8	42.6	48.0	2.2	-34.6	-9.5	0.03

BINOMIAL ANALYSIS: Families showing the most extreme deviation from the expected for medicinal use (Statistical significance < 0.05).						
Family	**Total #**	**Medicinal**	**Expected**	**Difference**	**PROB**	**%**
Ranunculaceae	43	23	7.9	15.1	<0.001	0.53
Lamiaceae	66	22	12.1	9.9	0.003	0.33
Apiaceae	46	15	8.5	6.5	0.015	0.33
Asteraceae	322	75	59.2	15.8	0.016	0.23
Fabaceae	154	39	28.3	10.7	0.020	0.25
Cornaceae	5	3	0.9	2.1	0.046	0.60
Cyperaceae	142	7	26.1	-19.1	<0.001	0.05
Poaceae	261	8	48.0	-40.0	<0.001	0.03

BAYESIAN ANALYSIS: Families showing the most extreme deviation from the expected for medicinal use (Based on the inferior and superior levels of 0.17 and 0.20).					
Family	**Total #**	**Medicinal**	**Inferior Levels**	**Superior Levels**	**Over/Under**
Cornaceae	5	3	0.222778	0.881883	Over
Dryopteridaceae	12	5	0.192232	0.684222	Over
Ranunculaceae	43	23	0.388472	0.675386	Over
Salicaceae	16	6	0.184437	0.616716	Over
Apiaceae	46	15	0.208641	0.471181	Over
Lamiaceae	66	22	0.231533	0.453951	Over
Fabaceae	154	39	0.191206	0.327529	Over
Asteraceae	322	75	0.190102	0.282137	Over
Cyperaceae	142	7	0.024459	0.098256	Under
Poaceae	261	8	0.015826	0.059276	Under

was 18.4%, which constituted the expected frequency. The plant families that showed significant (defined in Table 1) deviation from the expected are presented in Table 1.

Linear Regression

Regressing the number of medicinal species per family against total species showed that the numbers of medicinal plants per family differs significantly from that expected from chance, with an R^2 of 68.6%. The table shows only those families with residuals and standardized residuals greater than one standard deviation, which is how Moerman et al. (1999) defined significance. The direction of the residuals (positive or negative) indicates whether a family is over- or under-represented regarding the number of medicinal species it contains. Seven families showed significant variation from the expected value. Five of the families

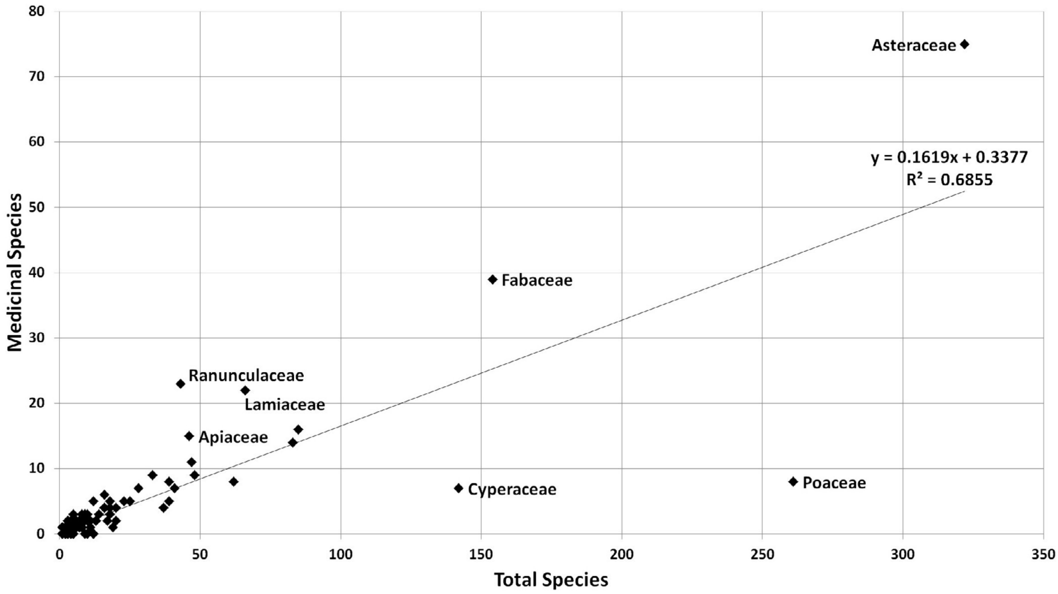

Figure 1. Scatter plot of families plotting medicinal species vs. total species.

had higher than expected frequencies, with Asteraceae having the highest residuals (that is having a greater than average number of medicinal plant species), and two families had lower than expected frequencies. A scatter plot (Figure 1) shows the distribution of species per family and indicates those diverging from the regression line.

Binomial Approach

A contingency table analysis was conducted on the entire model, where a goodness of fit test on the contingency table for Kansas flora revealed a significant departure from homogeneity (P = 0.003). This indicated that medicinal species were not evenly distributed among families, consistent with the results from regression.

The binomial approach identified the same subset of families as the regression analysis, but also included Cornaceae as one that deviated significantly from the expected frequency with a high number of medicinal species. Figure 2 shows the medicinal plant percentages for those families. The order of significance differed from that of the regression analysis, with Ranunculaceae being the most significant, followed by Lamiaceae, with Asteraceae dropping to be the fourth most significant family to be over-represented with medicinal plant species.

Bayesian Analysis

Bayesian analysis identified the same families as regression and binomial analysis, and added Dryopteridaceae and Salicaceae. The order of significance was different as well, with Cornaceae and Dryopteridaceae having greater numbers represented than either Ranunculaceae (top binomial) or Asteraceae (top regression). Asteraceae, being identified by regression analysis as the family most over-represented, became much less significant. As in the other methods, Cyperaceae and Poaceae had significantly fewer medicinal plant species than expected, with Poaceae having by far the fewest (see Table 1).

Discussion

We found a higher rate of species (18%) in Kansas having medicinal uses compared to those previously documented for North America, which found 12% of species having medicinal uses (Moerman 1979a; Moerman et

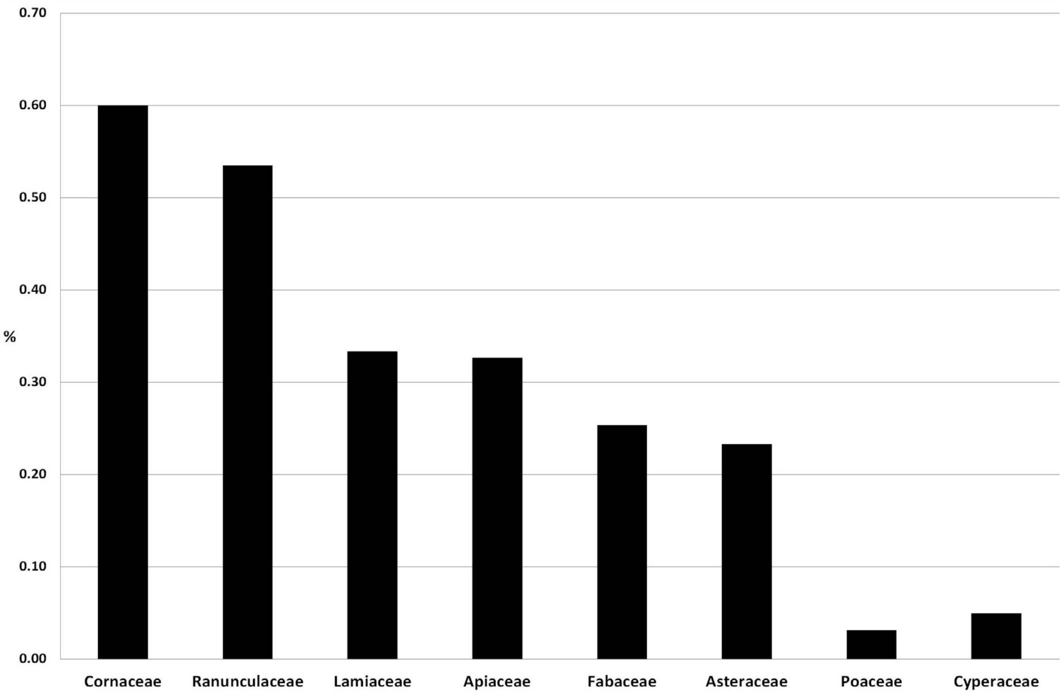

Figure 2. Percentages of medicinal plants per plant family showing those families identified as deviating significantly from the expected rate using binomial analysis.

al. 1999). This result may reflect the facts that we and other researchers have continued to add species and medicinal plant uses to the data set, and the significant levels of ethnobotanical research now occurring in the Great Plains. We also found that some families that were over-represented regarding ethnobotanical usage in Kansas were over-represented across North America in general. Kansas had fewer such families (5 versus 14): Asteraceae, Ranunculaceae, Fabaceae, Lamiaceae, and Apiaceae. In contrast, Moerman (1979a) found the additional families to be over-represented ethnobotanically in North America: Rosaceae, Pinaceae, Caprifoliaceae, Liliaceae, Polemoniaceae, Chenopodiaceae, Betulaceae, Berberidaceae, and Pyrolaceae. The greater number of families identified by Moerman also reflects the expanse of North America and its significant range of ecological zones and habitats, coupled with significant study and recording of Native Americans uses of medicinal plants.

Asteraceae (with 322 species) and Fabaceae (154 species) are the two largest plant families in Kansas. The Asteraceae have many well-known genera of medicinal plants in Kansas, including *Echinacea* (coneflower), *Achillea* (yarrow), *Artemisia* (mugwort/wormwood), *Erigeron* (fleabane), *Solidago* (goldenrods) and *Symphyotrichum* (asters). Fabaceae also has genera with medicinal applications, such as *Amorpha* (leadplant), *Astragalus* (locoweed), *Baptisia* (wild indigo), and *Pediomelum* (prairie turnip). The large numbers of medicinal species in these two families can perhaps be partially explained by the fact that the full light of a grassland habitat, with little shadowing by a canopy, might drive higher than normal levels of photosynthesis and energy capture for the development of metabolically expensive secondary compounds (Bjorkman et al 2011).

Three other families in Kansas (Ranuculaceae, Lamiaceae, and Apiaceae) also are over-represented. All three are rich in secondary metabolites, with several species-rich and important genera, and Kansas is particularly rich with these species because of the open meadow/prairie habitats. For Ranunculaceae the important genera are *Anemone* (windflower), *Delphinium* (larkspur), *Thalictrum* (meadow rue), and such well-known species as goldenseal (*Hydrastis canadensis*). For Lamiaceae, the important genera are well-known herbal plant groups and include *Agastache* (hyssop), *Hedeoma* (false pennyroyal), and *Monarda* (beebalm). Apiaceae contains a large group of species with herbal/medicinal uses, including *Osmorhiza* (sweetroot). The numbers of species in Kansas for the other families with medicinal uses included Cornaceae (5), Dryopretidaceae (12), and Salicaceae (16). Although the number of species identified with medicinal uses in each is relatively low (3, 5, and 6 respectively), they resulted in high percentages.

It is notable that grasses (Poaceae) and grass-like plants (Cyperaceae) are both abundant and species-rich, yet under-represented in Kansas in their medicinal uses. Moerman (1979a, 1999) also found Poaceae and Cyperaceae to be under-represented medicinally. However, it is known that these families have relatively fewer secondary compounds (Bourgaud et al. 2001). It is widely thought that selection did not favor the evolution of secondary compounds because these families evolved with grazers during the Miocene (Stebbins 1981), have basal leaf growth (rather than from the end of stems as in most species, and have high levels of silica bodies, which deter grazing by many insects and mammals. Given that grasses are abundant in the region, a random medicinal use of species in these families would suggest higher numbers of species used than our analysis revealed. Our analysis confirms that traditional knowledge and medicinal practices of Plains Indians were not just based on ceremony or proximity and profusion, but were indeed based on efficacy. Plains Indians were not oblivious to grasses, however. The onset of growth of grasses such as *Hesperostipa comata* (needle-and-thread grass) often coincided with the best time for hunting buffalo cows, for instance (Johnston 1987). In this case, Plains Indian knowledge of these grasses was utilitarian and based on a valid relationship. There is no reason to assume that their knowledge of medicinal plants was any less valid.

All three statistical methods identified some plant families as having more medicinal plant species than expected. Bayesian analysis was the most inclusive method of the three, whereas regression analysis was the most exclusive. Regression analysis tended to exclude small families containing few total species, as noted previously (Bennett and Husby 2008; Weckerle et al. 2011; Moerman 2012). For the purposes of the Native Medicinal Plant Research Program, all statistical approaches are useful. The most efficacious method would be the regression analysis originally proposed by Moerman. We would be more likely to target collections in a family with many medicinal species, with the expectation that such a family includes a relatively high number of medicinal species, providing us with more species to examine chemically.

We would be more likely to examine non-medicinal plant species for active compounds from the families identified as over-represented using the regression method. Though the Cornaceae, Dryopteridaceae, and Salicaceae all showed a high percentage of medicinal plants, the overall numbers of species are low, and in fact they all have fewer total medicinal plant species than the larger Cyperaceae and Poaceae families, which show lower than expected numbers of medicinal plants. Thus a strategic application of our analysis would be to focus the study of Great Plains medicinal plants on the Asteraceae and Fabaceae.

Literature Cited

Araya, J., Binns, F., Kindscher, K. and Timmermann, B. 2012. "Verticillosides A–M: Polyoxygenated pregnane glycosides from *Asclepias verticillata* L." Phytochemistry 78:179–89.

Bennett, B.C. and Husby, C.E. 2008. Patterns of medicinal plant use: an examination of the Ecuadorian Shuar medicinal flora using contingency table and binomial analyses. Journal of Ethnopharmacology 116:422–430.

Björkman, M., Klingen, I., Birch, A.N.E., Bones, A.M., Bruce, T.J.A., Johansen, T.J., Meadow, R., Molmann, J., Seljasen, R., Smart, L.E. and Stewart, D. 2011. Phytochemicals of Brassicaceae in plant protection and human health–Influences of climate, environment and agronomic practice. Phytochemistry 72:538–556.

Bourgaud, F., Gravot, A., Milesi, S. and Gontier, E. 2001. Production of plant secondary metabolites: A historical perspective. Plant Science 161(5):839-851.

Freeman, C.C. 2006. Keys to the Vascular Plants of Kansas. R. L. McGregor Herbarium, Division of Botany, Natural History Museum and Biodiversity Research Center, and Kansas Biological Survey, University of Kansas, Lawrence, Kansas, 292 pp.

Johnston, A. 1987. Plants and the Blackfoot. Lethbridge Historical Society, Lethbridge, Alberta, Canada 68 pp.

Kindscher, K., Long, Q., Corbett, S., Bosnak, K., Loring, H., Cohen, M. and Timmermann, B.N. 2012. The ethnobotany and ethnopharmacology of *w*ild Tomatillos, *Physalis longifolia* Nutt*.,* and related *Physalis* species: A Review. Economic Botany 66:298–310.

Kindscher, K. 1992. *Medicinal Wild Plants of the Prairie – An Ethnobotanical Guide*. University Press of Kansas, Lawrence, 336 pp.

Kindscher, K. 1987. *Edible Wild Plants of the Prairie – An Ethnobotanical Guide*. University Press of Kansas, Lawrence, 288 pp.

Kindscher, K., and Hurlburt, D. 1998. Huron Smith's Ethnobotany of the Hocak (Winnebago). Economic Botany 52:352–372.

Moerman, D.E. 1979a. Symbols and selectivity: a statistical analysis of Native American medical ethnobotany. Journal of Ethnopharmacology 1:111–119.

Moerman, D.E. 1979b. Anthropology of symbolic healing. Current Anthropology 20:59-80.

Moerman, D.E. 2012. Commentary: Regression residual vs. Bayesian analysis of medicinal floras. Journal of Ethnopharmacology 139:693–694.

Moerman, D.E., Pemberton, R.W., Kiefer, D. and Berlin, B. 1999. A comparative analysis of five medicinal floras. Journal of Ethnobiology 19:46-67.

Saslis-Lagoudakis, C.H., Williamson, E.M., Savolainen, V. and Hawkins, J.A. 2011. Cross-cultural comparison of three medicinal floras and implications for bioprospecting strategies. Journal of Ethnopharmacology 135:467-487.

Stebbins, G. L. 1981. Coevolution of grasses and herbivores. *Annals of the Missouri Botanical Garden* 1981:75-86.

Weckerle, C.S., Cabras, C., Castellanos, M.E, and Leonti, M. 2011. Quantitative methods in ethnobotany and ethnopharmacology: Considering the overall flora - hypothesis testing for over- and underused plant families with the Bayesian approach. Journal of Ethnopharmacology 137(1):837–843.

Zhang, H., Samadi, A.K., Gallagher, R.J., Araya, J.J., Tong, X., Day, V.W., Cohen, M.S., Kindscher, K., Gollapudi, R. and Timmermann, B.N. 2011. Cytotoxic withanolide constituents of *Physalis longifolia*. Journal of Natural Products 74:2532−2544.

Zhang, X., Samadi, A.K., Roby, K.F., Timmermann, B. and Cohen, M.S. 2012. Inhibition of cell growth and induction of apoptosis in ovarian carcinoma cell lines CaOV3 and SKOV3 by natural withanolide withaferin A. Gynecologic Oncology 124:606–612.